SYMPOSIUM: HIGHLIGHTS FROM THE FIRST COMBINED 2011 MEETING OF THE MUSCULO-SKELETAL TUMOR SOCIETY AND CONNECTIVE TISSUE ONCOLOGY SOCIETY

Using the CUSUM Test to Control the Proportion of Inadequate Open Biopsies of Musculoskeletal Tumors

David J. Biau MD, PhD, Kurt R. Weiss MD, MSc, Rej S. Bhumbra PhD, FRCS (Tr & Orth),
Darin Davidson MD, MSc, Chris Brown MD, PhD, Jay S. Wunder MD, MSc,
Peter C. Ferguson MD, MSc

Published online: 12 September 2012
© The Association of Bone and Joint Surgeons® 2012

Abstract

Background Biopsies of musculoskeletal tumors lead to alterations in treatment in almost 20% of cases. Control charts are useful to ensure that a process is operating at a predetermined level of performance, although their use has not been demonstrated in assessing the adequacy of musculoskeletal biopsies.

Questions/purposes We therefore (1) assessed the incidence of and the reasons for inadequate musculoskeletal biopsies when following guidelines for performing the procedure; and (2) implemented a process control chart, the

CUSUM test, to monitor the proportion of inadequate biopsies.

Methods We prospectively studied 116 incisional biopsies. The biopsy was performed according to 10 rules to (1) minimize contamination in the tissues surrounding the tumor; and (2) improve accuracy. A frozen section was systematically performed to confirm that a representative specimen was obtained. Procedures were considered inadequate if: (1) another biopsy was necessary; (2) the biopsy tract was not appropriately placed; and (3) the treatment provided based on the diagnosis from the biopsy was not appropriate.

Results Five (4.3%) of the 116 incisional biopsy procedures were considered failures. Three patients required a second repeat open biopsy and two were considered to receive inappropriate treatment. No alarm was raised by the control chart and the performance was deemed adequate over the monitoring period.

Conclusions The proportion of inadequate musculoskeletal open biopsies performed at a referral center was low. Using a statistical process control method to monitor the failures provided a continuous measure of the performance.

Each author certifies that he or she, or a member of their immediate family, has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.

Each author certifies that his or her institution approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

This work was performed at the University Musculoskeletal Oncology Unit, Mount Sinai Hospital, Toronto, Ontario, Canada.

D. J. Biau
Department of Orthopaedic Surgery,
Hospital Cochin, Paris, France
e-mail: djmbiau@yahoo.fr

K. R. Weiss
Division of Musculoskeletal Oncology, Department
of Orthopaedic Surgery, Shadyside Medical Center,
Pittsburgh, PA, USA

R. S. Bhumbra
London Sarcoma Service, Royal National
Orthopaedic Hospital, Stanmore, UK

D. Davidson
Department of Orthopaedics and Sports Medicine,
University of Washington, Seattle, WA, USA

C. Brown, J. S. Wunder, P. C. Ferguson (✉)
University Musculoskeletal Oncology Unit, Mount Sinai
Hospital, 600 University Avenue, Room 476G,
Toronto, ON M5G 1X5, Canada
e-mail: pferguson@mtsinai.on.ca

C. Brown, J. S. Wunder, P. C. Ferguson
Sarcoma Site Group, Department of Surgical Oncology,
Princess Margaret Hospital, Toronto, ON, Canada

Introduction

It is estimated that approximately one adult in four will develop a soft tissue mass over their lifetime but only one in 200,000 of these will be a sarcoma [21]. The incidence of soft tissue sarcoma in the United States is approximately seven per 100,000 people. Primary bone tumors are even less common, occurring in one per 100,000 people [16]. Imaging modalities such as plain radiographs, CT, and MRI are often adequate to diagnose benign bone and soft tissue lesions in the context of corroborative findings on history and physical examination. However, patients who present with a soft tissue or bone lesion that raises concern for malignancy will often require a biopsy to confirm pathology that requires treatment [12]. Although a definitive diagnosis will not always be reached with a biopsy, it is expected that enough information will be added to the clinical and radiological data so that appropriate care can be provided. Despite the availability of percutaneous core biopsies, incisional biopsies provide unaltered pretreatment tissue specimens for research purposes and are sometimes necessary when a core or needle biopsy fails to provide a definitive diagnosis [20].

In two seminal reports [12, 13], Mankin and colleagues documented the hazards of the biopsy. In these studies, conducted among members of the Musculoskeletal Tumor Society, as many as 8% to 10% of biopsies were nonrepresentative or technically inadequate, 14% to 18% were associated with major errors in diagnosis, 18% to 19% led to alterations in treatment, and 9% to 10% adversely affected the patient's outcome [12, 13]. No substantial improvement was reported over a 15-year period despite extensive physician education [12, 13]. Therefore, minimizing the proportion of inadequate biopsies remains paramount in providing adequate care to patients.

Control charts are increasingly being used to monitor the quality of health care provided to patients [3]. Control charts are designed to detect unusual sources of variability in a series of values collected over time. Patients, doctors, and public health authorities are using these methods to detect and, if necessary, correct inadequate performance [6, 9, 10, 23, 24]. The use of these methods such as the cumulative summation (CUSUM) test [17] has been proposed in various surgical and interventional specialties [2, 5, 11, 15, 18]. This control chart is designed to detect the deviation from target of some characteristic of a product. For instance, it may be used to detect an increase in the postoperative infection rate or in the proportion of THAs that fails to meet some prespecified standards [2]. These methods provide ongoing feedback to physicians learning a technique [18], using new implants [2], or simply attempting to reach some target level of technical competence [6, 11]. Because open biopsies impose some risk

potentially adversely affecting the final outcome of the patient [12, 13], using a tool that can provide ongoing feedback regarding the adequacy of this procedure is relevant and has not been done before.

We therefore (1) prospectively assessed the incidence of and the reasons for inadequate musculoskeletal biopsies when following guidelines for performing the procedure; and (2) implemented a process control chart, the CUSUM test, to monitor the proportion of inadequate biopsies.

Patients and Methods

This prospective study took place from January 2010 to December 2010 at one tertiary care referral center for the treatment of bone and soft tissue tumors after approval by the local research ethics board. During that time we performed 266 biopsies of bone and soft tissue tumors. For this study we included patients if they underwent an open biopsy of a bone or soft tissue lesion of the extremity or trunk. We excluded patients with retroperitoneal sarcoma and excluded approximately 150 patients who underwent only a needle/core biopsy or an excisional biopsy. These exclusions left 116 incisional biopsies during the monitoring period (Table 1). There were 53 biopsies in women (46%); the median age was 48 years (first quartile to third quartile, 33–62 years). Fifty-seven of the 116 patients (49%) presented with a soft tissue lesion and 26 (22%) had previously undergone an unsuccessful needle or core biopsy. The location of bone and soft tissue lesions varied widely (Table 2). In eight patients (7%), the initial frozen sections showed that the tissue was not representative and would likely not be diagnostic and additional tissue samples had to be collected. A specimen was considered representative if the sample showed tissue compatible with the area targeted by the surgeon based on clinical and imaging workup and on intraoperative findings and showed viable lesional tissue. A specimen showing only normal tissue was not considered as representative. For instance, in one patient the MR study showed local recurrence of a previously resected well-differentiated liposarcoma (Fig. 1). Although most of the tumor shows signal compatible with a well-differentiated liposarcoma on both T1 (Fig. 1A) and T2 (Fig. 1B) -weighted sequences, part of the tumor suggests the presence of a higher grade component (red area on Fig. 1C–D) indicating a possible dedifferentiated liposarcoma. At the time of the frozen section, the specimen will be considered representative if the tissue shows tumor cells compatible with a high-grade sarcoma. If the pathologist only finds cells suggestive of a low-grade liposarcoma, the surgeon will have to resample the tumor. The final pathology of the biopsy reported malignancy in 72 patients (62%), no evidence of

Table 1. Characteristics of patients, biopsies, and outcome per month and overall

Patient characteristic, biopsy, and outcome	January	February	March	April	May	June	July	August	September	October	November	December	Total
Number of patients	12	8	12	11	12	11	6	13	7	11	10	3	116
Age (median years)	48	44	40	44	60	50	39	48	48	52	50	51	48
Sex (male)	7	4	10	5	7	4	3	6	3	8	5	1	63
Previous biopsy	2	2	4	1	1	0	2	1	2	4	6	1	26
Type (soft)	5	4	5	4	9	4	2	6	4	3	9	2	57
Frozen section (lesional)	10	7	11	9	12	10	6	12	7	11	10	3	108
Pathology: malignant	6	5	7	8	11	5	1	9	4	8	5	3	72
Pathology: no malignancy	4	3	5	3	1	5	4	3	1	2	4	0	35
Pathology: indeterminate	2	0	0	0	0	1	1	1	2	1	1	0	9
Outcome (failure)	1	1	0	0	0	1	0	1	0	1	0	0	5

Table 2. Location of the lesions

Soft-tissue lesions		Bone lesions	
Area	Number	Bone	Number
Neck	2	Vertebra	1
Trunk	8	Rib	1
Shoulder girdle	3	Clavicle	3
Arm	4	Scapula	2
Elbow region	2	Humerus	5
Forearm	1	Radius	1
Buttock	2	Ulna	1
Thigh	28	Carpal bones, metacarpals, phalanges	1
Leg	6	Sacrum	4
Ankle/foot	1	Pelvis	11
		Femur	14
		Tibia	11
		Fibula	4

malignancy in 35 patients (30%), and a lesion with indeterminate malignant potential in nine (8%) (Table 3; observations 1 to 9; two cartilaginous tumors, two tumors with necrotic debris, two atypical lymphocytic infiltrates, one solitary fibrous tumor, one smooth muscle tumor, and one soft tissue giant cell rich tumor) (Appendix 2). Some patients had a lesion with indeterminate malignant potential or one of various discrepancies (Table 3). Seventy-three patients (63%) underwent further surgical treatment of the lesion.

All patients underwent evaluation with a thorough clinical history and physical examination followed by analysis of all imaging studies including plain radiographs in the case of bone tumors and MRI in both bone and soft tissue tumors. For patients who appeared to have benign

aggressive or primary bone malignancies and high-grade soft tissue sarcomas, especially if associated with images suggestive of internal necrosis (reflected by low intensity signal on T1-weighted images; high signal intensity on T2-weighted images; no enhancement on gadolinium), we recommended incisional biopsy. Patients with smaller tumors that were adjacent to neurovascular structures generally underwent image-guided core biopsies and were excluded if a diagnosis was rendered. We included patients who underwent open biopsy after nondiagnostic needle biopsy.

All biopsies were performed by one of five musculo-skeletal oncology clinical fellows after a direct discussion with the attending surgeon in charge of the patient. The biopsy was performed according to 10 rules to (1) minimize contamination in the tissues surrounding the tumor; and (2) improve accuracy (Table 4). These rules were based on previous reports in the literature [4, 12, 13, 21] and on the experience of participating senior surgeons and allow one to plan soft tissue (Fig. 2) and bone (Fig. 3) biopsy tract resection in continuity with the main lesion. Contamination can be of two different types based on the consequences it has; it is either known or unknown. Known contamination can be appreciated by the surgeon performing the resection should there be one. For instance, a transverse incision contaminates normal tissues in such a way that a wider resection is necessary; the contaminated volume can be identified and dealt with at the time of the resection. On the opposite, a hematoma will contaminate tissues but it is not possible to precisely locate and resect the volume contaminated. Increasing the known contamination requires an upgrade of the surgical procedure (a sarcoma that required resection and primary closure may then require resection and flap coverage); an unknown contamination increases the risk of local recurrence. The procedure was conducted, after informed consent, under spinal or general anesthesia with strict aseptic conditions.

Fig. 1A–D Magnetic resonance T1 (A, C) and T2 (B, D) -weighted sequences of a local recurrence of a previously resected well-differentiated liposarcoma. A higher grade component is suspected

(C–D, red area) with the fatty component, making this tumor suspicious for a dedifferentiated liposarcoma.

Whenever possible, we used a tourniquet. The biopsy tract was orientated close to, and in line with, an extensile incision should limb salvage through wide resection become necessary. The lesion was approached through the involved compartment, away from major neurovascular structures, and without developing planes. We biopsied what we presumed was the most representative region of the lesion based on imaging studies and intraoperative findings. Care was taken not to crush the specimen. We routinely performed a frozen section to confirm a representative specimen was obtained. An adequate amount of tissue (a minimum of 10 cc) was sent to pathology for definitive diagnosis and the remainder sent for research

purposes and for tumor banking because our center provides numerous laboratories with unaltered, pretreatment tissue specimens to support a variety of molecular biology investigations in sarcoma. After the biopsy, the tourniquet was released before wound closure, meticulous hemostasis was obtained, and a tight multilayer closure was performed. Additionally, a resorbable sterile sponge-like material was used to pack the wound if it was believed it could help hemostasis. A drain was used whenever the surgeon performing the procedure believed a hematoma could form in the cavity created.

All diagnostic and treatment decisions were made at dedicated sarcoma multidisciplinary tumor boards

Table 3. Details of the patient with lesions of indeterminate malignant potential and those in which the biopsy was inadequate

Observation number	Age (years)	Sex	Tissue	Site	Pathology class	Pathology first biopsy	Second biopsy	Pathology second biopsy	Resect	Pathology resection	Outcome
1	41	F	Soft	Thigh	Ind	GCRT	No		Yes	GCRT	Success
2	23	F	Bone	Ulna	Ind	Cartilaginous	No		Yes	Chondrosarcoma	Success
3	58	M	Soft	Thigh	Ind	Smooth muscle	No		Yes	Leiomyosarcoma	Failure
4	15	M	Bone	Fibula	Ind	Cartilaginous	No		Yes	Chondrosarcoma	Success
5	76	M	Soft	Neck	Ind	Solitary fibrous	No		Yes	Solitary fibrous	Success
6	57	M	Bone	Pelvis	Ind	Atypical lymphocytic infiltrate	Yes	Lymphoma	No		Failure
7	73	M	Soft	Thigh	Ind	Necrotic debris	No		Yes	Foreign body reaction	Success
8	23	F	Soft	Thigh	Ind	Necrotic debris	Yes	PVNS	Yes	PVNS	Failure
9	50	F	Bone	Humerus	Ind	Atypical Lymphocytic infiltrate	Yes	Lymphoma	No		Failure
10	76	M	Soft	Thigh	Mal	Liposarcoma	No		Yes	Myxoid liposarcoma	Failure

F = female; M = male; Ind = indeterminate; Mal = malignant; GCRT = giant cell rich tumor; PVNS = pigmented villonodular synovitis.

Table 4. Technical guidelines for an open biopsy with their rationale and adverse outcomes if not followed

Rule	Technique	Category	Rationale	Problem if not followed
1	Use a tourniquet	Oncologic	Minimize unknown contamination	Increase risk of local recurrence
2	Biopsy tract close to, and in line with, an extensile incision should limb salvage through wide resection become necessary	Oncologic	Minimize known contamination	Upgrade surgical treatment to a more radical one
3	Use involved compartment	Oncologic	Minimize known contamination	Upgrade surgical treatment to a more radical one
4	Avoid developing planes	Oncologic	Minimize unknown contamination	Increase risk of local recurrence
5	Stay away from significant neurovascular structures	Oncologic	Minimize known contamination	Upgrade surgical treatment to a more radical one
6	Aim for a representative specimen	Diagnostic	Improves accuracy	No diagnosis/misdiagnosis
7	Confirm previous point on frozen section	Diagnostic	Improves accuracy	No diagnosis/misdiagnosis
8	Obtain a significant amount of tissue for pathology	Diagnostic	Improves accuracy	No diagnosis/misdiagnosis
9	Release tourniquet before closure and perform meticulous hemostasis	Oncologic	Minimize unknown contamination	Increase risk of local recurrence
10	If using a drain, bring it out close to and in line with the incision	Oncologic	Minimize known contamination	Upgrade surgical treatment to a more radical one

involving surgeons, medical and radiation oncologists, pathologists, and musculoskeletal radiologists. During these meetings, the pathology from the biopsy specimen was correlated with clinical and radiological data and, whenever possible, a diagnosis was provided. For the purpose of this study, the specimen was classified as a lesion of malignant potential, lesion with no evidence of malignancy (including benign tumors and nontumoral conditions), and lesion of indeterminate malignant potential. For lesions that had a second pathology examination,

either after another biopsy or a resection, both pathologies were compared. Each incisional biopsy was classified as adequate or inadequate. Procedures were considered inadequate if (1) another biopsy (incisional or core) was necessary as a diagnosis could not be rendered; (2) the biopsy tract was not optimally placed as judged by the surgeon in charge of the patient when the eventual resection took place; and (3) if the treatment provided based on the diagnosis from the biopsy was not optimal based on the eventual resection pathology (ie, the resection pathology

Fig. 2 Fibrosarcoma of the fibula invading the lateral and anterior compartments of the leg. The red arrow marks the planned biopsy tract. edl = extensor digitorum longus; f = fibula; fl = fibularis longus; t = tibia; ta = tibialis anterior; tp = tibialis posterior; ellipses mark neurovascular bundles.

Fig. 3 Pleomorphic undifferentiated sarcoma lying in the rectus femoris muscle. The red arrow marks the planned biopsy tract. a = adductors; f = femur; rf = rectus femoris; s = sartorius; vl = vastus lateralis; vi = vastus intermedius; ellipse marks superficial femoral vascular bundle.

differed substantially from the biopsy result). Otherwise, procedures were classified as adequate. The biopsy tract was judged to be optimal if, should a wide resection occur,

the resection of the tract and surrounding tissues (approximately 2 cm around) had the least impact on the expected function of the patient. For instance, a biopsy tract that was close to, and in line with, the incision necessary for resection could be excised en bloc with the specimen with minimal additional soft tissue. On the contrary, an incision that was away from, or of aberrant direction in relation to, the usual incision chosen for resection may require resection of more soft tissue than would have been necessary otherwise; this could adversely affect the patient ultimate function. Similarly, if the biopsy tract contaminated major neurovascular structures requiring their sacrifice during the definitive resection, the expected function was adversely affected and the biopsy tract was not considered optimal. The procedures were classified when the final surgical treatment had been administered to the patients and the definitive pathology result became available. Therefore, for benign lesions that did not undergo resection, the results were issued after the biopsy pathology; for malignant lesions that underwent resection, the results were obtained after resection of the specimen.

The CUSUM test [17] provides a simple and intuitive graphical representation of the ongoing successes and failures of the procedure (Appendix 1). The CUSUM score, St (St represents the test statistic that is used for hypothesis testing), increases with each failure and decreases with each success. The score is plotted on the y-axis against the sequential observations on the x-axis. When the score St crosses a limit h , the null hypothesis of adequate performance is rejected. In quality control wording it is said that an alarm is emitted indicating that substantial variation in the process (biopsy) may have occurred. As long as the score remains below this limit h , performance is considered adequate and monitoring continues. An important feature of the CUSUM test is that it has a limiting barrier at zero that prevents the score from drifting away from the limit h if one accumulates successes. Therefore, past successes do not compensate for present failures and the test remains responsive at all times. For instance, if fellows only accumulate successes, the score St will remain at 0 and the distance between St and h will remain the same. Then, if new fellows start their fellowship and yield several failures, they may reach the limit h quickly. Had the score been allowed to decrease without that limiting barrier at 0, the new fellows may have benefited from all the accumulated successes of previous fellows. In the present case, based on reports from the literature [12, 13, 22] and on consensus among surgeons in the department, the adequate performance level (null hypothesis, also known as the target) was set at a 5% failure rate and the inadequate level of performance was set at 15% failure rate. The limit h is determined, based on computer simulations (see Appendix 1), because there is no analytical solution to yield certain

level of risks (see subsequently) depending on the adequate and inadequate performance level chosen; it is akin to the critical values of traditional hypothesis testing such as 1.97 for the *t* test or 3.84 for the chi square test at 1 degree of freedom. In the present study a limit $h = 4.45$ was chosen to yield a false discovery rate (FDR; akin to the Type I error) of no more than 5% over 150 procedures (estimated number of procedures performed over 1 year). The corresponding true discovery rate (TDR; akin to the power) was 97%. Meetings were held every 2 or 3 months to present the results of the CUSUM test and discuss procedures that failed. All patients completed the study; there were no missing data.

Results

Five (4.3%) of the 116 incisional biopsy procedures were considered failures (Table 3: observations 3, 6, 8, 9, and 10). Three patients required a second repeat open biopsy (Table 3: observations 6, 8, and 9) and two were considered to receive suboptimal treatment based on the results of the biopsy (Table 3: observations 3 and 10). Two patients (one pelvis, one humerus; Table 3: observations 6 and 9) presented with a poorly marginated lytic lesion with an adjacent soft tissue mass and low signal intensity on both T1- and T2-weighted images. The pathology report from the biopsy revealed an atypical lymphocytic infiltrate. Although a lymphoma was strongly suspected, a second biopsy was necessary in each case to confirm the diagnosis. A 23-year-old woman (Table 3: observation 8) presented with a large cystic mass in her buttock and posterior thigh. Clinical examination and imaging studies suggested a sarcoma, an ancient schwannoma, or a myxoma. Final pathology from the biopsy reported necrotic debris with no evidence of malignancy. However, as a result of the possibility of inadequate sampling, it was believed that another biopsy was necessary to rule out a malignancy and, if possible, reach a diagnosis. A CT-guided biopsy targeting a more representative solid area of the lesion was performed and pathology confirmed villonodular synovitis. A 76-year-old patient (Table 3: observation 8) presented with a lipomatous mass in the thigh with some heterogeneity on CT imaging (a pacemaker prevented MRI studies) and the biopsy was in favor of a well-differentiated liposarcoma. He underwent a marginal excision but the definitive pathology identified a focus of myxoid liposarcoma extending to the surgical margins. He subsequently received postoperative radiation. A 58-year-old patient (Table 3: observation 3) presented with a circumferential mass of the thigh highly suggestive of a soft tissue sarcoma. An extensive incisional biopsy revealed a smooth muscle tumor of indeterminate malignant potential. This

Fig. 4 CUSUM test for monitoring the adequacy of open biopsy. The following parameters are considered: adequate performance: 5% failure (target rate; null hypothesis); inadequate performance: 15% (alternative hypothesis). The limit $h = 4.45$ gives a TDR of 97% and FDR of 4.7% over 150 procedures. No alarm was emitted during the monitoring period.

presented a dilemma, because an amputation was not reasonable without histologic evidence of sarcoma and because a limb salvage procedure was not possible should this tumor be malignant. Therefore, a marginal resection was performed. Final pathology revealed a Grade 3 leiomyosarcoma. Although an amputation was planned, the patient developed pulmonary metastases and died with no evidence of local recurrence.

No alarm was raised by the CUSUM test over the course of this study of 116 open biopsies, and the performance was deemed adequate over the monitoring period (Fig. 4). Failures can be seen as a jump on the graph; however, the limit h was never crossed.

Discussion

Biopsies are associated with major errors in diagnosis in 14% to 18% of cases and led to alterations in treatment in 18% to 19% [12, 13]. Control charts are useful to ensure that a process is operating at a predetermined level of performance [14]. In the present study we prospectively assessed the proportion of inadequate biopsies and used a control chart, the CUSUM test, to continuously ensure this rate was maintained at an adequate level.

This study has several limitations. First, only open biopsies were included because this is the preferred procedure to obtain diagnostic tissue at the study institution for most benign aggressive and malignant bone tumors and high-grade soft tissue sarcomas. Although we recognize core needle biopsies are perfectly reasonable to reach a diagnosis, incisional biopsy provides pretreatment material in sufficient quantity to support translational research initiatives and is sometimes necessary when a core biopsy is nondiagnostic [20]. Excluding patients who underwent needle/core biopsies restricts the applicability of the findings

to patients who undergo incisional biopsies. Second, patients who did not undergo surgical resection of their lesion after the biopsy such as those with asymptomatic fibromatosis, lipomas, etc, did not have a definitive examination of the entire specimen. Therefore, a wrong diagnosis cannot be definitively excluded and only careful followup will ensure whether the treatment based on the biopsy was appropriate. Third, the final diagnosis of patients undergoing neoadjuvant treatment before surgical resection such as those with osteosarcoma undergoing neoadjuvant chemotherapy only becomes available a few months after the biopsy has been performed, after definitive resection. Therefore, during that time, the results of monitoring, presented at regular intervals, considered that the treatment provided was adequate only to be potentially reclassified later.

Five (4%) of the 116 procedures in our study were considered failures. Of these, there was one (1%) major error in diagnosis with a patient undergoing marginal resection of a myxoid liposarcoma. Another patient had a minor error in diagnosis (smooth muscle tumor of indeterminate malignant potential, which ultimately was called a leiomyosarcoma), although this did not alter the final outcome of the patient. Our findings are consistent with those in the literature. Mankin et al. [12, 13] reported an incidence of 18% (in 1982) and 14% (in 1996) major errors in diagnosis among members of the Musculoskeletal Tumor Society. However, when considering only biopsies performed at their study institutions, the incidence of major diagnostic errors was only 9% (in 1982) and 12% (in 1996). The discrepancy in outcome between biopsies performed at recognized treatment centers and those performed at referring centers has been demonstrated by others [19]. The use of intraoperative frozen section reportedly improves accuracy [21]. With routine use of frozen section we found the specimen was either necrotic or nondiagnostic in eight (7%) patients, thus allowing more tissue to be collected while the patient remained under anesthetic. This is similar to findings in other studies. Dupuy et al. [7] reported an increase in accuracy for core and needle biopsies when frozen sections were concomitantly used from 88% to 94%. Similarly, Ashford et al. [1] in a study of 104 core needle biopsies with systematic intraoperative frozen section reported no major error in diagnosis. Technical errors such as inappropriate location of the biopsy tract, thereby contaminating critical neurovascular structures or compromising limb salvage surgery, is a well-known problem that can be associated with open biopsies. In the present series, there were no such errors. Mankin et al. [12, 13] reported 18% (in 1982) and 19% (in 1996) alterations in patient treatment because of a poorly performed biopsy; eventually, the outcome of the patient was adversely altered in 9% and 10%. More recently,

Pollock and Stalley [19] reported that 38% of biopsies performed at referral centers were associated with errors that substantially altered the definitive treatment.

The use of a process control chart such as the CUSUM provides surgeons and centers with real-time information about the quality of the process of care they are part of and sometimes dependent on. In our study, two of the three repeat biopsies (lymphoma) were more likely associated with difficulties in obtaining histologic confirmation rather than with failures of the surgical technique. This is in keeping with previous reports, which noted the difficulties in making the diagnosis of round cell tumors [7, 8, 21]. Without ongoing close scrutiny, the performance of a process may deviate unbeknownst to the care providers participating in that same process. The Bristol Royal Infirmary [24] provides one such example: at that center, the mortality rate after open heart surgery in patients younger than 1 year operated on between 1984 and 1995 was twice the national average. It took years for the inadequate performance to be detected. A retrospective review showed that the use of statistical process control would have detected the inadequate performance much earlier [23]. Although monitoring a yearly failure rate is an option, the CUSUM test presents numerous advantages [14]. First it provides an immediate and ongoing measure of the performance allowing for quick action to be taken. Second, it is not affected by the particular timing of failures. For instance, if two biopsies failed in December and two others in January, they will count over different periods if one monitors a yearly failure rate; this is not the issue when using the CUSUM test. Last, the CUSUM test is not affected by multiple testing (inflation of Type I error).

In conclusion, the proportion of inadequate open biopsies performed at our referral center was low. Using a statistical process control method to monitor the failures provides a continuous measure of the performance.

Appendix 1. The CUSUM test

The CUSUM test was developed in the 1950s by Page [17] to monitor the quality of manufactured products. Let $X_1, X_2, X_3, \dots$ be the sequence of observations under surveillance and assume no serial correlation and μ the mean of these observations. These could be the times to complete a surgical procedure, the local recurrence rate, or any other quantity one is interested in monitoring. The CUSUM tests after each observation X_t ($t > 0$) $H_0: \mu = \mu_0$, ie, performance is adequate, versus $H_1: \mu \neq \mu_0$, ie, performance is inadequate. The value μ_0 is referred to as the target. In its one-sided formulation, $H_0: \mu = \mu_0$ is tested against $H_1: \mu = \mu_1$ ($\mu_1 > \mu_0$), and test is based on the statistic S_t computed after each observation X_t :

$$S_t = \max(0, S_{t-1} + W_t), S_0 = 0 \quad (1)$$

W_t , the sample weight, depends on the observation X_t , μ_0 and μ_1 (see Equation 2). Medical data are often expressed in a success/failure dichotomy, and in that case, the mean of the process μ is equal to the probability of failure of the event monitored, which is noted as p . Optimal choices for W_t are proportional to the log-likelihood ratio.

The test statistic S_t is compared with a predefined limit h . If S_t equals or exceeds h , the null hypothesis is rejected and performance is deemed inadequate; the CUSUM test is said to emit an alarm indicating that the process is out of control. As long as S_t remains below h , the null hypothesis cannot be rejected and monitoring continues under the assumption that performance is adequate.

$$W_t = \log \left[\frac{p_1^{X_t} (1 - p_1)^{1-X_t}}{p_0^{X_t} (1 - p_0)^{1-X_t}} \right] \quad (2)$$

Performances of CUSUM tests are not expressed with Type I and Type II errors because, as a result of its design and given enough time, the CUSUM test will always reject the null hypothesis regardless the true performance of the process. Therefore, eventually, the probability to reject the null hypothesis of adequate performance when it is in fact true (Type I error) is 100% and the probability not to reject the null hypothesis when it is in fact false (Type II) is 0%. Therefore, performances of CUSUM tests are expressed with the true and false discovery rates (TDR and FDR, akin to the power and Type I error, respectively), namely the probability of an alarm to be emitted under the alternative and null hypotheses within a defined number of observations.

In the present case, based on reports from the literature [12, 13, 23] and on consensus among surgeons in the department, the adequate performance level was set at a 5% failure rate and the inadequate level of performance was set at a 15% failure rate. Based on computer simulations (10,000 simulations), a limit $h = 4.45$ was chosen to yield a FDR of no more than 5% over 150 procedures (estimated number of procedures performed over 1 year). The limit h is determined, based on computer simulations because there is no analytical solution, to yield certain level of risks (FDR and TDR) depending on the adequate and inadequate performance level chosen; it is akin to the critical values of traditional hypothesis testing such as 1.97 for the t test or 3.84 for the chi square test at 1 degree of freedom. The corresponding TDR was 97%. In practical terms it means that, over 150 procedures, there was a 97% chance that an alarm would be raised if the performance was 15% failure (true alarm) and there less than 5% chance that an alarm would be raised if the performance of the process was 5% (false alarm). Table A1 shows the results

of the simulations performed. Figure A1 shows the CUSUM test applied for the fictive series of 120 biopsies at theoretical, ie, is randomly generated, 5%, 10%, 15%, and 20% failure rates. It can be seen that the CUSUM test indicates inadequate performance at the 10th, 46th, and 115th procedures for the 20%, 15%, and 10%, respectively; no alarm is raised at the 5% performance level.

Appendix 2. Pathology of the biopsy of all lesions

Pathology of malignant tumors: 19, undifferentiated pleomorphic sarcomas; 9, osteosarcomas; 6, lymphomas; 5, Ewing sarcomas; 4, chondrosarcomas, and metastatic

Fig. A1 CUSUM test for four series of 120 biopsies at performance levels of 5%, 10%, 15%, and 20%. The CUSUM test parameters are $p_0 = 5\%$, $p_1 = 15\%$, and $h = 4.46$. It can be seen that the CUSUM test indicates inadequate performance at the 10th, 46th, and 115th procedures for the 20%, 15%, and 10%, respectively; no alarm is raised at the 5% performance level.

Table A1. Number of alarms obtained at various limits over 10,000 random simulations of series of 150 procedures at 5%, 7%, 12%, 15%, and 18% failure rates with a CUSUM test with the following parameters: adequate performance = 5%, inadequate performance = 15%*

h	Failure rate				
	5%	7%	12%	15%	18%
4.2	643	2395	8672	9784	9983
4.25	615	2429	8614	9808	9981
4.3	561	2354	8509	9760	9980
4.35	527	2235	8487	9771	9978
4.4	568	2156	8481	9761	9970
4.45	511	2144	8425	9743	9976
4.46	471	2041	8378	9752	9967
4.5	430	1994	8348	9725	9971
4.55	441	1946	8300	9710	9969
4.6	416	1919	8313	9691	9970

* The limit $h = 4.46$ was chosen for the study because it provided no more than 5% false discovery rate; it also provided 97% chance to detect a performance of 15% failure over 150 procedures.

adenocarcinomas; 3, low-grade osteosarcomas, myxoid liposarcomas, and soft tissue sarcoma not otherwise specified; 2, low-grade liposarcomas, myelomas, and pleomorphic liposarcomas; 1, chordoma, fibrosarcoma, leiomyosarcoma, liposarcoma, low-grade fibromyxoid sarcoma, malignant solitary fibrous tumor, melanoma, mpnst, myxofibrosarcoma, and pleomorphic rhabdomyosarcoma.

Pathology of lesions showing no evidence of malignancy: 5, fibromatosis; 4, giant cell tumors; 3, atypical lymphoid infiltrates and myxomas; 2, benign fibrous lesions, benign tumour of paraspinal nerves, haemangiomas, and necrotic debris; 1, enchondroma, eosinophilic granuloma, fibrous lesion, giant cell rich lesion, gout, granulocytic hyperplasia, infection, low grade liposarcoma, myositis, nodular fasciitis, nonossifying fibroma, osteochondroma, and spindle cell lipoma.

Pathology of lesions with indeterminate malignant potential: 2, cartilaginous tumors, tumors with necrotic debris, and atypical lymphoid reactions; 1 solitary fibrous tumor, one smooth muscle tumor, and one soft tissue giant cell-rich tumor.

References

1. Ashford RU, McCarthy SW, Scolyer RA, Bonar SF, Karim RZ, Stalley PD. Surgical biopsy with intra-operative frozen section. An accurate and cost-effective method for diagnosis of musculoskeletal sarcomas. *J Bone Joint Surg Br.* 2006;88:1207–1211.
2. Biau DJ, Meziane M, Bhumbra RS, Dumaine V, Babinet A, Anract P. Monitoring the quality of total hip replacement in a tertiary care department using a cumulative summation statistical method (CUSUM). *J Bone Joint Surg Br.* 2011;93:1183–1188.
3. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard R, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care.* 2007;16:203–207.
4. Bickels J, Jelinek JS, Shmookler BM, Neff RS, Malawer MM. Biopsy of musculoskeletal tumors. Current concepts. *Clin Orthop Relat Res.* 1999;368:212–219.
5. Collins GS, Jibawi A, McCulloch P. Control chart methods for monitoring surgical performance: a case study from gastro-oesophageal surgery. *Eur J Surg Oncol.* 2011;37:473–480.
6. de Leval MR, Francois K, Bull C, Brawn W, Spiegelhalter D. Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations. *J Thorac Cardiovasc Surg.* 1994;107:914–923.
7. Dupuy DE, Rosenberg AE, Punyaratabandhu T, Tan MH, Mankin HJ. Accuracy of CT-guided needle biopsy of musculoskeletal neoplasms. *AJR Am J Roentgenol.* 1998;171:759–762.
8. Ehrlich PF, Friedman DL, Schwartz CL. Monitoring diagnostic accuracy and complications. A report from the Children's Oncology Group Hodgkin lymphoma study. *J Pediatr Surg.* 2007;42:788–791.
9. Institute of Medicine. *To Err Is Human: Building a Safer Health System.* Washington, DC, USA: National Academy Press; 2000.
10. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC, USA: National Academy Press; 2001.
11. Macpherson GJ, Brenkel IJ, Smith R, Howie CR. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. *J Bone Joint Surg Am.* 2011;93:81–88.
12. Mankin HJ, Lange TA, Spanier SS. The hazards of biopsy in patients with malignant primary bone and soft-tissue tumors. *J Bone Joint Surg Am.* 1982;64:1121–1127.
13. Mankin HJ, Mankin CJ, Simon MA. The hazards of the biopsy, revisited. Members of the Musculoskeletal Tumor Society. *J Bone Joint Surg Am.* 1996;78:656–663.
14. Montgomery DC. *Introduction to Statistical Quality Control.* 5th ed. New York, NY, USA: John Wiley and Sons; 2005.
15. Morton AP, Whitby M, McLaws ML, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *Qual Clin Pract.* 2001;21:112–117.
16. National Cancer Institute. A snapshot of sarcoma. Available at: <http://nci.nih.gov/aboutnci/servingpeople/snapshots/sarcoma.pdf>. Accessed December 22, 2011.
17. Page ES. Continuous inspection schemes. *Biometrika.* 1954;41:100–115.
18. Papanna R, Biau DJ, Mann LK, Johnson A, Moise KJ. Use of the Learning Curve-Cumulative Summation test for quantitative and individualized assessment of competency of a surgical procedure in obstetrics and gynecology: fetoscopic laser ablation as a model. *Am J Obstet Gynecol.* 2011;204:1–9.
19. Pollock RC, Stalley PD. Biopsy of musculoskeletal tumours—beware. *ANZ J Surg.* 2004;74:516–519.
20. Rimondi E, Rossi G, Bartalena T, Ciminari R, Alberghini M, Ruggieri P, Errani C, Angelini A, Calabro T, Abati CN, Balladelli A, Tranfaglia C, Mavrogenis AF, Vanel D, Mercuri F. Percutaneous CT-guided biopsy of the musculoskeletal system: results of 2027 cases. *Eur J Radiol.* 2011;77:34–42.
21. Rougraff BT, Aboulafia A, Biermann JS, Healey J. Biopsy of soft tissue masses: evidence-based medicine for the Musculoskeletal Tumor Society. *Clin Orthop Relat Res.* 2009;467:2783–2791.
22. Skrzynski MC, Biermann JS, Montag A, Simon MA. Diagnostic accuracy and charge-savings of outpatient core needle biopsy compared with open biopsy of musculoskeletal tumors. *J Bone Joint Surg Am.* 1996;78:644–649.
23. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care.* 2003;15:7–13.
24. The inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary. The Report of the Public Inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995. 2001. Available at: www.bristol-inquiry.org.uk/; Accessed May 21, 2012.