Journal of Veterinary Internal Medicine, 2026, 40(3), aalag121
https://doi.org/10.1093/jvimsj/aalag121
Original Research | Small Animal Internal Medicine Hematology

Multisite sampling improves consensus cytologic interpretation in bone marrow evaluation of dogs

Kristi M. Smiley[1], Michael F. Rosser[1], Amy N. Schnelle[1], Sara L. Connolly[1], Rose Raskin[2], Nicolas Lopez-Villalobos[3], Arnon Gal[1,*]

[1]Department of Veterinary Clinical Medicine, University of Illinois Urbana-Champaign, Urbana, IL 61802, United States
[2]Department of Comparative Pathobiology, Purdue University, West Lafayette, IN 47907, United States
[3]School of Agriculture and Environment, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand
*Corresponding author: Arnon Gal, Department of Veterinary Clinical Medicine, University of Illinois Urbana-Champaign, 1008 W Hazelwood Drive, Urbana, IL 61802, United States (agal2@illinois.edu).

Abstract

Background: **Bone marrow (BM) evaluation is essential for diagnosing hematologic abnormalities in dogs. Standard practice often involves single-site sampling, despite some evidence suggesting that focal BM lesions may occur. We assessed the probability of obtaining an unrepresentative cytologic interpretation when BM cytology interpretation depends on a single site vs multiple sites.**

Hypothesis/Objectives: **Single-site BM sampling increases the probability of unrepresentative cytologic interpretation in dogs.**

Animals: **Sixteen client-owned dogs with hematologic disease requiring BM sampling were prospectively enrolled.**

Methods: **Randomized, masked clinical trial. We collected 4 BM aspirates per dog: right and left proximal humeri and iliac crests. Interpretative concordance was assessed by comparing individual site interpretations with a composite multisite interpretation. The primary outcome was the probability of a truth-consistent interpretation based on the number of sites sampled ($k = 1$-4) under 3 evaluation rules, accounting for inter-clinical pathologist agreement.**

Results: **Single-site sampling yielded a truth-consistent interpretation in only 50.0%-81.7% of cases, depending on diagnostic stringency. Evaluating 2 sites significantly improved the probability of a truth-consistent interpretation by 17.2%-23.3% ($P < .002$). Additional sites (2 or 4) offered only marginal further benefit. Inter-clinical pathologist agreement was moderate (kappa = 0.574). No single anatomical site provided superior diagnostic yield for overall interpretation.**

Conclusions and clinical importance: **A single BM sampling site carries a higher probability of unrepresentative cytologic interpretation in dogs. Two-site sampling significantly improves recovery of a truth-consistent interpretation, providing a more reliable assessment for hematologic disease in dogs. This approach accounts for potential focal disease distribution, improving diagnostic confidence.**

Keywords cytology, sampling strategy, truth-consistent interpretation, clinical pathology, canine

Abbreviations BM bone marrow; Bx_Site, biopsy site; CBC, complete blood count; GLIMMIX, generalized linear mixed model (SAS procedure); ICC, intraclass correlation coefficient; ILINK, inverse link transformation; IQR, interquartile range; LH, left humerus; LI, left iliac crest; MER, myeloid-to-erythroid ratio; MDS, myelodysplastic syndrome; OR, odds ratio; RI, right iliac crest; RH, right humerus

Introduction

Bone marrow (BM) evaluation is a cornerstone diagnostic tool in small animal internal medicine, particularly when peripheral blood findings cannot fully explain hematologic abnormalities.[1] In dogs, BM aspiration and core biopsy samples are indicated for the investigation of unexplained cytopenias, suspected hematopoietic neoplasia, hematologic myelopathies, or the staging of systemic malignancies.[1,2] Sampling is most frequently performed from the proximal humerus or dorsal iliac crest, with the femur and sternum being less commonly used because of patient size, accessibility, and operator familiarity.[3] Site selection is influenced by patient conformation, age, and underlying disease: the proximal humerus offers a large medullary cavity and minimal overlying soft tissue in larger or obese dogs, whereas the iliac crest is easily accessible in leaner dogs.[3] In very small dogs, the proximal femur may be preferred,[3] whereas sternal aspiration (historically underutilized because of perceived safety concerns) has

Received: September 9, 2025. **Revised:** May 13, 2026. **Accepted:** May 20, 2026

been shown to be safe and diagnostically comparable to other sites.[4] Bone marrow composition varies with age and anatomical location. In older dogs, appendicular marrow tends to be more fatty (inactive), making certain sites less representative of active hematopoiesis.[1] Given these variations, site choice can influence sample quality, and potentially, diagnostic yield.

In veterinary medicine, standard practice typically relies on a single-site BM sample, assuming uniform distribution of hematopoietic pathology. However, focal or multifocal distribution of BM lesions is well recognized in humans, where bilateral or multisite sampling is recommended in conditions such as lymphoma, metastatic disease, or myelofibrosis to improve detection rates.[5] Veterinary evidence, although limited, supports the possibility of clinically relevant site-to-site variation.[6] In a study evaluating lymphoma and high-grade mast cell tumor in dogs, multisite aspirates yielded discordant results for marrow involvement in some mast cell tumor cases, suggesting that a single-site approach would have missed the disease.[4] Conversely, in apparently healthy dogs, humeral and iliac crest aspirates showed high diagnostic concordance, with only an approximately 5% discordance rate and minor differences in myeloid-to-erythroid ratios.[7] That study, however, did not include diseased dogs and recommended further investigation in clinical populations.[7]

Little evidence quantifies the probability of obtaining an unrepresentative cytologic interpretation when relying on a single BM sampling site in dogs with hematologic disease. Given the potential for patchy marrow involvement, the diagnostic implications of site selection remain unclear in dogs with hematologic diseases. Thus, we hypothesized that relying on a single, standard BM sampling site would increase the probability of an unrepresentative cytologic interpretation compared with the evaluation of multiple sites. Our objective was to determine the probability of obtaining a truth-consistent interpretation when BM cytology is based on one randomly selected site compared with the composite interpretation derived from multiple sites in the same dog.

Materials and methods

Study design and population

This randomized, masked clinical trial (IACUC#19119) was conducted at the University of Illinois Veterinary Teaching Hospital. From July 2020 to April 2023, 16 client-owned dogs (>7 kg) with hematologic disease requiring BM sampling as part of their diagnostic evaluation (eg, bi- or pancytopenia; suspected infectious, neoplastic, or immune-mediated precursor-directed BM disease) were prospectively enrolled with written owner consent.

Bone marrow sampling

A board-certified internist (A.G.) performed all procedures using the ARROW OnControl Powered Driver with a 15-gauge needle (Teleflex, Morrisville, NC). Four aspirates per dog were obtained from the right (RH) and left (LH) proximal humeri and iliac crests (RI, LI) in randomized complete block order (Table 1), with a concurrent CBC. Landmarks, needle orientation, citrate anticoagulation, aspirate volume (≤0.2 mL), and slide preparation are described in the Appendix. Two to 5 spicule-containing smears per site were stained with Wright–Giemsa. Samples received randomized accession numbers; dog identity and site were masked from 2 board-certified clinical pathologists (A.S., M.R.).

Table 1 Sequence of sampling (randomized complete block design).

Dog no.	First	Second	Third	Fourth
1	LI	LH	RH	RI
2	LI	RH	LH	RI
3	LH	RH	LI	RI
4	LH	LI	RI	RH
5	RH	LI	RI	LH
6	LH	RI	LI	RH
7	LH	RI	RH	LI
8	LI	RI	LH	RH
9	LI	RH	RI	LH
10	LI	LH	RI	RH
11	RH	RI	LH	LI
12	LH	RH	RI	LI
13	RI	LH	LI	RH
14	RH	RI	LI	LH
15	LI	RI	RH	LH
16	LH	LI	RH	RI

Abbreviations: LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest.

Table 2a Cytologic interpretation code list.

Code no.	Explanation
1	Hyperplasia of 1 or more than 1 lineage
2	Myelodysplastic syndrome
3	Leukemia/round cell neoplasia
4	Hypoplasia of 1 or more than 1 lineage
5	BM inflammation/infection
6	BM toxicity
7	Metastatic neoplasia
8	Nondiagnostic sample
9	Other

Abbreviation: BM = bone marrow.

Outcomes

The primary outcome was site-specific clinical interpretation, with up to 3 diagnostic codes per site (Table 2a) integrating cytomorphology, CBC, and signalment. Secondary outcomes were spicule cellularity, spicules per slide, myeloid-to-erythroid ratio, blasts (%), megakaryocytes per spicule, ordinal sample quality, lymphocytosis > 5%, and plasmacytosis > 2%.

Statistical analysis

Analyses were conducted using SAS 9.4. For each dog, a truth set comprised the cytology code(s) with the highest frequency across sites and pathologists (nondiagnostic reads excluded). Sites were classified as truth-consistent under 3 rules: ANY, BOTH-any, and BOTH-same. For k = 1-4 sites sampled without replacement, the

Table 2b Cytologic code distribution and nondiagnostic summary.

Category	Code	*n*/128	%
BM hyperplasia	1	52	40.6
Nondiagnostic sample	8	30	23.4
BM hypoplasia	4	23	18.0
Leukemia/round cell neoplasia	3	16	12.5
Other	9	2	1.6
BM inflammation/infection	5	1	0.8
Myelodysplastic syndrome	2	0	0
BM toxicity	6	0	0
Metastatic neoplasia	7	0	0
Missing data	—	4	3.1

Abbreviation: BM = bone marrow Nondiagnostic distribution. Dogs affected: 11/16; median within-dog nondiagnostic proportion 25% (range 0%-100%); by site: RI 11/30, RH 8/30, LI 6/30, LH 5/30; by pathologist: Path 1 = 18/30, Path 2 = 12/30.

probability of ≥ 1 truth-consistent site was computed per dog and averaged, with nonparametric bootstrap 95% confidence intervals (CIs; 2000 replicates). The minimum clinically important difference was prespecified as $\geq 10\%$ absolute increase (number needed to treat ≤ 10), consistent with hematopathology benchmarks used in humans.[5] Adjacent within-dog contrasts (2-1, 3-2, and 4-3) were tested using 2-sided Wilcoxon signed-rank tests.

Per-read truth consistency, continuous secondary outcomes (log-transformed where skewed), binary outcomes, and ordinal quality score were modeled using mixed-effects regression (MIXED/GLIMMIX; Laplace approximation), with fixed effects of site, pathologist, and their interaction, and a random intercept for dog. Type III tests, marginal means with 95% CIs (ILINK), and Tukey–Kramer-adjusted pairwise site comparisons are reported. Inter- and intra-observer agreements were summarized by percentage agreement (Clopper–Pearson CIs), Cohen's κ (bootstrap 95% CIs), and Gwet's agreement coefficient 1 (AC1) as a κ-paradox sensitivity analysis. Secondary analyses are exploratory.

Results

Sixteen dogs met the inclusion criteria and were enrolled (Table S1). Across 128 possible clinical pathologists' ratings (16 dogs $\times$ 4 sites $\times$ 2 clinical pathologists), BM hyperplasia predominated (52/128, 40.6%), followed by nondiagnostic samples (30/128, 23.4%), hypoplasia (23/128, 18.0%), and leukemia or round-cell neoplasia (16/128, 12.5%); BM inflammation or infection and other were rare (1/128 [0.8%] and 2/128 [1.6%]), and 4/128 (3.1%) were missing data (ie, dry taps that yielded no sample; Table 2b; Table S2). Nondiagnostic results occurred in 11 of 16 dogs. The nondiagnostic code was assigned in 1 dog (ID 6) to all 4 sites by both clinical pathologists (8/8, 100%); dog 14 had 5/8 (62.5%), dog 16 had 3/8 (37.5%); 6 dogs (2, 8, 9, 10, 12, and 13) had 2/8 (25%), and 2 dogs (1, 5) had 1/8 (12.5%). By site, nondiagnostic reads were most frequent at RI (11/30), followed by RH (8/30), LI (6/30), and LH (5/30). Nondiagnostic samples were excluded from further analyses.

Across dogs, the probability of obtaining at least 1 truth-consistent interpretation increased as more sampling sites were sampled (Table 3; Figure 1). With 1 site, success was high under the most permissive rule and increased markedly when 2 sites were sampled. Under stricter rules (requiring both clinical pathologists to align with the truth set), 3 sites generally were needed to approach the upper 90% range. Gains were minimal when increasing from 3 to 4 sites.

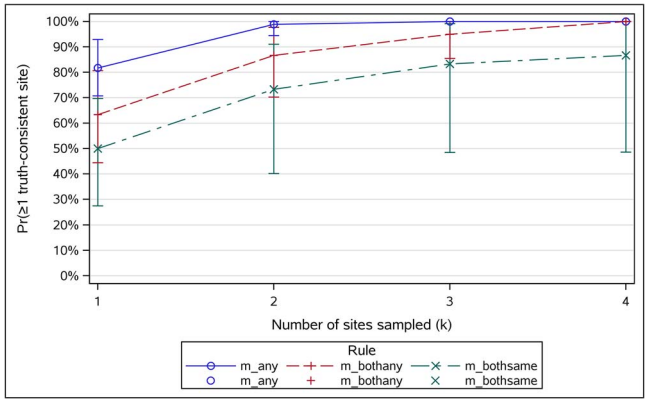

Figure 1 Probability of capturing ≥ 1 truth-consistent site vs number of sites sampled (*k*), by rule. Lines show means across dogs; bands show 95% bootstrap CIs. m_any = $\geq$1 clinical pathologist's read in the truth set; m_bothany = both in the truth set (not necessarily equal); m_bothsame = both agree and the agreed code is in the truth set. With m_any, 2 sites $\approx$ 99% (94%-100%); with m_bothany, 3 sites $\approx$ 95% (86%-100%); with m_bothsame, 4 sites average $\approx$ 87% (49%-100%).

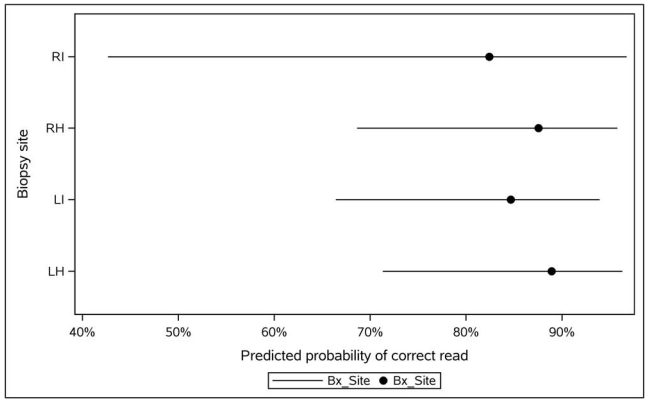

Figure 2 Model-based site effects on per-reading truth consistency. GLIMMIX least-squares means (points) with 95% CIs (whiskers) for LH, LI, RH, RI on the probability scale. No site effect ($P = .904$) and no clinical pathologist effect ($P = .988$). Abbreviations: LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest.

A mixed effects model did not indicate that per-reading truth-consistency differed by anatomical site or by clinical pathologist (Table 4; Figure 2). Model-based, site-specific predicted probabilities were similar, and Tukey-adjusted pairwise comparisons were not significant.

Within-dog paired comparisons confirmed these patterns (Table 5). The most significant improvement was from 1 to 2 sites across all rules. Under stricter rules, a smaller but still meaningful gain was observed from 2 to 3 sites, but adding a fourth site provided little additional benefit.

Between the 2 clinical pathologists, agreement on site-level codes was in the moderate to substantial range overall (Table 6;

Table 3 Probability of capturing ≥ 1 truth-consistent site by number of sites sampled (k) under three evaluation rules; values are means across dogs with 95% bootstrap CIs.

Rule	$k = 1$	$k = 2$	$k = 3$	$k = 4$
m_any (≥1 clinical pathologist in truth set)	81.7% (70.7-93.0)	98.9% (94.4-100.0)	100.0% (100-100)	100.0% (100-100)
m_bothany (both in truth set)	63.3% (44.4-80.7)	86.7% (70.2-97.7)	95.0% (85.5-100.0)	100.0% (100-100)
m_bothsame (both agree + in truth set)	50.0% (27.4-69.6)	73.3% (40.1-91.1)	83.3% (48.4-99.2)	86.7% (48.6-100.0)

Mean probability (across dogs) of obtaining ≥ 1 site consistent with the dog's truth set when sampling k sites without replacement from 4 sites; CIs are 2.5th-97.5th bootstrap percentiles.

Table 4 Site-specific predicted probability that a single read is truth consistent (GLIMMIX least-squares means on the probability scale).

Bx site	Mean	95% CI
LH	0.889	0.713-0.963
LI	0.847	0.664-0.939
RH	0.875	0.686-0.958
RI	0.824	0.426-0.967

Abbreviations: LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest. Type III test for site: $P = .904$; clinical pathologist: $P = .988$.

Table 5 Paired differences in $P(k)$ between adjacent k's, by rule (mean ΔP with bootstrap 95% CI across dogs; 2-sided P-values).

Rule	Contrast	Mean ΔP (95% CI)	Wilcoxon P	Sign P	t-test P
m_any	2-1	17.2% (6.8-25.0)	.002	.002	.001
m_any	3-2	1.1% (0.0-5.6)	1	1	.33
m_any	4-3	0.0% (0.0-0.0)	—	—	—
m_bothany	2-1	23.3% (13.6-28.8)	.001	.001	<.001
m_bothany	3-2	8.3% (1.6-16.9)	.031	.031	.01
m_bothany	4-3	5.0% (0.0-13.7)	.25	.25	.08
m_bothsame	2-1	23.3% (12.7-30.5)	.001	.001	<.001
m_bothsame	3-2	10.0% (3.4-17.5)	.008	.008	.002
m_bothsame	4-3	3.3% (0.0-12.0)	.5	.5	.16

$\Delta P = P(k\_2) - P(k\_1)$ computed within dog; means and 95% CIs are from the paired bootstrap across dogs (2000 reps). Wilcoxon = signed-rank test of median $\Delta P = 0$; "Sign P" = sign test; t-test shown for reference. "—" indicates all within-dog differences were zero (test not applicable).

Table 6 Inter-clinical pathologist agreement (overall).

Metric	Estimate	95% CI
Simple κ	0.574	0.340-0.808
Gwet's AC1	0.6314	0.460-0.803
Percent agreement	70.5%	54.8%-83.2%

Abbreviation: AC1 = agreement coefficient 1. Exact Clopper–Pearson interval for percent agreement; asymptotic 95% CIs for κ; Gwet's agreement coefficient 1 (95% CI).

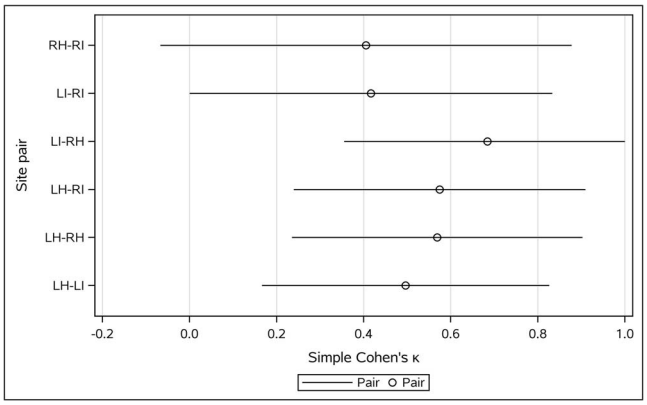

Figure 3 Inter-clinical pathologist agreement by site pair. Simple κ with 95% CIs for each site pair pooled across clinical pathologists, with N overlaid. Highlights that LI–RH shows the highest agreement. Abbreviations: LI = left iliac crest; RH = right humerus.

Figure 3), with by-site percentage agreement showing similar patterns (Table 7). A symmetry test indicated no directional bias in the distribution of codes between raters.

Within-clinical pathologist (intra-observer) agreement across site pairs was also generally moderate to substantial, although precision varied where pair counts were small (Table 8). The LI–RH pair tended to show the highest within-rater consistency, whereas LI–RI and RH–RI were typically lower. Pooling across clinical pathologists yielded the same ranking (Table 9).

Across models, sampling site did not have a meaningful effect on the secondary outcomes (all $P \geq .15$). In contrast, the clinical pathologists differed for several quantitative measures: total

Table 7 Percent agreement by sampling site (exact 95% CIs).

Bx site	Percent agree	95% CI
LH	76.9%	46.2%-95.0%
LI	66.7%	34.9%-90.1%
RH	66.7%	34.9%-90.1%
RI	71.4%	29.0%-96.3%

Site-wise exact binomial CIs for the proportion of matching codes between clinical pathologists. Abbreviations: LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest.

spicule counts, mean megakaryocytes per spicule, and the proportion of blasts (all $P \leq .004$), whereas no clear clinical pathologist effect was seen for myeloid-to-erythroid ratio (MER) or overall marrow cell percentage. Dog-to-dog differences were substantial (intraclass correlation coefficients, approximately 51%-79%). Interactions between site and clinical pathologist generally were not significant; 1 model (MER) showed a borderline site × clinical pathologist interaction ($P = .06$), and thus site comparisons for MER should be interpreted cautiously (see Table 10 for details).

In the mixed-effects logistic models, sampling site did not meaningfully affect the odds of diagnosing lymphocytosis or plasmacytosis, whereas the 2 clinical pathologists differed in their overall calling rates for both outcomes; no site-by-clinical pathologist interaction was observed (Table 11). On the probability scale, site estimates had wide, overlapping CIs, consistent with the nonsignificant site tests (Table 11). Readings from the same dog were strongly correlated, especially for lymphocytosis (intraclass correlation coefficient [ICC] approximately 0.92), indicating substantial between-dog variability (Table 11). Overall, differences appear to be driven more by clinical pathologist and dog-level factors than by sample site. Also, no evidence indicated that sampling site, clinical pathologist, or their interaction affected the quality score (biopsy site $P = .25$; clinical pathologist $P = .75$; site × clinical pathologist $P = .36$).

Discussion

Our study was designed to evaluate how effectively single-site vs multisite BM sampling recovers truth-consistent cytologic interpretation. The reference standard used was an internal consensus-based composite derived from the most frequently recurring diagnostic codes across all available reads and sites. This framework was selected to capture the interpretation most consistently supported by the total marrow information obtained from each dog and to provide a patient-level comparator against which different sampling strategies could be evaluated. The core hypothesis was that relying on a single BM sample increases the probability of an unrepresentative cytologic interpretation compared with a composite interpretation derived from multiple sites. To investigate this hypothesis, 4 BM aspirates were collected from each dog: RH, LH, RI, and LI.

Our results indicated a substantial decrease in interpretive robustness when only a single, randomly selected BM site was evaluated. Our approach provides a structured estimate of how often within-dog different site-sampling strategies recover the consensus cytologic interpretation supported by the total marrow information obtained. Under the strictest evaluation criteria, the probability that a single site reflected the final composite interpretation was 50.0%, compared with 81.7% under the most lenient criteria. Increasing the number of sampled sites improved this probability substantially (by a mean of 17.2%-23.3%; $P < .002$), with the addition of a second site providing the largest gain. These findings show that broader site sampling more consistently recovers the dominant marrow interpretation than reliance on a single site alone.

The observed discordance likely arose from 2 distinct sources: biologic and sampling-related heterogeneity among marrow specimens collected from different sites within a dog and reader-dependent variability when the same cytologic specimen was interpreted by different clinical pathologists. First, regarding site variation, the data indicated that diagnostic agreement among different sampling sites within the same dog (as assessed by individual clinical pathologists) was imperfect, ranging from 63.6% to 83.3% (Table 9). This observation supports the premise that hematopoietic pathology can be focal or patchy, meaning a single sample may not be representative of the overall disease state. Second, regarding interpretation, the overall agreement between the 2 participating clinical pathologists was moderate, with a simple κ statistic of 0.574 and 70.5% agreement. Gwet's AC1, a prevalence-adjusted agreement metric, yielded similar or slightly higher estimates than Cohen's κ across all site pairs and clinical pathologists, indicating that the observed agreement was robust to diagnostic skewness and not artificially decreased by unequal code distribution. This level of agreement was consistent across all 4 anatomical sites, suggesting no single location was inherently more difficult to interpret than another. Although interobserver variability is an intrinsic challenge of morphologic assessment, the 2 board-certified clinical pathologists in our study had similar training backgrounds and extensive experience with routine BM interpretation, minimizing the likelihood that the observed differences were primarily attributable to major differences in expertise.

Despite the observed differences in interpretations among sites for individual dogs, we found no evidence that any single anatomical location was systematically superior. Statistical analysis showed that the specific sampling site (humerus vs iliac crest) was not a significant predictor of truth consistency ($P = .90$) or of key cytologic features such as MER, cellularity, lymphocytosis, or plasmacytosis. In other words, the marginal variation observed among sites (<5%) was minimal in comparison with the substantial 17%-23% improvement achieved by increasing the number of sampling sites. These findings indicate that which site is sampled matters far less than how many sites are sampled. These findings also align with a recent necropsy study,[8] which reported that BM composition was broadly comparable across humeral, iliac, sternal, and costal sites in dogs with spontaneous disease, with most variation attributable to the individual animal rather than the site. However, although gross BM compositional metrics in both studies suggest within-dog similarities across BM sampling sites, our study emphasizes that the ability to detect cytologic interpretive discordance remains a substantial risk when relying on a single site, reinforcing the value of multisite sampling regardless of the chosen sampling site location. In contrast, the individual clinical pathologist reviewing the sample was a significant source of variation for several quantitative assessments, including megakaryocyte counts, blast percentages, and the presence of lymphocytosis or plasmacytosis. This observation reinforces that

Table 8 Intra-observer agreement by site pair and clinical pathologist.

Path	SiteA	SiteB	*N*	Cohen's κ (95% CI)	Cohen's κ (bootstrap percentile 95% CI)	Mean $\pm$ SD bootstrapped κ	Gwet's AC1 (95% CI)	% Agreement (95% CI)
1	LH	LI	11	0.34 (−0.14-0.82)	0.34 (−0.14-0.88)	0.31 ± 0.30	0.58 (0.24-0.92)	63.6% (30.8%-89.1%)
1	LH	RH	10	0.50 (0.02-0.98)	0.50 (−0.23-0.94)	0.45 ± 0.33	0.65 (0.31-0.98)	70.0% (34.8%-93.3%)
1	LH	RI	7	0.53 (0.03-1.00)	0.53 (0.00-1.00)	0.49 ± 0.32	0.67 (0.28-1.00)	71.4% (29.0%-96.3%)
1	LI	RH	9	0.43 (−0.08-0.93)	0.43 (−0.12-0.92)	0.38 ± 0.31	0.61 (0.25-0.98)	66.7% (29.9%-92.5%)
1	LI	RI	6	0.14 (−0.38-0.67)	0.14 (−0.26-0.84)	0.14 ± 0.30	0.42 (−0.05-0.90)	50.0% (11.8%-88.2%)
1	RH	RI	4	0.20 (−0.61-1.00)	0.20 (−0.36-1.00)	0.19 ± 0.41	0.41 (−0.17-0.99)	50.0% (6.8%-93.2%)
2	LH	LI	13	0.49 (0.10-0.87)	0.49 (−0.03-0.93)	0.46 ± 0.26	0.64 (0.34-0.94)	69.2% (38.6%-90.9%)
2	LH	RH	11	0.54 (0.09-0.98)	0.54 (−0.09-1.00)	0.49 ± 0.31	0.68 (0.37-0.99)	72.7% (39.0%-94.0%)
2	LH	RI	9	0.49 (0.04-0.94)	0.49 (−0.10-0.92)	0.46 ± 0.30	0.60 (0.23-0.97)	66.7% (29.9%-92.5%)
2	LI	RH	11	0.67 (0.28-1.00)	0.67 (0.04-1.00)	0.64 ± 0.27	0.79 (0.52-1.00)	81.8% (48.2%-97.7%)
2	LI	RI	10	0.41 (−0.02-0.85)	0.41 (−0.25-0.91)	0.36 ± 0.30	0.52 (0.15-0.89)	60.0% (26.2%-87.8%)
2	RH	RI	7	0.53 (0.03-1.00)	0.53 (0.00-1.00)	0.48 ± 0.33	0.67 (0.28-1.00)	71.4% (29.0%-96.3%)

Abbreviations: AC1 = agreement coefficient 1; LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest. Cohen's κ (95% CI), percent agreement with exact (Clopper–Pearson 95% CI), Gwet's AC1 (95% CI), Cohen's κ (bootstrap percentile 95% CI), mean $\pm$ SD bootstrapped κ; Path = clinical pathologist; *N* = number of paired site readings for that clinical pathologist.

Table 9 Pooled (across clinical pathologists) agreement by site pair.

SiteA	SiteB	*N*	Cohen's κ (95% CI)	Cohen's κ (bootstrap percentile 95% CI)	Mean $\pm$ SD bootstrapped κ	Gwet's AC1 (95% CI)	% Agreement (95% CI)
LH	LI	24	0.42 (0.12-0.73)	0.42 (−0.02-0.81)	0.40 ± 0.22	0.61 (0.39-0.84)	66.7% (44.7%-84.4%)
LH	RH	21	0.52 (0.19-0.85)	0.52 (−0.03-0.89)	0.49 ± 0.24	0.66 (0.44-0.89)	71.4% (47.8%-88.7%)
LH	RI	16	0.51 (0.16-0.85)	0.51 (0.03-0.91)	0.48 ± 0.24	0.63 (0.36-0.90)	68.8% (41.3%-89.0%)
LI	RH	20	0.55 (0.22-0.89)	0.55 (0.01-0.91)	0.51 ± 0.25	0.71 (0.49-0.93)	75.0% (50.9%-91.3%)
LI	RI	16	0.33 (−0.04-0.69)	0.33 (−0.19-0.80)	0.30 ± 0.26	0.48 (0.19-0.77)	56.3% (29.9%-80.2%)
RH	RI	11	0.41 (−0.07-0.88)	0.41 (−0.27-0.89)	0.37 ± 0.31	0.57 (0.24-0.91)	63.6% (30.8%-89.1%)

Abbreviations: AC1 = agreement coefficient 1; LH = left humerus; LI = left iliac crest; RH = right humerus; RI = right iliac crest. Cohen's κ (95% CI), Cohen's κ (bootstrap percentile 95% CI), mean $\pm$ SD bootstrapped κ, Gwet's AC1 (95% CI), percent agreement with exact (Clopper–Pearson 95% CI); *N* = number of paired site readings pooled across both pathologists.

Table 10 Mixed-effects models: main effects and clustering by dog.

Outcome (dependent variable)	*n*	Sampling site (type III *P*)	Clinical pathologist (type III *P*)	Site × path (type III *P*)	ICC
logMER	89	.82	.75	.06	0.70
log(Number of spicules)	94	.45	.003	.54	0.55
log(Mean megakaryocytes per spicule)	87	.16	<.001	.24	0.51
logit(% bone-marrow cells)	94	.68	.12	.20	0.59
logit(% blasts)	92	.17	<.001	.09	0.79

Abbreviation: ICC = intraclass correlation coefficient. Type III *P*-values are from PROC MIXED. ICC = Var(Dog_ID)/(Var(Dog_ID) + Residual); higher ICC means more between-dog clustering. (ICCs above are: 70.4%, 56.0%, 51.2%, 59.1%, and 79.3%, respectively).

both the biological heterogeneity of the marrow and the subjective nature of cytologic interpretation contribute to diagnostic uncertainty.

Considering the demonstrated risk of single-site interpretive error, our data support a general recommendation for multi-site sampling. In our study population, the largest and most clinically meaningful improvement was achieved by adding a second site, which provided a 17-23 percentage gain in the probability of a truth-consistent interpretation while limiting additional procedural burden. Beyond 2 sites, the incremental benefit was smaller. Accordingly, in cases with a high index of suspicion for disease processes known or suspected to have patchy BM involvement, such as certain malignancies (eg, mast cell tumor,[4] histiocytic sarcoma[9,10]) or focal infectious processes,[11–13] sampling a third site might be warranted. In contrast, in cases in which the suspected process is expected to be more diffuse, such as toxic injury or myelodysplastic syndromes[14,15] or in re-evaluation after treatment when the

Table 11 Mixed-effects logistic models: main effects and clustering by dog.

Outcome	*n*	Fixed effect	Num DF	Den DF	*F* value	*P*-value
Lymphocytosis (Yes/No)	92	Bx_Site	3	70	1.12	.35
Lymphocytosis (Yes/No)		Path	1	70	4.71	.03
Lymphocytosis (Yes/No)		Path∗Bx_Site	3	70	0.83	.48
Plasmacytosis (Yes/No)	92	Bx_Site	3	70	0.82	.49
Plasmacytosis (Yes/No)		Path	1	70	8.73	.004
Plasmacytosis (Yes/No)		Path∗Bx_Site	3	70	0.94	.43

Abbreviations: Bx_Site = biopsy site; DF = degrees of freedom; ICC = intraclass correlation coefficient; Path = clinical pathologist. Type III tests of fixed effects from mixed-effects logistic regression (binomial, logit link) with a random intercept for dog (Dog_ID). Models were fit with PROC GLIMMIX (Laplace approximation; containment degrees of freedom). Outcomes are the presence/absence of lymphocytosis or plasmacytosis. ICC = Var(Dog_ID)/(Var(Dog_ID) + Residual); $ICC_{lymphocytosis} = 0.92$; $ICC_{plasmacytosis} = 0.62$.

baseline pattern of marrow involvement has already been established, sampling 2 sites might be a reasonable starting approach. However, this distinction remains exploratory and will require confirmation in studies specifically designed and adequately powered to assess disease-specific sampling strategies.

Our study had some limitations. The primary limitation is the small cohort of 16 dogs, which decreases the statistical power to identify more subtle effects, especially within specific disease subgroups. Furthermore, the study population included only dogs that required BM evaluation for existing clinical signs, meaning the results may not be generalizable to the broader range of hematopoietic conditions seen in veterinary medicine. The study's single-center design also impacted the external validity of the results; procedures were performed by experienced personnel following a standardized protocol, and these findings may not be fully reproducible in other clinical environments where levels of expertise and equipment may vary.[2,16–21] A learning curve is associated with mastering multisite sampling, and differences in sample handling among institutions could affect outcomes.

Another key consideration is the methodology used to define the reference comparator used in our study. Rather than relying on an unavailable external gold standard, we used an internal consensus-based composite derived from the most frequently recurring diagnostic codes across all reads and sampling sites. This approach allowed each site-sampling strategy to be evaluated against the cytologic interpretation most reproducibly supported by the total marrow dataset obtained for that dog. As such, our analysis does not simply report percentage agreement among reads but rather provides a structured estimate of how often single-site and multisite sampling strategies recover the consensus cytologic interpretation supported by the full marrow information available in that patient. We also acknowledge that the frequency of nondiagnostic samples is clinically relevant and may limit interpretability in individual cases. Nondiagnostic reads were excluded from truth-set construction because the primary analysis was designed to assess recovery of a truth-consistent interpretation among evaluable cytologic samples.

Although BM trephine core biopsies also were collected from each patient at the time of aspiration, these histopathologic data are reported separately[22] to allow for adequate detailed analysis of both modalities. Consequently, direct cytologic-histologic correlation was not performed using this specific dataset, and future integrated studies will be valuable to fully define the combined diagnostic yield. Also, we did not evaluate the impact of improved recovery of a truth-consistent interpretation on clinical outcomes. Although a higher diagnostic yield is presumed to lead to better patient management, further research is needed to determine if multisite sampling directly translates into more effective therapeutic choices and improved patient health. An additional limitation of our study is that all sampled sites were from the appendicular skeleton, specifically the proximal humerus and iliac crest. We did not include true axial sites such as the sternum. Although a recent study[8] found similar marrow composition between appendicular and axial sites (eg, sternum and rib), it observed slightly higher cellularity in the sternum. Future clinical research comparing sternal aspiration with our current appendicular method could help determine if the patterns of patchy distribution of certain disorders vary among these regions. Finally, although the clinical pathologist was masked to the sampling site and reviewed samples in a randomized, nonsequential order, the necessity of providing CBC and signalment data for clinical interpretation precludes absolute blinding, as unique patient profiles theoretically could be recognized. Also, although our study design provided a controlled interpretive framework by including 2 highly experienced, board-certified clinical pathologists from the same institution, inter-institutional variability was not specifically assessed. Importantly, however, our objective was not to model the full range of observer variability in clinical practice, but rather to compare the relative robustness of single-site vs multisite sampling strategies within a defined expert-read framework.

To conclude, we demonstrated that reliance on a single BM sample in dogs is substantially less robust than sampling multiple sites. Single-site sampling is less reliable for recovering the dominant cytologic interpretation in an individual dog. The addition of a second sample from a different anatomical location significantly improves recovery of that interpretation, supporting multisite sampling as a more informative cytologic sampling strategy in veterinary hematology.

Acknowledgments

The authors gratefully acknowledge the Small Animal Internal Medicine Technicians at the University of Illinois for assisting with the bone marrow procedures.

Author contributions

Kristi M. Smiley (Data curation, Formal analysis, Investigation, Writing—original draft, Writing—review & editing), Amy L. Schnelle

(Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing—review & editing), Michael F. Rosser (Formal analysis, Investigation, Methodology, Writing—review & editing), Sara L. Connolly (Conceptualization, Funding acquisition, Methodology, Writing—review & editing), Rose E. Raskin (Methodology, Writing—review & editing), Nicolas Lopez (Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing—review & editing), and Arnon Gal (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Veterinary Internal Medicine* online.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

University of Illinois Companion Animal Research Grant Program and the Wayne and the Josephine Spangler Fund.

Off-label antimicrobial declaration

The authors declare no off-label use of antimicrobials.

Institutional animal care and use committee or other approval

The study was approved by the University of Illinois Institutional Animal Care and Use Committee (IACUC #19119).

Human ethics approval declaration

The authors declare human ethics approval was not needed.

Appendix

Methods

Our randomized, masked clinical trial (IACUC#19119) was conducted at the University of Illinois Veterinary Teaching Hospital. Sixteen client-owned dogs with hematologic disease requiring bone marrow (BM) sampling were prospectively enrolled from July 2020 to April 2023. Inclusion required body weight $>$ 7 kg. Bone marrow evaluation was part of each patient's diagnostic evaluation for hematologic disorders (eg, bi- or pancytopenia, suspected infectious or neoplastic BM disease, suspected immune-mediated precursor-directed disease). Owners provided written informed consent. A board-certified internist (A.G.) performed all procedures.

The ARROW OnControl Powered Driver system (Teleflex, Morrisville, NC, USA) was used to expedite BM acquisition and minimize procedure time. Four BM aspirates were obtained per dog (RH, LH, RI, and LI) according to a randomized complete block design (Table 1). A CBC with differential and blood smear evaluation was collected at the time of BM sampling to support cytologic interpretation.

Bone marrow was sampled from the widest portion of the cranial–dorsal iliac crest and the flat craniolateral surface of the proximal humerus, lateral and distal to the greater tubercle.[1] After a $<$ 1 cm skin incision, the needle was advanced into the iliac crest parallel to the long axis of the ilial wing.[1] At the humerus, needle orientation was perpendicular to the craniolateral surface.[1] A 15-gauge BM aspiration needle (OnControl system) was used. For each site, a 5-mL syringe was primed with 4% sodium citrate (leaving $<$ 0.1 mL of dead space) to provide anticoagulation without clinically relevant dilution. Upon entering the marrow cavity, constant negative pressure was applied until marrow was visible in the hub, limiting total aspirate volume to $\leq$ 0.2 mL to minimize peripheral blood contamination. Small marrow droplets were immediately dispensed onto approximately 20 glass slides, and smears were prepared using a standard pull-apart technique. Two to 5 high-quality smears containing visible spicules were selected for analysis per site and stained with Wright–Giemsa (Hematek 3000 by Siemens) at the University of Illinois Veterinary Clinical Pathology laboratory. Skin incisions were closed with tissue adhesive.

Each sample received a unique randomized accession number; dog identity and sampling site were masked from the board-certified clinical pathology investigators (A.S., M.R.). Masking was maintained throughout slide evaluation and interpretation.

Primary outcome was the site-specific clinical interpretation, determined by BM cytomorphology integrated with the CBC and the patient's signalment (age, sex). For analysis, each site within each dog received a prespecified numerical code. Clinical pathologists assigned up to three diagnostic codes per site (Table 2a).

Secondary outcomes included: % BM cells (spicule cellularity); average number of spicules per slide; myeloid-to-erythroid ratio (MER; 300-cell differential); blast cells (%; 300-cell differential); mean megakaryocytes per spicule (product of mean megakaryocyte count [average of 2 sets of 10 random 40$\times$ fields] divided by number of spicules); sample quality (ordinal: nondiagnostic, poor, adequate, and excellent); lymphocytosis $>$ 5% (yes/no); plasmacytosis $>$ 2% (yes/no).

Statistical methods

Statistical power was estimated (by authors A.G. and N.L.V.) using bootstrap simulations in R v3.5.1 (R Foundation for Statistical Computing). We ran 10 replicate experiments, each with 1000 resamples. In each resample, 15 dogs were simulated; discordant BM samples were generated using a Bernoulli process (rbinom) with probability = 0.05, reflecting the discordant error rate previously observed in healthy dogs.[2] For each dog, 4 anatomical sites were sampled twice (to model 2 raters; 8 samples/dog). The observed proportion of discordant events per resample was compared with an expected discordant rate of 0.10 (based on investigator experience) using a χ^2 test at α = 0.05. Under this design, n = 15 dogs yielded a mean power of 0.908 (95% CI, 0.902-0.914) to detect a difference between observed and expected discordance. Because 12/36 (33%) BM samples were non-diagnostic in our prior study,[2] we targeted n = 20 dogs to allow for non-diagnostic samples and attrition.

The power sample size estimation addressed the primary hypothesis regarding overall discordance rates. The subsequent

mixed-effects models and agreement analyses were conducted to characterize the specific sources of diagnostic variation (eg, site vs pathologist effects) underlying the primary outcome. These analyses provide robust quantitative estimates of the variance components, but they were not a priori powered to detect specific effect sizes among individual subgroups.

All statistical analyses were performed (by authors A.G. and N.L.V.) using SAS 9.4 (SAS Institute Inc., Cary, NC). Descriptive statistics (mean, SD, median, and range) were obtained for continuous variables and counts (%) for categorical variables. For the analysis of the primary hypothesis, the data were kept in long form with 1 record per dog, sampling site, and clinical pathologist (2 clinical pathologists per site). For each dog, we defined a truth set of cytology codes as the codes with the highest overall frequency across the 4 sites and both clinical pathologists (δ = 0; ties allowed; nondiagnostic reads excluded). A read was marked truth-consistent if its code belonged to that dog's truth set.

At the site level, we evaluated 3 ways a site could be considered truth-consistent: ANY (at least 1 clinical pathologist's code is in the dog's truth set), BOTH-any (both clinical pathologists' codes are in the truth set, not necessarily identical), and BOTH-same (the 2 clinical pathologists agree and that agreed code is in the truth set). For each dog and rule, we counted $m \in [0, 1, 2, 3, 4]$ (ie, the number of truth-consistent sites, ranging from 0 to 4). If k sites are sampled without replacement from 4, the probability of obtaining at least 1 truth-consistent site for that dog is $P(\geq 1) = 1 - P(0 \text{ correct}) = 1 - \frac{\binom{4-m}{k}}{\binom{4}{k}}$ (ie, the probability of getting at least 1 truth-consistent site is calculated by taking 1 minus the probability of getting no truth-consistent sites). We computed this probability for each dog and then averaged across dogs. Uncertainty was quantified with a nonparametric bootstrap resampling dogs (2000 replicates), performed separately by rule and k; percentile 95% CIs are reported.

We defined the minimum clinically important difference (MCID) for the probability of a truth-consistent interpretation as a $\geq$ 10% absolute increase in that probability. This threshold corresponds to a number needed to treat of 10 or fewer (ie, for every 10 dogs sampled at multiple sites, at least one unrepresentative cytologic interpretation is prevented). Given the substantial clinical consequences of an unrepresentative cytologic interpretation, a 10% improvement in yield is widely accepted in comparable hematopathology guidelines used in humans (eg, bilateral vs unilateral staging) as sufficient to justify the additional procedural time.[3]

To compare success across numbers of sites, we formed within-dog paired differences in probability for adjacent contrasts (eg, 2–1, 3–2, and 4–3) under each rule and summarized mean differences with bootstrap 95% CIs as above. As a complementary hypothesis test that does not assume normality, we used Wilcoxon signed-rank tests on these within-dog differences (2-sided).

To assess whether per-read truth consistency differed by anatomical site or clinical pathologist, we fit a mixed-effects logistic regression to model binary outcome variables. The model included the fixed effects for sampling site (LH, LI, RH, and RI) and clinical pathologist, and a random intercept for dog to account for clustering within dog. The data were analyzed after logit transformation with Laplace approximation. We reported type III tests of fixed effects, model-based marginal means on the probability scale (inverse-link transformation and 95% CI), and Tukey–Kramer-adjusted pairwise site comparisons.

Agreement was examined in 2 settings: inter-observer (same site, different pathologist) and intra-observer (same pathologist, different sites within the same dog). Because standard Cohen's κ is undefined if the contingency table is not square, we enforced square tables for all agreement analyses by populating zero-weight rows and columns for any code observed in the global dataset but missing from a specific rater pairing. We calculated the percentage of agreement with exact Clopper–Pearson 95% CI. To assess chance-corrected agreement, we computed Cohen's simple κ. Given the relatively small sample size and known instability of asymptotic standard errors for κ in sparse tables, we derived robust 95% CIs for κ using a nonparametric bootstrap (2000 replicates) stratified by site and clinical pathologist. As a sensitivity analysis to address the kappa paradox (where high agreement yields low κ because of skewed prevalence), we also calculated Gwet's AC1 statistic. For intra-observer agreement, we analyzed all 6 possible site-pairs (LH–LI, LH–RH, LH–RI, LI–RH, LI–RI, and RH–RI) for each pathologist. To provide a summary measure of site-to-site consistency, we pooled data across both pathologists and recalculated these metrics for each site pair using frequency weights. Figures display the mean probability of obtaining $\geq$ 1 truth-consistent site vs the number of sampled sites included in the combination (k) with bootstrap 95% CIs, model-based per-read truth-consistency by site with 95% CIs, and κ "forest" plots by site pair.

Standard diagnostic plots, including residuals and influence plots, were examined to verify the assumptions underlying the models. Mixed-effects models were used to compare sampling sites, accounting for the fact that each dog contributed multiple measurements. The mixed model included the fixed effect of sampling site (LH, LI, RH, and RI), the clinical pathologist (1 vs 2), and their interaction, and the random effect of dog to account for baseline differences among dogs. Skewed counts and ratios were analyzed after logarithmic transformation, and proportions were evaluated on the log odds (logit) scale to ensure that predicted values remained within the 0 to 1 range. The models were fitted using the MIXED procedure. Pairwise comparisons between sites were adjusted using the Tukey–Kramer adjustment.

We analyzed the yes and no outcomes with mixed-effects logistic regression using the GLIMMIX procedure. This method handles binary data while accounting for clustering of measurements within dogs by adding a random intercept for dog. Fixed effects were sampling site, clinical pathologist, and their interaction. The data were analyzed after a logit transformation with Laplace approximation. We report overall (type III) tests, obtained predicted probabilities with 95% CIs from least-squares means on the probability scale (ILINK), and used Tukey–Kramer adjustment for pairwise site comparisons. Model convergence and standard residual diagnostics were reviewed to confirm adequate fit. Lastly, the ordinal sample-quality score (1-4) was analyzed using a proportional odds mixed-effects cumulative logit model (GLIMMIX, Laplace approximation), with sampling site, clinical pathologist, and their interaction as fixed effects and a random intercept for dog to account for repeated samples. Results are reported as type III tests and cumulative odds ratios with 95% CIs.

For multiplicity control, the primary hypothesis employs a single estimand (probability of $\geq$ 1 truth-consistent site by k under 3 diagnostic rules) evaluated with bootstrap CIs, which do not

incur type I error. Paired within-dog contrasts (2–1, 3–2, and 4–3) were evaluated using two-sided Wilcoxon signed-rank tests without additional multiplicity adjustment, because they form a single family of related paired comparisons.

Secondary characterization analyses (mixed-effects models, κ estimates, and site comparisons) employed Tukey–Kramer adjustment for pairwise site comparisons and were interpreted at exploratory significance levels where applicable ($P < .10$ for marginal effects suggesting direction of variation). Kappa estimates were reported with 95% CIs rather than hypothesis tests and reflect descriptive agreement structures. All secondary type III fixed-effect tests were reported with CIs and should be interpreted as exploratory summaries of variation rather than confirmatory hypothesis tests.

References

1. Raskin RE, Messick JB. Bone marrow cytologic and histologic biopsies: indications, technique, and evaluation. *Vet Clin North Am Small Anim Pract*. 2012;42:23-42. https://doi.org/10.1016/j.cvsm.2011.10.001
2. Lee SH, Erber WN, Porwit A, Tomonaga M, Peterson LC, International Council for Standardization in Hematology. ICSH guidelines for the standardization of bone marrow specimens and reports. *Int J Lab Hematol*. 2008;30:349-364. https://doi.org/10.1111/j.1751-553X.2008.01100.x
3. Defarges A, Abrams-Ogg A, Foster RA, Bienzle D. Comparison of sternal, iliac, and humeral bone marrow aspiration in beagle dogs. *Vet Clin Pathol*. 2013;42:170-176. https://doi.org/10.1111/vcp.12036
4. Aubry OA, Spangler EA, Schleis SE, Smith AN. Evaluation of bone marrow aspirates from multiple sites for staging of canine lymphoma and mast cell tumours. *Vet Comp Oncol*. 2014;12:58-66. https://doi.org/10.1111/j.1476-5829.2012.00331.x
5. Wang J, Weiss LM, Chang KL, et al. Diagnostic utility of bilateral bone marrow examination: significance of morphologic and ancillary technique study in malignancy. *Cancer*. 2002;94:1522-1531. https://doi.org/10.1002/cncr.10364
6. Raskin RE, Krehbiel JD. Histopathology of canine bone marrow in malignant lymphoproliferative disorders. *Vet Pathol*. 1988;25:83-88. https://doi.org/10.1177/030098588802500112
7. Gal A, Burchell RK, Worth AJ, Lopez-Villallobos N, Marshall JC, MacNeill AL. The site of bone marrow acquisition affects the myeloid to erythroid ratio in apparently healthy dogs. *Vet Pathol*. 2018;55:853-860. https://doi.org/10.1177/0300985818780469
8. Diamantino GML, Beeler-Marfisi J, Foster RA, Sears W, Defarges A, Bienzle D. Comparison of the bone marrow composition in samples from rib, ilium, sternum, and humerus of dogs. *J Vet Diagn Invest*. 2026;38:50-58. https://doi.org/10.1177/10406387251387786
9. Moore PF. A review of histiocytic diseases of dogs and cats. *Vet Pathol*. 2014;51:167-184. https://doi.org/10.1177/0300985813510413
10. Affolter VK, Moore PF. Localized and disseminated histiocytic sarcoma of dendritic cell origin in dogs. *Vet Pathol*. 2002;39:74-83. https://doi.org/10.1354/vp.39-1-74
11. Abreu R, Martinho A, Noiva R, et al. Osteomyelitis caused by *Aspergillus terreus* complex in a dog: a case report. *BMC Vet Res*. 2023;19:76. https://doi.org/10.1186/s12917-023-03628-x
12. Armstrong PF, Sigler L, Sutton DA, Grooters AM, Hitt M. Fungal myelitis caused by *Phialosimplex caninus* in an immunosuppressed dog. *Med Mycol*. 2012;50:509-512. https://doi.org/10.3109/13693786.2011.633106
13. Levy A, Harran N, Hammer M, Bennaim M. Fungal osteomyelitis caused by *Candida glabrata* in a Groenendael dog. *Vet Rec Case Rep*. 2020;8:e001329. https://doi.org/10.1136/vetreccr-2020-001329
14. Feeney DA, Sharkey LC, Steward SM, et al. Applicability of 3T body MRI in assessment of nonfocal bone marrow involvement of hematopoietic neoplasia in dogs. *J Vet Intern Med*. 2013;27:1165-1171. https://doi.org/10.1111/jvim.12151
15. Fry MM, Forman MA. 5-Fluorouracil toxicity with severe bone marrow suppression in a dog. *Vet Hum Toxicol*. 2004;46:178-180.
16. Swords RT, Anguita J, Higgins RA, et al. A prospective randomised study of a rotary powered device (OnControl) for bone marrow aspiration and biopsy. *J Clin Pathol*. 2011;64: 809-813. https://doi.org/10.1136/jclinpath-2011-200047
17. Berenson JR, Yellin O, Blumenstein B, et al. Using a powered bone marrow biopsy system results in shorter procedures, causes less residual pain to adult patients, and yields larger specimens. *Diagn Pathol*. 2011;6:23. https://doi.org/10.1186/1746-1596-6-23
18. Miller LJ, Philbeck TE, Montez DF, et al. Powered bone marrow biopsy procedures produce larger core specimens, with less pain, in less time than with standard manual devices. *Hematol Rep*. 2011;3:e8. https://doi.org/10.4081/hr.2011.e8
19. Reed LJ, Raghupathy R, Strakhan M, et al. The OnControl bone marrow biopsy technique is superior to the standard manual technique for hematologists-in-training: a prospective, randomized comparison. *Hematol Rep*. 2011;3:e21. https://doi.org/10.4081/hr.2011.e21
20. Marinelli LM, Fang H, Howard MT, et al. Bone marrow biopsy operator experience and impact on aspirate, biopsy, and ancillary testing quality. *Mayo Clin Proc Innov Qual Outcomes*. 2018;2:241-247. https://doi.org/10.1016/j.mayocpiqo.2018.06.007
21. Brestoff JR, Toland A, Afaneh K, et al. Bone marrow biopsy needle type affects core biopsy specimen length and quality and aspirate hemodilution. *Am J Clin Pathol*. 2018;151:185-193. https://doi.org/10.1093/ajcp/aqy126
22. Smiley KM, Connolly SL, Raskin R, et al. Multi-site bone marrow core biopsy improves diagnostic accuracy in dogs with hematologic disease. *PloS One*. 2026;21:e0343251. https://doi.org/10.1371/journal.pone.0343251

Journal of Veterinary Internal Medicine, 2026, 40, aalag121
https://doi.org/10.1093/jvimsj/aalag121
Original Research