12-2021

INVESTIGATION OF SEMIQUANTITATIVE MARROW IRON SCORING AND RETICULOCYTE HEMOGLOBIN CONTENT (CHr) IN HEALTHY AND ILL CANINES WITH AND WITHOUT CLINICAL EVIDENCE OF BLOOD LOSS.

Grace A. Pawsat
gpawsat@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Grace A. Pawsat entitled "INVESTIGATION OF SEMIQUANTITATIVE MARROW IRON SCORING AND RETICULOCYTE HEMOGLOBIN CONTENT (CHr) IN HEALTHY AND ILL CANINES WITH AND WITHOUT CLINICAL EVIDENCE OF BLOOD LOSS.." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Comparative and Experimental Medicine.

Deanna M. Schaefer, Major Professor

We have read this thesis and recommend its acceptance:

Mike Fry, Shelly Olin

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

INVESTIGATION OF SEMIQUANTITATIVE MARROW IRON SCORING AND RETICULOCYTE HEMOGLOBIN CONTENT (CHr) IN HEALTHY AND ILL CANINES WITH AND WITHOUT CLINICAL EVIDENCE OF BLOOD LOSS.

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Grace Pawsat
December 2021

ACKNOWLEDGEMENTS

I would like to thank Dr. Deanna Schaefer, Department of Biomedical and Diagnostic Sciences, University of Tennessee, Knoxville, TN for encouragement and support throughout my master's research and course work. I would not be where I am today without her.

I would like to thank Dr. Shelly Olin, Department of Small Animal Clinical Sciences, University of Tennessee, Knoxville, TN for encouragement and support throughout my research endeavor. Her clinical expertise and teaching abilities were instrumental to my success.

I would like to thank Dr. Michael Fry, Department of Biomedical and Diagnostic Sciences, University of Tennessee, Knoxville, TN for always being there for me. He has encouraged me and pushed me to higher heights.

Thank you to Dr. Jennifer Stokes, Department of Small Animal Clinical Sciences, University of Tennessee, Knoxville, TN for her support, clinical expertise, and thoughtfulness.

Thank you to Michelle Story and Aimee Hebrard, Department of Biomedical and Diagnostic Sciences, University of Tennessee, Knoxville, TN for all their assistance in sample processing and quality assurance.

I would like to thank Xiaojuan Zhu, Department of Office of Information Technology, University of Tennessee, Knoxville, TN for her help in research, statistical consulting. I learned so much from her in a short amount of time.

ABSTRACT

Background: Insufficient iron for erythropoiesis can occur in multiple conditions, including absolute iron deficiency, which is often caused by chronic external hemorrhage in dogs. Distinguishing this from other causes of iron-restricted erythropoiesis allows appropriate intervention. Decreased marrow iron assessed by Prussian blue staining is considered the gold standard to diagnose absolute iron deficiency but scoring systems are not validated in dogs. The performance of standardized semiquantitative marrow scoring of iron deposits in dogs has not been previously reported or correlated to iron biomarkers in blood.[1-4]

Objectives: Objectives included evaluation of the effects of different staining sequences, evaluation the technical performance of bone marrow scoring systems used in human medicine, and assessment of the clinical utility of marrow iron scoring (Gale method) and reticulocyte hemoglobin content (CHr) in healthy and clinically ill canines with and without evidence of blood loss.

Results: Established human scoring via the Gale iron bone marrow classification method (scale 0-6) can be applied to healthy and ill canine bone marrow aspirates stained with Prussian blue with acceptable intraobserver (substantial to perfect) and interobserver agreement (fair).[5] The Gale score did not correlate with the CHr [reticulocyte hemoglobin content] as expected, given that CHr can be affected by different types of iron restricted erythropoiesis. The odds of a low CHr (equal or less than 24.4 pg) in a canine patient is approximately 3 times higher in patients that have a low median Gale score (2 or less). Descriptive statistics for Gale marrow scoring in clinically healthy research canines were as follows: range (4-5), mean (4.3), and median (4.17). Most dogs with a low Gale score and low CHr (equal or less than 24.4 pg) had absolute iron deficiency; however, a few dogs had evidence of additional iron sequestration and functional iron deficiency. A low Gale score and normal CHr could indicate a subclinical iron deficiency.

Conclusions: The Gale scoring method performs well on canine bone marrow aspirates but should always be interpreted together with CHr and other laboratory data and clinical findings for accurate evaluation for one or more potential processes of iron restricted-erythropoiesis in the patient.

PREFACE

Bone marrow iron grading is important to identify absolute iron deficiency, which in dogs is commonly due to chronic external blood loss, and to distinguish it from other causes of iron deficiency. Veterinary medicine does not have a standardized semi-quantitative method of grading iron stores during microscopic evaluation of bone marrow samples. Marrow iron is often only reported as present or absent. Establishing a grading system for iron scoring in dog marrow and evaluating its clinical utility will help with diagnosis and treatment of absolute iron deficiency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
ASVCP	American Society for Veterinary Clinical Pathology
CBC	Complete Blood Count
CHr	Reticulocyte Hemoglobin Content
CRP	C Reactive Protein
CUHA	Cornell University Hospital for Animals
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
IL 6	Interleukin 6
ICC	Intraclass correlation coefficient
MCHC	Mean Cell Hemoglobin Concentration
MCV	Mean Cell Volume
NSAID	Non-steroidal anti-inflammatory
Pg	Picograms
SOP	Standard Operating Procedure
TIBC	Total Iron Binding Capacity
UTVMC	University of Tennessee Veterinary Medical Center
%Sat	Percent Saturation of transferrin

CHAPTER ONE
INTRODUCTION AND GENERAL INFORMATION

Background

Microscopic evaluation for decreased marrow iron stores has been considered the practical gold standard for diagnosis of absolute iron deficiency, compared to blood biomarkers of iron such as reticulocyte hemoglobin content (CHr), soluble transferrin receptor, serum iron, total iron binding capacity (TIBC), and ferritin.[6-8] Veterinary pathologists commonly only classify marrow iron as present or absent. However, human medicine utilizes two different iron marrow scoring systems: the Gale method and sideroblast method.[6, 9] Respectively, these reflect the total stored iron in the marrow and cellular utilization of iron.[6, 9] These scoring systems require Prussian blue staining to highlight extracellular and intracellular iron deposits.[7, 10] Prior to this study, there have not been any published studies in canine marrow samples documenting differences in staining sequences and the technical performance of these human iron grading systems.[5] Further investigation of marrow iron along with blood biomarkers of iron may give additional information about the type(s) of iron deficiency processes in the canine patient and determine if the less invasive blood biomarkers are acceptable alternatives to marrow sampling.

Because bone marrow sampling is an invasive and painful procedure, less invasive blood biomarkers have been investigated as an alternative to aid in determination of marrow iron status.[1-3, 11, 12] CHr has previously been evaluated as a promising, non-invasive reflection of iron status in canines. [1, 4, 14-17] However, it is not a consistent indicator of absolute iron deficiency in canines and does not discriminate between iron restricted erythropoiesis categories (see below). [16] CHr is a measurement of the amount of hemoglobin in reticulocytes and can be decreased in conditions of iron deficiency since iron is an essential component of hemoglobin. Reticulocytes are newly formed red cells that have been recently released from the bone marrow; therefore, CHr is a more direct reflection of the patient's current red cell hemoglobin content and iron status when compared to the hemoglobin content of older red cells.[1, 4, 6] CHr can have high sensitivity and specificity for diagnosis of absolute iron deficiency in certain situations in humans.[13, 14] However, in both humans and canines, CHr can be affected by iron sequestration (inflammation) and functional iron deficiency.[4, 11] Evaluating Gale iron marrow scores concurrently with CHr may elucidate how effectively CHr can be used as a surrogate marker of marrow iron stores.

There are three broad categories of iron deficiency described below, and distinction of these categories can be difficult, especially if only blood biomarkers are available.[1-3, 11, 12] This study investigates utility of concurrent

semiquantitative iron marrow scoring, along with CHr and other blood iron biomarkers, for classification of iron deficiency categories in the canine patient.

Categories of Iron Restricted Erythropoiesis

Iron restricted erythropoiesis is a condition in which there is inadequate iron available for erythropoiesis to proceed normally. The three categories of iron restricted erythropoiesis are absolute iron deficiency, iron sequestration, and functional iron deficiency.[11] In dogs, absolute iron deficiency most commonly occurs due to chronic blood loss, but may also occur with feeding of imbalanced homemade or vegetarian diets.[11]

Iron sequestration applies to anemia of inflammation and chronic disease.[11, 14] It occurs due to inflammatory cytokines (e.g. IL 6) driving an increase in hepcidin, which is a main regulator of iron in the body.[11, 14, 15] Hepcidin causes the degradation of portal exits (ferroportin) for intracellular iron. Without ferroportin, iron is sequestered within cellular storage sites, such as intestinal epithelial cells, macrophages, and hepatocytes, resulting in an iron deficient state and consequential anemia.[14]

Functional iron deficiency occurs when there are adequate iron stores but inadequate delivery of iron to erythroid precursors.[4, 7] It is seen when erythropoietin (erythrocyte growth factor) is increased, promoting usage of iron for erythropoiesis at an increased rate so that iron demand for erythropoiesis exceeds supply via mobilization from tissue stores. This can be seen in patients treated with exogenous erythropoietin or in patients with hemolytic anemia [3]

Diagnosis of Iron Deficiency Anemia

Diagnosis of the different types of iron restricted erythropoiesis is dependent upon a holistic approach including history, clinical presentation, blood biomarkers, and iron marrow stores. Documentation of chronic blood loss, diet, inflammatory conditions, and administration of any type of erythropoietic stimulant must be reported. Clinical presentation of symptomatic anemia such as pallor, weakness, heart murmur, and possible melena must be noted.[16] Serum iron and CHr are usually decreased in canines with anemia of inflammation or absolute iron deficiency. Serum TIBC (an indicator of concentrations of serum transferrin, the main serum iron transport protein) is not usually increased in absolute iron deficiency or anemia of inflammation, and may be normal or decreased.[17] Stainable marrow iron and serum ferritin (storage iron in blood) are decreased in absolute iron deficiency, but both are adequate to increased in anemia of inflammation. Patients with functional iron deficiency will have adequate iron marrow stores but decreased iron available to erythroid precursors.[4, 7] Semiquantification of marrow iron may be especially helpful along with the blood biomarkers to determine the type of iron deficiency for appropriate patient treatment.

Utility of Destaining Marrow Aspirates

Many veterinary clinical pathology laboratories stain routine bone marrow aspirates with a Romanowsky stain such as Wright or Giemsa. These stains highlight cellular detail to allow for evaluation of cell morphology, but iron is not highlighted and can be difficult to see, especially in thickly prepared smears. Additionally, in some cases only a small amount of marrow is collected, which may yield only one or two diagnostic smears in which Romanowsky staining is often prioritized based on the patient's clinical needs. To avoid having to resample the marrow in those cases, it would be useful to be able to destain Romanowsky-stained smears after evaluation of cell morphology, then restain the smears with Prussian blue if semi-quantification of iron is needed.

Evaluation of Iron Marrow

Qualitative assessment of marrow iron stores is a routine component of microscopic marrow evaluation in dogs, where marrow iron stores are often simply reported as present or absent. To date there are no published guidelines for interpretation of semiquantitative iron assessments such as the Gale scoring method in veterinary medicine. Therefore, it is unclear whether anemic dogs with very small amounts of microscopically detectable iron should be diagnosed with absolute iron deficiency or should be evaluated for other causes of anemia.

Microscopic marrow iron assessment does have notable limitations, though, including that some studies in humans had significant interobserver variation in assessment of iron stores. [6-8] To date there have been no publications to evaluate the technical performance of the Gale or sideroblast iron grading schemes in canine marrow samples. Additionally, there have been no studies to evaluate the effect of destaining slide previously stained with a Romanowsky stain before restaining with Prussian blue.

CHAPTER TWO

LITERATURE REVIEW

Iron Marrow Scoring in Human Literature

Marrow iron scoring in humans is based on two scoring systems. The Gale scoring system is semiquantitative and based on identification of large deposits of storage iron within macrophages. The sideroblast scoring system is based on identification of small iron granules in the cytoplasm of erythroid precursors. Prussian blue staining is required for optimal visualization of iron by either scoring system.

Gale Scoring

A semiquantitative scheme for assessing bone marrow iron stores, the Gale method, has been described for humans.[18] It grades large deposits of hemosiderin in macrophages of Prussian blue stained bone marrow smears on a scale of 0 to 6, with grades of 1-3 considered normal (Table 1). [18, 19] Scores are expected to be decreased with absolute iron deficiency, but may be normal or increased in other causes of iron-restricted erythropoiesis.[7]

Sideroblast Scoring

A second method, the sideroblast method, has also been described in humans. This method involves quantification of the percentage of nucleated red cell precursors containing 1-5 fine cytoplasmic iron granules (“sideroblasts”), with greater than 20-30% of erythroid precursors expected to be sideroblasts in humans (Figure 1). [7] [9] Sideroblastic iron is theorized to provide a better indication of cellular utilization of iron, and together with the Gale iron score might aid in diagnosis of iron sequestration and functional iron deficiency. Both of these conditions are associated with adequate iron scores (normal Gale score) but insufficient iron availability to meet the increased demand.[7]

Challenges of Diagnosing Absolute Iron Deficiency

Several challenges may arise when diagnosing absolute iron deficiency based solely on routine erythrocyte values from the Complete Blood Count (CBC) and on iron biomarkers in blood.[2, 7, 20] Patients with iron sequestration and those with absolute iron deficiency can both have a low CHr and low serum iron.[2, 3, 20] Additionally, a microcytic erythrocyte population (indicated by decreased Mean Cell Volume [MCV] on the CBC and decreased CHr), is usually suggestive of absolute iron deficiency, but can also be noted in patients with portosystemic shunting and hereditary microcytosis (i.e. Akitas, Jindos, Shiba Inus, and Chow breed canines).[2, 20] Using erythrocytes values and iron biomarkers are a less

invasive way to estimate marrow iron stores. However, a bone marrow aspirate stained with Prussian blue is often required for definitive diagnosis of absolute iron deficiency.[7]

Correlation of Biomarkers and Marrow Iron in Humans and Canines

Many human studies have reported the correlation between blood iron biomarkers and marrow iron stores. These blood biomarkers include but are not limited to CHr, serum ferritin, MCV, Mean Cell Hemoglobin Concentration (MCHC), low density hemoglobin, and TIBC.[7, 13, 15] Findings are similar in canines and humans with absolute iron deficiency and include microcytosis (decreased MCV), hypochromasia (decreased MCHC), low ferritin, and low CHr. One notable difference is that TIBC is increased in humans in response to iron deficiency anemia, whereas it is not commonly increased in canines.[17, 20]

One human study demonstrated that CHr had the highest sensitivity and specificity for iron deficiency when compared to other iron biomarkers.[13] Additionally some human studies incorporate C-reactive protein (CRP) concentrations (increased in iron sequestration) and soluble transferrin receptor concentrations (increased in absolute iron deficiency) as part of an intensive marrow grading method.[7] This intensive method considerably improves the ability to distinguish causes of iron restricted erythropoiesis in humans.[7] However, the clinical utility of CRP in distinguishing causes of iron restriction in dogs has not been thoroughly evaluated [4], and an assay for soluble transferrin receptor is not available for canines.

Core Biopsy Verses Marrow Aspirates for Iron Scoring

In low numbers of human cases, iron assessment can be discrepant between the marrow core biopsy and aspirate smears.[19] In some cases, iron may be detectable on bone marrow core biopsy when none is detected on aspirate smears.[19] This may be partially explained because at least seven marrow particles must be present on aspirate smears to establish absolute iron deficiency by the Gale method.[7, 19] Acquiring this number of particles may be a challenge when only a small sample volume of bone marrow is collected. Also, aspirates tend to have irregular distribution of iron which can further complicate iron assessment.[6] Iron assessment by core biopsies may also be erroneous at times. Iron staining by Prussian blue is recommended on plastic-embedded samples.[6] Paraffin-embedded core biopsies can be used, but when combined with decalcification, some iron can be lost.[19]

CHAPTER THREE

PILOT AND UTILITY STUDY

Objectives

1. Evaluate the effect of different staining sequences on the iron scoring systems.
2. Evaluate interobserver agreements of Gale score in healthy and ill canines, with and without blood loss.
3. Investigate the correlation between Gale iron scores and CHr in healthy and those ill canines with and without blood loss.
4. Provide descriptive statistics for iron Gale scores and CHr in healthy research dogs and those ill dogs that are hematologically normal.
5. Determine how often that dogs with low marrow iron +/- low CHr have evidence of absolute and subclinical iron deficiency based on clinical findings and routine bloodwork.

Hypotheses

1. Destaining will decrease Gale iron scores by stripping stainable iron.[21]
2. The interobserver agreement will be fair to substantial with the Gale method in both healthy canines and those ill canines with and without blood loss.
3. The Gale score will not correlate to CHr in healthy and ill canines, including categories of canines with and without blood loss because ill dogs may commonly have iron sequestration due to hepcidin (e.g. anemia of inflammatory disease) which may decrease the CHr but not decrease the Gale score.
4. In healthy dogs and in clinically ill dogs that are hematologically normal, the Gale iron score will be at least 1, and the CHr values will be similar to those published reference intervals 24.5 to 28.6 pg (picograms).[2, 3]
5. Dogs with low Gale score and low CHr will commonly have other clinical or laboratory evidence of absolute iron deficiency, while those with low Gale score and normal CHr would uncommonly have other evidence of absolute iron deficiency.

CHAPTER FOUR

MATERIALS AND METHODS

Scoring Performance

Pilot Study

Bone marrow material remaining from aspirates collected as part of the clinical evaluation of 20 canine patients at the University of Tennessee Veterinary Medical Center (UTVMC) were selected prospectively by convenience sampling (samples where nearby and easily attainable) from April 2017 to September 2018. As part of a standard clause on the UTVMC admissions form, clients of these patients acknowledged by signature that any remaining diagnostic material could be used for research. Eight canine patients were excluded because at least one of the paired aspirate smears contained less than 7 marrow particles. Seven particles is the minimum recommended number to determine lack of stainable iron, as seen with absolute iron deficiency.[6]

The remaining 12 canines ranged in age from 3 to 11 years. These represented a variety of breeds, including Golden Retriever (2), Beagle (2), Labrador Retriever (2), West Highland White Terrier (1), mixed breed (1), Staffordshire Terrier (1), Collie (1), Great Dane (1), and Bichon Frisé (1). Clinical diagnoses included leukemia (5), immune-mediated thrombocytopenia (2), vector-borne disease (2), precursor-directed immune-mediated anemia (1), chronic kidney disease (1), and anemia of undetermined origin (1).

Staining

All marrow aspirates were collected by UTVMC veterinarians in the Department of Small Animal Clinical Sciences. Two smears were prepared from each of the 12 canine samples, generating a total of 24 smears. One smear from each canine was stained directly with Prussian blue, while the other was first stained with Wright stain, destained, and then restained with Prussian blue.

The initial Wright staining of the restained smears was performed via automatic slide stainer (Aerospray Pro, ELITech Group Wescor, Puteaux, France). The slides were then decolorized with an ethanol/hydrochloric acid (9:1) solution until colorless, followed by a rinse with deionized water.

All 24 smears were then stained with Perls' Prussian blue by a standard, staining protocol performed at UTVMC Histology laboratory. This stain included steps of 1) fixation in methanol for 10 minutes, 2) incubation in a 1:1 solution of 10% potassium ferrocyanide and 20% hydrochloric acid for 10 minutes, and 3) counterstaining by flooding with 1% safranin for 10 seconds. Distilled water

rinses were used after each step. All smears were mounted with CoverSeal™-X (Cancer Diagnostics, Durham, NC) and a coverslip applied. Positive controls for each run included a histological sample of canine liver with increased amount of large globular iron deposits and a canine blood film with moderate numbers of siderocytes. These controls represented stainable storage iron for Gale scoring and intracellular small iron granules for sideroblast scoring, respectively.

Scoring

The resulting 24 Prussian blue stained smears were randomized, and all labels were concealed with a random number creating a double-blind study with three independent observers. To assess intraobserver repeatability, after one round of scoring, samples were re-randomized and relabeled for a second round of scoring in which observers where blinded to previous scores. Observers included two experienced boarded veterinary clinical pathologists and one veterinary clinical pathology resident. Each observer used the same model of microscope (Olympus BX41, Olympus Corporation, Tokyo, Japan). All marrow aspirate smears were examined microscopically using the following protocols for both Gale and Sideroblast scores. Each observer independently recorded both Gale and Sideroblast scores for each sample without knowledge of results from the other observers.

Gale

A published Gale scoring system from human literature was utilized (Image 1).[18] A short training session for all observers along with complied microscopic images (Figure 2) corresponding to each Gale score (Table 1) was provided by author DS. The entire smear was scanned using the 10x objective. Samples with variable iron in different marrow particles were scored based on the highest iron score, as they reflect the maximum amount of iron stores.

If iron was visible during this low power scan, Grades 3 to 6 were assigned, and a high-power scan was not performed for that sample. A grade 2 was assigned if there were small, sparsely distributed iron particles on low power, which were confirmed as iron using a high-powered scan with the 50x or 100x objective. If no iron was visible at low power, a high-powered scan evaluated at least 7 marrow particles for scant iron, distinguishing grades 0 and 1.

Sideroblast

The sideroblastic method was performed using the 100X oil immersion objective in a well spread area, avoiding areas of excess macrophagic iron deposits. One hundred nucleated erythrocyte precursors were counted, enumerating the percentage that were sideroblasts. Any erythroid precursor was considered a sideroblast (Figure 1) if it contained 1-5 small blue iron granules in the cytoplasm. Iron granules occurring in a ring around the erythroid nucleus or diffusely, are

termed ring sideroblasts and were also included in the sideroblast percentage.[10]

Statistical Analysis
All 24 marrow smears, including those directly stained and those destained /restained, and repeated scores recorded by all three observers were included in the analysis, yielding a total of 144 observations. The statistical software (SAS v 9.4, Cary, NC) was utilized for all analyses. A two-way analysis of variance (ANOVA) was used to test the effect of destaining on the Gale and Sideroblast scoring methods. Staining and observer were designated as treatments, blocking on patient. The null hypotheses stated that all treatment means for staining and observer were equal for both the Gale and Sideroblast scoring methods. Cohen's kappa was used to determine intraobserver agreement between each observer's first and second scorings of the Gale method for each patient. Intraobserver agreement of the sideroblast method was evaluated by Pearson correlation for continuous data. Fleiss' kappa was used to determine interobserver agreement of the three observers based on Gale method scorings. Interobserver variability was analyzed via observer effect on sideroblast score by two-way analysis of variance (ANOVA). The statistical software SPSS (IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp) was used to perform an intraclass correlation coefficient (ICC) to determine interobserver reliability. ICC has more power and is based on continuous data and was added on as an additional way to audit interobserver agreement.[22, 23]

Cohen's Kappa (intraobserver variability) and Fleiss' Kappa (interobserver variability) values were interpreted with the Landis and Koch's kappa agreement scoring system. Interpretation of weighted kappa (K_ω) values were as follows: <0 no agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1 near perfect or perfect agreement.[24] The ICC was interpretated with published, qualitative guidelines were as follows: <0.5 poor, 0.5 – 0.75 moderate, 0.75 – 0.9 good, and >0.9 excellent correlation.[25]

For intraobserver agreement, correlation (rho) was classified as positive or negative. Interobserver variability was analyzed via observer effect on sideroblast score by two-way analysis of variance (ANOVA). Correlation strength was based on the rho values. A correlation (rho) of 0 to 0.3 was considered weak correlation, 0.3 to 0.7 as moderate correlation, and above 0.7 as strong correlation.[26]

Larger Study Set

A larger and separate set of canine marrow aspirates and concurrent blood samples included 139 clinically ill canines from Cornell University Hospital for

Animals (CUHA), 20 healthy research canines from UTVMC, and 12 canines enrolled in a previous UTVMC nutritional iron deficiency study from 2005 [1].

Inclusion criteria for all sample groups were at least 7 marrow particles present on the Prussian blue stained bone marrow smear and a CBC including reticulocyte parameters obtained within 1 day of marrow sampling.

Additional inclusion criteria for the healthy research dogs were that they must be at least one year of age and must be clinically healthy on physical exam, with no evidence of external bleeding and no medications other than routine parasite prevention within 10 days prior to sample collections. Samples from the healthy research dogs were excluded if there was more than one CBC value or serum iron biomarker outside of the reference intervals.

For dogs with experimentally induced iron deficiency, an additional inclusion criterion was that the dog had to have been on an iron-restricted diet for at least one month before marrow sampling.

After application of these inclusion and exclusion criteria, the final sample set included samples from 118 clinically ill dogs from CUHA, 12 healthy research dogs from UTVMC, and 4 canines from the nutritional iron deficiency study.

Canine marrow aspirate smears from the 118 patients at CUHA were retrospectively identified from 2010 to 2014. These were collected and stained with Prussian blue as part of the diagnostic evaluation of the patient, so required no IACUC approval. Ages spanned 1 to 14 years. Breeds included mixed breed (23), Golden Retriever (16), Labrador Retriever (15), Bernese Mountain dog (5), Boxer (4), Dachshund (3), German Shepherd (3), Rottweiler (3), and 1-2 each of various other breeds. There were a variety of clinical diagnoses, which included neoplasia, inflammation, inflammation concurrent with neoplasia, immune mediated disease, hemolytic anemia, external blood loss, and iron deficiency anemia.

Blood and bone marrow samples were collected from sedated healthy research dogs from a colony maintained by the UTCVM in 2020 and from research dogs with experimental nutritional iron deficiency at UTCVM in 2005. All samples were collected in accordance with an approved IACUC protocol (UT IACUC protocols 2791-1020 and 1385) Marrow was collected from a proximal humerus by IACUC approved veterinarians in all cases. The healthy research colony included mostly Beagles and 2 hounds, all aged 1 to 6 years. After application of inclusion and exclusion criteria, the sample of healthy research dogs included 12 canines aged 1 to 4 years old, and all were Beagles. The research dogs with experimental iron deficiency were young (ages 3-7 months) and mixed breed. After application of inclusion and exclusion criteria, this group included 4 dogs.

All marrow samples from the CUHA patients were stained with Prussian blue at the CUHA clinical pathology laboratory. Samples from the healthy UTVMC research canines were shipped to CUHA for Prussian Blue staining in order to mitigate interlaboratory variability. However, the samples from the dogs with experimentally induced nutritional iron deficiency had already been stained by Prussian blue through the UTVMC histology laboratory using the same protocol as in the pilot study. The Cornell clinical pathology standard operating procedure (SOP) for Prussian blue staining is adapted from published protocol, similar to that used in the UTVMC histology laboratory for the pilot study.[27] Initially slides were fixed for 10 minutes with methanol, incubated for 10 minutes with freshly prepared and equal parts of 2% potassium ferrocyanide and 2% hydrochloric acid, and finally counterstained by flooding with 1% safranin for 10 minutes. Between each step and after the counter stain, slides were rinsed with distilled water. After the final step slides were left to air dry and mounted with a standard glue and covered with a coverslip.

For all dogs, a concurrent blood sample was placed in EDTA purple top tube (for CBC with reticulocytes) and into a no-additive red top tube (for measurement of serum iron, TIBC, and percent saturation of transferrin (%Sat). Concurrent CBC data with reticulocyte parameters were analyzed for all dogs either via an ADVIA 2120 or 120 hematology analyzer (Siemens Healthcare Diagnostics Inc.,Tarrytown, NY, USA). The CBCs for each group were performed using different instruments, as follows: an ADVIA 2120 at CUHA for the CUHA patients, an ADVIA 2120 at UTVCM for the healthy research dogs, and an ADVIA 120 at UTVMC for the dogs with nutritional iron deficiency.

Serum iron, TIBC, and %Sat were measured on all UTCVM healthy research canines, all dogs with experimental iron deficiency, and most CUHA patients using the Cobas C501 chemistry analyzer (Roche Diagnostics, Indianapolis, IN, USA) at the CUHA clinical pathology laboratory.

For the larger set, the statistical software SPSS (IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp) was used for all data analysis. For intraclass correlation coefficient (ICC) analyses, published, qualitative, interpretative guidelines were used: <0.5 poor, 0.5 – 0.75 moderate, 0.75 – 0.9 good, and >0.9 excellent correlation.[25] The previous pilot study employed a Fleiss Kappa, which is based categorical ratings. Using the Fleiss' Kappa, given that one rater observed a Gale score of 4 and the other a score of 5, these categorically do not agree, and some power could be lost even when the weighted kappa (K_ω) was used.[22, 23] ICC utilizes a continuous scale, and accounts for the closeness of score of 4 and 5 between different observers. However, Fleiss' Kappa was more appropriate is some instances where variability was interpreted with fewer categories i.e. observer agreement at lower

iron Gale scores. Interpretation of Fleiss' kappa was based on K_ω values, which were as follows: <0 no agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1 near perfect or perfect agreement.[24] Rationale for this statistical analysis of samples in the lower half of the Gale score scale was to determine agreement between observers at scores around the clinical decision limits for detection of iron deficiency.

Interobserver Variability

ICC was used to determine interobserver reliability of all 134 samples. Additionally, the ICC for all directly Prussian blue stained smears from the pilot study and larger set smears was determined.

Assessment of Clinical Utility

All smears were assessed via Gale method by three observers. These included two, experienced, board certified veterinary clinical pathologists and one veterinary clinical pathology resident. All observers used the same model microscope (Olympus BX41, Olympus Corporation, Tokyo, Japan). As defined by the previous study, bone aspirates had at least 7 marrow particles that were moderately sized, defined as at least occupying ¾ of a low power field (10x objective) for particle standardization. Scoring was the same as our pilot study using the Gale scoring method based on categorical values of 0 to 6 (Figure 2).[5] Based on results of the pilot study, sideroblast scoring was not performed on the larger data set because of the significant observer effect.

Interobserver Variability at low iron scores

Fleiss' kappa was also used to determine interobserver agreement of the three observers only on those samples on the lower half of the Gale method scoring scale, i.e. a Gale score of 0-2, with a total number of 25 samples. Rationale for this statistical analysis of samples in the lower half of the Gale score scale was to determine agreement by a stricter method between observers at scores around the clinical decision limits for detection of iron deficiency.

Correlation of Gale Score and CHr

Spearman correlation on all 134 samples was used to investigate correlation between the Gale scores and CHr. For each sample, the mean Gale iron score for the three observers was utilized for comparison to the CHr. Interpretation of the Spearman correlation coefficient was based on strength (1 to 0.9- perfect to very high correlation, 0.9 to 0.7 high correlation, 0.5 to 0.7 moderate correlation, 0.5 to 0.3 low correlation, 0.3 to 0.0 negligible to no correlation) and direction (negative verses positive).[28]

Odds Ratio

Odd ratio was used to determine the odds of canines with a low CHr, defined as equal to or less than the lower reference limit established in the UTVMC clinical pathology laboratory (24.4 picograms), having a Gale score in the lower half of the grading scale, i.e. 0-2.[18, 19, 29] The published canine reference range for CHr of 24.5 to 28.6 pg was used.[2]

Healthy Canine Descriptive Statistics

Ideally, a sample size of at least 40 canines is needed to calculate reference intervals based on recommendations from the American Society for Veterinary Clinical Pathology (ASVCP), but reference interval estimates can be provided with as few as 20 samples. [30] Since bone marrow sampling is a painful and expensive procedure, 20 canines were targeted. However, since fewer than 20 samples could be included after applying the inclusion and exclusion criteria, results from this group are presented only as descriptive statistics per ASVCP guidelines. Descriptive statistics were used to determine the 12 healthy research canine Gale scores and CHr values, including minimum, median, maximum, mean, and standard deviation values.

Hematologically Normal, Clinically Ill Canines

The limitation of using healthy research dogs to define normal marrow iron stores by the Gale method is that these dogs do not likely reflect the population of sick canine patients that may have marrow sampled during clinical evaluation. In order to address this limitation, an additional comparison group of dogs expected to have adequate marrow iron stores was identified that included clinically ill dogs without any evidence of iron-restricted erythropoiesis. Two proposed ways were investigated for the definition of these hematological normal canines. The first definition was more strict, requiring all CBC and iron biomarker values to be within reference intervals. This included normal erythrocyte values (RBC, HGB, MCV, MCH, MCHC, absolute reticulocyte count, CHr), leukocyte values (WBC with absolute leukocyte counts (segmented neutrophils, band neutrophils, lymphocytes, monocytes, eosinophils, and basophils), platelet values, and serum biochemistry values (serum iron, TIBC, and percentage saturation). Additionally, patients were excluded from this group if there was medical history of a condition that could affect iron metabolism or erythropoiesis, including evidence of external bleeding noted in the medical history that could lead to decreased marrow iron, diagnosis of chronic kidney disease which could have affected erythropoiesis, evidence of portosystemic shunting which could alter iron metabolism, and lastly specific breeds such as Akita that have been documented to have a hereditary microcytosis.[2, 3, 15, 31, 32] There were 13 CUHA canines that met these strict criteria.

The less strict definition included a slightly larger set of 20 canines. Exclusion criteria based on CBC and serum biochemistry in this sample were less strict,

allowing inclusion if there were changes in WBC values or platelet values that were unlikely to be associated with changes in iron metabolism, such as a stress leukogram, mild thrombocytopenia, or thrombocytosis. Descriptive statistics were performed for the CUHA as defined by both the strict and less strict definition of hematologically normal but clinically ill canines. Values for both proposed sets investigated Gale scores and CHr values based on minimum, median, maximum, mean, standard deviation.

Differences in Gale scores

A Mann Whitney U test was performed on all of the mean Gale score from all observers between 1) healthy research dogs verses all ill canines, 2) hematologically normal ill dogs versus hematologically abnormal ill dogs, 3) canines with and without clinically reported blood loss, and 4) dogs on an experimentally iron-restricted diet versus healthy research dogs.

Difference in CHr

A Mann Whitney U test was performed on all of the mean Gale score from all observers between 1) healthy research dogs verses all ill canines, 2) hematologically normal ill dogs versus hematologically abnormal ill dogs, 3) canines with and without clinically reported blood loss, and 4) dogs on an experimentally iron-restricted diet versus healthy research dogs.

Concurrent Use of Gale and CHr

Utilization of the Gale score and concurrent CHr was used to identify dogs that have either overt iron deficiency or those that could have some other underlying pathology (e.g. covert blood loss) that might predispose to iron deficiency and warrant clinical investigation for causes of blood loss. Determination of which clinically-ill dogs had evidence of possible absolute iron deficiency was based on two categories that include Gale scoring and CHr (Table 2). Within each CHr/iron score category (1-2), the number of dogs that have evidence of each type of iron-restricted erythropoiesis, using the (Table 3) were identified. These categories included RBC parameters, iron values, and clinical information as diagnostic criteria.

CHAPTER FIVE

RESULTS AND DISCUSSION

Scoring Performance

Pilot Study

Staining

The Gale score was significantly higher for the directly stained smears when compared to the destained smear from the same patient (p=0.004). The differences were modest, averaging from 0.25 to 0.40 points higher in the non-destained smears depending on observer and patient.[5] The difference in sideroblast scores between destained and directly stained smears was not statistically significant (p=0.45).[5] Results from all control slides were acceptable, with liver and blood samples showing large iron deposits and small iron granules, respectively, in all staining runs.[5]

Scoring

There was not a statistically significant difference in iron scores between observers by the Gale method (p=0.21) (Figure 3).[5] Fleiss' kappa was utilized for interobserver agreement, with directly stained and destained smears analyzed separately since there was an effect of destaining on the Gale method. For directly stained smears there was fair agreement, with an ordinal measure $K_\omega = 0.23$ and a coefficient of concordance (CC) of 0.82. For destained smears there was also fair agreement, with an ordinal measure $K_\omega = 0.31$ and a CC of 0.81.[5] For directly stained smears there was fair agreement, with an ordinal measure $K_\omega = 0.23$ and a coefficient of concordance (CC) of 0.82. For destained smears there was also fair agreement, with an ordinal measure $K_\omega = 0.31$ and a CC of 0.81.[5] The ICC for interobserver reliability among the three observers for unstained smears was 0.930 CI (0.862, 0.967). This indicates excellent reliability of all three observers.[25] The ICC for interobserver reliability among the three observers for destained smears was 0.933 CI (0.868, 0.969). This indicates excellent reliability of all three observers.[25]

For the sideroblast method, the majority of sideroblasts had morphology typical of that described in human marrow samples, with 1-5 fine iron granules. Very rarely, possible ring sideroblasts were identified (Figure 1), but the maximum number identified in a sample during the sideroblast count was two.[5] There was a significant difference in scores between observers (p=0.002), with the scores from one observer (average of about 9-11% sideroblasts per sample) being slightly higher than the other two observers (average of about 4-5% sideroblasts per sample) (Figure 4).[5]

Results for Cohen's weighted kappa for the duplicate scores from each observer by the Gale method indicate substantial to perfect agreement (Table 4). Results for non-destained and destained smears were analyzed separately since there was an effect of destaining on the Gale method. Because the data for the sideroblast method was continuous, intraobserver agreement for the duplicate scores was evaluated using Pearson correlation and indicated moderate to strong positive correlation (Table 5).

Larger Set

Interobserver Variability

The ICC for interobserver reliability among the three observers for all 134 samples was 0.935 CI (0.914, 0.952). This indicates excellent correlation of all three observers.[25] For the pilot study directly stained smears and larger set smear the overall ICC was 0.938 CI (0.919, 0.953).

Assessment of Clinical Utility

There is a statistically significant difference in iron scores between observers for the 25 samples with lower Gale iron scores (score of ≤2 by at least one observer) (p=0.001). Fleiss' kappa was utilized for interobserver agreement. There was fair agreement, with an ordinal measure K_{ω} = 0.306, CI (0.162 to 0.450).[24] Within these 25 samples, all three observers gave the same grade in 8/25, while observer(s) disagreed on the grade in 17/25, with the following details:

- Grade 0 by all three observers, n=6/25
- Grade 1 by all three observers, n=0/25
- Grade 2 by all three observers, n=2/25
- Grade 0 given by at least one observer and Grade 1-3 given by other observer(s), n=10/25
- No observers gave a grade of 0, but observers gave varying scores ranging from 1-3, n=7/25

In summary observers have the best agreement on Gale scores of 0 and the worst agreement on Gale scores of 1 and 2. Additionally there was some disagreement of Gale score of 3 when other observer scored a 0-2, however it was not statistically significant when comparing all observers' scores (p=0.014).

Correlation of Gale Score and CHr

When testing the correlation between the CHr to the Gale score (mean of the 3 observers) in the 134 samples, a spearman correlation coefficient was 0.132, indicating a weak positive correlation that was not statistically significant (p=0.134).

Odds Ratio
The odds of a low CHr (equal or less than 24.4 pg) in a canine patient is 3.346 times higher in patients that have a low median Gale score (2 or less) compared to those with a higher median Gale score (3 or higher) (OR) = 3.346, Cl (1.204 to 9.303).

Descriptive Statistics
Descriptive statistics of Gale scores are provided for the healthy research dogs and the clinically-ill dogs that were defined as hematologically normal using a more strict (group A) and a less strict (group B) definition (Table 6).

Differences in Gale scores
There was statistically significant difference in Gale iron scores (mean of all observers) between healthy controls and ill canines (p=0.001) as tested by the Mann Whitney U test. There was no statistically significant difference in the mean Gale score of the strict definition (N=13) of hematologically normal verses hematologically abnormal canines (p=0.145). However, Gale score in clinically-ill dogs with the less strict definition (N=20) were significantly higher than clinically-ill and hematologically abnormal canines (p=0.019). There was not a statistically significant difference in the mean Gale score based on the clinical presentation of blood loss verses no blood loss (p=0.813). Dogs with experimentally induced nutritional iron deficiency had a significantly lower Gale score compared to healthy controls (p=0.003) as tested by the Mann Whitney U test.

Differences in CHr
There was not a statistically significant difference CHr of all observers in healthy controls verses ill canines (p=0.719) as tested by the Mann Whitney U test. There was no statistically significant difference in the CHr based the strict definition (N=13) of hematologically normal verses hematologically abnormal canines (p=0.214). There was not a statistically significant difference in the CHr based on the less strict definition (N=20) of hematologically normal verses hematologically abnormal canines (p=0.094). There was not a statistically significant difference in the CHr based on the clinical presentation of blood loss verses no blood loss (p=0.388). Dogs with experimentally induced nutritional iron deficiency had a significantly lower CHr compared to healthy controls (p=0.004) as tested by the Mann Whitney U test.

Concurrent Use of Gale and CHr
Based on criteria defined in Table 2 regarding CHr and marrow Gale score, there were 9 canines in Group 1 (possible absolute iron deficiency, i.e. low CHr and low Gale score) and 11 canines in Group 2 (possible subclinical absolute iron deficiency, i.e. normal CHr and low Gale score). All laboratory and clinical information from these patients were evaluated according to criteria in Table 3 to

determine how many of these dogs had evidence of different causes of iron-restricted erythropoiesis.

GROUP 1 (Low CHr and low Gale score, n=9)
One dog in Group 1 met the criteria for absolute iron deficiency associated with blood loss. This patient had microcytic and hypochromic anemia, low serum iron, low %Sat, with evidence of external blood loss. Blood loss was secondary to a small intestinal leiomyoma.

Three dogs in Group 1 met the criteria for absolute, nutritional iron deficiency. These patients had microcytic and hypochromic anemia, low serum iron, low %Sat, and had been on an experimental iron deficient diet for at least 77 days.

Two dogs in Group 1 had clinical evidence of predisposition to absolute iron deficiency (chronic external blood loss and/or chronic NSAID administration but with normal MCV, MCHC, serum iron, TIBC, and %Sat). One of these dogs had recurrent episodes of epistaxis for two months, which was likely due to hyperviscosity associated with a diagnosis of multiple myeloma. The other had immune-mediated thrombocytopenia with about a one-week history of melena, and had also been on the NSAID Metacam for osteoarthritis.

Two dogs in Group 1 were grouped in the iron sequestration category because they had neoplasia or inflammation. One these dogs had multicentric large cell lymphoma and was on thyroxine for hypothyroidism with a microcytic hypochromic mildly regenerative mild anemia, low serum iron, and low %Sat. The other dog had a pericardial mass, probable diaphragmatic hernia, normocytic normochromic non-regenerative mild anemia, low serum iron, and low %Sat.

No dogs in Group 1 were classified in the functional iron deficiency category which included dogs with hemolytic disease. One of the experimental nutritional deficient dogs was not sub classified into the nutritional deficient categories as all-inclusive categories were met except the anemia was normocytic hyperchromatic. This patient's hemolytic index was low therefore not likely a cause. Electrolytes and lipemic index were not reported therefore the cause of the aberrant normocytic hyperchromatic erythrocyte's findings were unclear.

GROUP 2 (normal CHr and low Gale score, n=11)
In Group 2, subclinical absolute iron deficiency category there were 8 canines with clinical or laboratory evidence of iron sequestration (inflammatory disease or neoplasia). Diagnoses included multicentric large cell lymphoma (2), hypereosinophilia of Rottweilers, severe stomatitis, mast cell tumor stage III, chronic pancreatitis with PIMA, and splenic or liver histiocytic sarcoma.

In Group 2, one dog was classified as functionally iron deficient. This patient was diagnosed with probable immune mediated hemolytic anemia and had low numbers of spherocytes and a weakly positive Coombs' test.

No dogs fit the criteria for the absolute iron deficiency with blood loss, absolute iron deficiency (functional), and predisposition to absolute iron deficiency. The remaining dogs not classified included 2 canines with an open diagnosis (possible PIMA) (1), and regenerative anemia of undetermined origin (1).

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

The pilot study evaluated the technical performance of two methods for scoring iron stores in canine bone marrow aspirate samples stained by Prussian blue, including assessments of interobserver agreement, intraobserver agreement, and the effect of destaining Wright stained smears before staining with Prussian blue. Our findings indicate that destaining prior to Prussian blue staining may cause a modest decrease in iron scores by the Gale method. It has been reported that the acids used in fixatives and decalcifiers in human histologic marrow samples may remove some iron stores.[21] Therefore, it is possible that the destaining procedure of acid and ethanol used in this study may have stripped some iron from the samples. Alternatively, the difference noted in our small sample set may reflect type I statistical error. In future studies, evaluation of destaining methods that do not employ an acid, such as a microwave technique, would be warranted.[33] Based on our results, performing Prussian blue staining on unstained marrow slides is advisable.[5]

The Gale method of marrow iron scoring grades large deposits of hemosiderin within macrophages on a scale of 0 (absent) to 6 (very heavy), and correlates moderately well with chemical measures of marrow iron in humans.[18] This method for iron scoring is very commonly used for evaluating marrow iron stores in humans, but to the authors' knowledge has not been evaluated for use in canine marrow samples.[6, 7] Results from our study indicate that Gale scoring has acceptable interobserver and intraobserver agreement, and differences between observers were not statistically significant.

The sideroblast method of iron scoring evaluates for small iron granules in nucleated erythroid precursors of any stage of maturation (sideroblasts).[7, 9] The sideroblast score is considered a better indicator of cellular iron availability that the Gale score.[7] However, the sideroblast score is more labor intensive and difficult. In our experience, it was difficult to identify faint iron granules in some cells and sometimes difficult to distinguish intracellular iron from superimposed extracellular iron granules. There was a significant difference in sideroblast scores between observers in our study, with one observer commonly counting about twice as many sideroblasts compared to the other observers.[5] This could be due to different visual sensitivities among the observers for detecting sideroblasts. Also, observers likely counted different 100 erythroid subsets, which may yield very different sideroblast percentages. This difference between observers suggests that it is unlikely that sideroblast iron scores would be highly reproducible by different pathologists, potentially limiting the clinical utility of this measurement.[5]

An additional notable finding was that the samples included in this study appear to have much lower sideroblast scores than what is considered normal for human marrow samples.[5] The dogs included in this study were clinically ill, so it is possible that they may have had iron sequestration or functional iron deficiency leading to decreased sideroblastic iron. Alternatively, there may be a species difference in normal amounts of sideroblastic iron. Assessment of sideroblast iron scores on additional canine samples would be useful, including samples from healthy dogs.

Limitations of this pilot study include that the sample size was small and that most samples fell in a limited range of Gale iron scores of 3-4. It would be useful to evaluate interobserver and intraobserver agreement of more samples that are at the most important clinical decision cutoff, which extrapolated from human marrow samples would be between 0 and 1 for the Gale score. Of note in our results, there was only one dog with Gale iron scores below 3, and scores for all observers in the destained and non-destained smears from this patient ranged from 0 to 2. Additionally, the method for defining a sideroblast was refined after the first set of observations due to the difficulty in distinguishing superimposed extracellular iron from sideroblast iron granules. This may have affected intraobserver variability of the sideroblast score between the first and second observations, although it should not have affected the interobserver variability for each replicate.[5]

Based on these results, the Gale method for iron scoring performs adequately, particularly when Prussian blue staining is performed on previously unstained smears.[5] However, the sideroblast iron scoring method should be used with caution due to significant interobserver variability.[5]

Interobserver variability of the larger set of 134 samples and in addition with the directly Prussian blue stained smear from the pilot study indicated excellent reliability among all three observers. Given the larger numbers of observations, this larger set has more robust power. This enforces that the Gale method is reproducible among observers. Additionally, reliability of the Gale method among observers was acceptable at clinically relevant low iron scores (2 or less) even though minor differences were seen between observers of Gale scores of 1 and 2 as seen from the Fleiss kappa. These differences would still classify the patient with low iron. There were rare instances where some observers scored a Gale score of 3 and others a Gale score of 0-2. This was not statistically significant; however, could result in some patients being misclassified as iron deficient and vice versa which could have a negative clinical impact. These differences could be attributable to different visual sensitivities of the observers, the locations of high-power scans, and length of time examining the smear.

CHr correlates weakly with Gale score, and it is not an acceptable indicator of marrow iron stores. This was an expected finding as different pathological processes other than absolute iron deficiency can alter the CHr. Most notably in iron sequestration with inflammatory processes, CHr is decreased and the Gale score maybe normal. Alternatively, the odds of a canine patient given a low CHr are approximately 3 times as likely to have a low Gale score. However, this is merely a measure of association and is not predictive of every patient. Therefore, CHr cannot be a surrogate for marrow iron stores and must be interpretated considering clinical findings and other blood work parameters.

Descriptive statistics for healthy "control" canines suggest that Gale scores of 4 to 5 and CHr of 24.5 pg to 26 pg would likely be normal. Given that the Gale range for healthy controls was 4 to 5 it may suggest that normal Gale scores for canines maybe higher compared to the normal Gale scores in humans i.e. 1 to 3.[18, 19] The CHr range of 24.5 pg to 26 pg seems to correlate with previous studies.[2, 3] The limitation of these values is that the number of observations is low which limits the robustness of the data. Additionally, since all patients are Beagles between the ages of 1- and 4-years old, variation is reduced. Therefore, this sample set is likely not applicable to a population of various breed and ages of canines. However, it is a starting place as a larger number of marrow aspirates from patients may be difficult to sample in a controlled setting. Also as mentioned obtaining marrow aspirates can be a very painful and expensive procedure to perform. It would also be ideal to have core biopsies to compare to the marrow aspirates. The core biopsies could provide and extra layer of enforcement of the Gale scoring, as distribution iron could be more even. However, to avoid repeated painful sampling a successful core was not taken on most of the healthy canines.

The strict (Group A) and less strict (Group B) definition of hematologically normal but clinically ill canines were similar with the less strict having slightly increased mean and median Gale scoring. Both the strict and less strict subcategories tended to mirror the mean and median Gale scores and CHr in the healthy control groups. However, for both the strict and less strict hematological normal but clinical ill canines the range of Gale scores were more variable than the healthy controls. Results from the Mann Whitney U test showed that Gale scores were higher with less strict (group B) canines compared to those hematologically abnormal canines. Therefore, by using the less strict criteria which allowed for changes in WBC values or platelet values, it is unlikely to be associated with alterations in iron metabolism, the patient was more likely to have a Gale score closer to the mean of the healthy control canines i.e. 4. Given the ranges of Gale scores (Table 3.4.), Gale score of <2 in a clinically ill dog or <4 in a healthy dog may warrant evaluation for a condition associated with iron restricted erythropoiesis, such as external blood loss, nutritional deficiency, or conditions

associated with iron sequestration or functional iron deficiency (e.g. inflammation, neoplasia, or rapid erythropoiesis).

The fact that dogs with clinical evidence of blood loss could not be distinguished from those dogs without clinical blood loss by the Gale score could be due to differences in type of bleeding. For example, internal bleeding into body cavities would allow for recycling of iron verses external to the body bleeding where iron would be lost. Additionally, a limitation is that the evaluation for blood loss was restricted to those with any amount of clinical evidence of bleeding documented in the medical record. Therefore, no objective measurement of the amount of external blood loss was available. Additionally, we cannot rule out that dogs without clinically reported blood loss could have had subclinical external blood loss.

Healthy control canines could be distinguished from the clinically ill canines by the Gale score but not the CHr. Both the Gale score and CHr were significantly lower in canines with experimental nutritional iron deficiency as compared to healthy controls. Dogs with low CHr and low Gale score in our data set commonly had evidence to support absolute iron deficiency. However, some dogs in this category had additional evidence to support concurrent iron sequestration or functional iron deficiency. None of the dogs with normal CHr and low Gale score in our data set had evidence to support absolute iron deficiency. Therefore, low Gale score may not be specific for absolute iron deficiency, and it might be best to interpret it in conjunction with CHr. A limitation, though, is that we cannot rule out that those dogs with normal CHr and low Gale score may have early or subclinical iron deficiency.

In summary, the Gale score is a reliable method to semi-quantitate canine iron marrow. Directly stained Prussian blue marrow aspirates are preferred. Descriptive statistics for healthy Beagles include Gale scores of 4 to 5 and CHr of 24.5 pg to 26 pg. The Gale score and CHr should be interpreted with all bloodwork and clinical data can ensure appropriate classification of iron restricted erythropoiesis as subclinical iron deficiency or more than one restrictive iron process may be present.

LIST OF REFERENCES

1. Fry, M.M. and C.A. Kirk, *Reticulocyte indices in a canine model of nutritional iron deficiency.* Vet Clin Pathol, 2006. **35**(2): p. 172-181.
2. Schaefer, D.M.W. and T. Stokol, *The utility of reticulocyte indices in distinguishing iron deficiency anemia from anemia of inflammatory disease, portosystemic shunting, and breed-associated microcytosis in dogs.* Vet Clin Pathol, 2015. **44**(1): p. 109-119.
3. Schaefer, D.M. and T. Stokol, *Retrospective study of reticulocyte indices as indicators of iron-restricted erythropoiesis in dogs with immune-mediated hemolytic anemia.* J Vet Diagn Invest, 2016. **28**(3): p. 304-8.
4. Melendez-Lazo, A., et al., *Evaluation of the Relationship between Selected Reticulocyte Parameters and Inflammation determined by Plasma C-reactive Protein in Dogs.* J Comp Pathol, 2015. **152**(4): p. 304-12.
5. Pawsat, G.A., et al., *Comparison of iron staining and scoring methods on canine bone marrow aspirates.* Veterinary clinical pathology, 2021. **50**(1): p. 132-141.
6. Hughes, D.A., S.E. Stuart-Smith, and B.J. Bain, *How should stainable iron in bone marrow films be assessed?* Journal of clinical pathology, 2004. **57**(10): p. 1038-1040.
7. Phiri, K.S., et al., *Improved method for assessing iron stores in the bone marrow.* J Clin Pathol, 2009. **62**(8): p. 685-9.
8. Archer, N.M. and C. Brugnara, *Diagnosis of iron-deficient states.* Crit Rev Clin Lab Sci, 2015. **52**(5): p. 256-72.
9. Smock, K.J., *Examination of the blood and bone marrow*, in *Wintrobe's Clinical Hematology*, J.P. Greer, et al., Editors. 2019, Wolters Kluwer: Philadelphia, PA.
10. Harvey, J.W., *Veterinary Hematology: A Diagnostic Guide and Color Atlas*. 2011, St. Louis: Elsevier.
11. Fuchs, J., et al., *Canine reticulocyte hemoglobin content (RET-He) in different types of iron-deficient erythropoiesis.* Vet Clin Pathol, 2017. **46**(3): p. 422-429.
12. Fuchs, J., et al., *Evaluation of reticulocyte hemoglobin content (RET-He) in the diagnosis of iron-deficient erythropoiesis in dogs.* Vet Clin Pathol, 2017. **46**(4): p. 558-568.
13. Mast, A.E., et al., *Clinical utility of the reticulocyte hemoglobin content in the diagnosis of iron deficiency.* Blood, 2002. **99**(4): p. 1489-91.
14. Latimer, K.S., *Duncan and Prasse's Veterinary Laboratory Medicine: Clinical Pathology*. 2011, Hoboken: John Wiley & Sons, Incorporated.
15. Weiss, G., T. Ganz, and L.T. Goodnough, *Anemia of inflammation.* Blood, 2019. **133**(1): p. 40-50.
16. Cote, E., *Clinical veterinary advisor.: (Dogs and cats)*. 2014: Mosby.
17. Foy, D.S., K.R. Friedrichs, and J.F. Bach, *Evaluation of Iron Deficiency Using Reticulocyte Indices in Dogs Enrolled in a Blood Donor Program.* J Vet Intern Med, 2015. **29**(5): p. 1376-80.

18. Gale, E., J. Torrance, and T. Bothwell, *The quantitative estimation of total iron stores in human bone marrow.* J Clin Invest, 1963. **42**: p. 1076-82.
19. Bain, B.J., et al., *Special techniques applicable to bone marrow diagnosis*, in *Bone Marrow Pathology*, B.J. Bain, et al., Editors. 2001, Blackwell Science: Oxford, UK. p. 51-89.
20. Stockham, S.L. and M.A. Scott, *Fundamentals of veterinary clinical pathology, Second edition*. 2nd ed. ed. 2013: Wiley-Blackwell.
21. Travlos, G.S., *Normal structure, function, and histology of the bone marrow.* Toxicol Pathol, 2006. **34**(5): p. 548-65.
22. McHugh, M.L., *Interrater reliability: the kappa statistic.* Biochemia medica, 2012. **22**(3): p. 276-282.
23. Watson, P.F. and A. Petrie, *Method agreement analysis: A review of correct methodology.* Theriogenology, 2010. **73**(9): p. 1167-1179.
24. Landis, J.R. and G.G. Koch, *The Measurement of Observer Agreement for Categorical Data.* Biometrics, 1977. **33**(1): p. 159-174.
25. Liljequist, D., B. Elfving, and K. Skavberg Roaldsen, *Intraclass correlation – A discussion and demonstration of basic features.* PloS one, 2019. **14**(7): p. e0219854-e0219854.
26. Gaddis, M.L. and G.M. Gaddis, *Introduction to biostatistics: Part 6, Correlation and regression.* Ann Emerg Med, 1990. **19**(12): p. 1462-8.
27. Orchard, G.E., *Pigments and Minerals*, in *Bancroft's Theory and Practice of Histological Techniques*, S.K. Suvarna, C. Layton, and J.D. Bancroft, Editors. 2013, Churchill, Livingstone, Elsevier: United Kingdom. p. 239-270.
28. Hinkle, D.E., W. Wiersma, and S.G. Jurs, *Applied statistics for the behavioral sciences*. Vol. 663. 2003: Houghton Mifflin College Division.
29. Steinberg, J.D. and C.S. Olver, *Hematologic and biochemical abnormalities indicating iron deficiency are associated with decreased reticulocyte hemoglobin content (CHr) and reticulocyte volume (rMCV) in dogs.* Veterinary clinical pathology / American Society for Veterinary Clinical Pathology, 2005. **34**(1): p. 23-27.
30. Friedrichs, K.R., et al., *ASVCP reference interval guidelines: determination of de novo reference intervals in veterinary species and other related topics.* Veterinary clinical pathology / American Society for Veterinary Clinical Pathology, 2012. **41**(4): p. 441-453.
31. Simpson, K.W., et al., *Iron status and erythrocyte volume in dogs with congenital portosystemic vascular anomalies.* Journal of veterinary internal medicine / American College of Veterinary Internal Medicine, 1997. **11**(1): p. 14-19.
32. Beutler, E., *Chapter 40, Disorders of iron metabolism*, in *Williams Hematology*, M.A. Lichtman, et al., Editors. 2006, McGraw-Hill: New York. p. 511-553.

33. Marcos, R., et al., *Use of destained cytology slides for the application of routine special stains.* Veterinary clinical pathology, 2009. **38**(1): p. 94-102.

APPENDIX

Table 1 Gale scoring table

Score	Summary	Description
0	None	No visible iron, even under high power (oil) magnification
1	Very slight	Any small intracellular iron particles only visible under high power magnification (50x or 100x objective)
2	Slight	Small, sparsely distributed iron particles just visible under low power magnification (10x objective)
3	Moderate	Numerous small intracellular iron particles throughout sample
4	Moderately heavy	Larger iron particles throughout sample with tendency to aggregate in clumps
5	Heavy	Dense, large clumps of iron throughout the sample
6	Very heavy	Very large deposits of iron obscuring cellular detail in sample

Table 2 Absolute iron deficiency

Categorization of probability of dogs having absolute iron deficiency, based on CHr and Gale score.

	CHr	Gale Score	Hypothetical causes
Group 1	<24.4 pg	**≤ 2***	Probably has absolute iron deficiency
Group 2	24.5-28.6 pg	**≤ 2***	Suspicious for mild or incipient absolute iron deficiency

*≤ 2 indicating a Gale score that was less than that of dogs that were clinically ill but hematologically normal.

Table 3 Iron deficiency categories

	Diagnostic Criteria
Absolute Iron Deficiency, blood loss	Microcytic and hypochromic anemia, decreased serum iron, decreased %Sat, AND external blood loss reported in the medical record
Absolute Iron Deficiency, Nutritional	Microcytic and hypochromic anemia, decreased serum iron, decreased %Sat, AND experimental nutritional iron deficiency (only those on the iron deficient diet for at least one month)
Predisposition to Absolute Iron Deficiency	Clinical evidence of chronic external blood loss and/or chronic NSAID administration but with normal MCV, MCHC, serum iron, TIBC, and/or %Sat
Iron Sequestration	Clinical or laboratory evidence of inflammatory disease or neoplasia
Functional Iron Deficiency	Laboratory evidence of hemolytic disease, e.g. spherocytes, RBC ghosts, agglutination or positive Coombs' test, more than rare Heinz bodies or eccentrocytes, RBC parasites, hemoglobinuria (positive heme w no sediment RBCs and normal serum CK), or high indirect bilirubin without other evidence of cholestasis
None of the Above	Does not meet any of the above diagnostic criteria. Tabulate and describe conditions included in this category.

Table 4 Gale variation

Intraobserver variation for duplicate Gale scores on Prussian blue stained canine marrow aspirates, by Cohen's weighted kappa. Non-destained slides were stained initially with Prussian blue stain, while destained slides were previously stained with Wright stain then decolorized before Prussian blue staining. Agreement is based upon Landis and Koch.[8]

Observer	**Stain**	**Kappa**	**Agreement**
1	Non-destained	1.00	Perfect
1	Destained	0.78	Substantial
2	Non-destained	0.71	Substantial
2	Destained	1.00	Perfect
3	Non-destained	0.68	Substantial
3	Destained	0.74	Substantial

Table 5 Sideroblast variation

Intraobserver variation for duplicate sideroblast scores on Prussian blue stained canine marrow aspirates, by Pearson correlation. Non-destained slides were stained initially with Prussian blue stain, while destained slides were previously stained with Wright stain then decolorized before Prussian blue staining. Degree of correlation is based upon Gaddis and Gaddis[23]				
Observer	**Stain**	**P value**	**Rho value**	**Degree of Correlation**
1	Non-destained	0.004	0.76	Strong positive
1	Destained	<0.0001	0.92	Strong positive
2	Non-destained	0.0003	0.87	Strong positive
2	Destained	0.0012	0.82	Strong positive
3	Non-destained	0.05	0.57	Moderate positive
3	Destained	0.05	0.57	Moderate positive

Table 6 Descriptive statistics

	Healthy Research	**Clinically-ill but hematologically normal (Group A)**	**Clinically-ill but hematologically normal (Group B)**
Sample size	12	13	20
Range of Gale scores	4-5	2.67-5.0	2.67-5.67
Mean Gale score	4.3	3.9	4.03
Median Gale score	4.17	4.00	4.10
Standard Deviation of mean	0.3	0.83	0.8

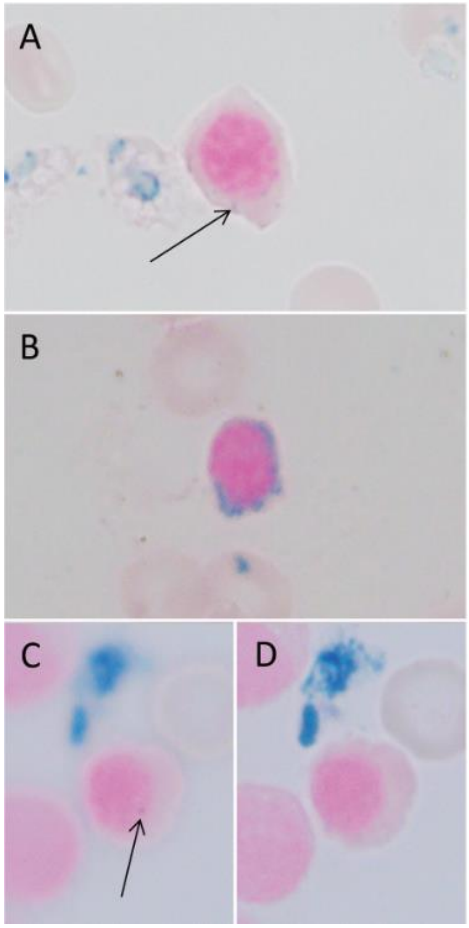

Figure 1 Sideroblasts

Sideroblasts in canine marrow aspirate smears. A) A typical sideroblast containing several very fine blue cytoplasmic iron granules, one of which is indicated by the arrow. B) an apparent ring sideroblast, defined as an erythroid precursor with greater than five cytoplasmic iron granules that encircle at least two-thirds of the nucleus. C-D) an erythroid precursor with a single small granule indicated by the arrow that is outside of the same focal plane as the surrounding cells, so likely represents a superimposed extracellular iron granule. This cell would not be included in the sideroblast count. 100x objective, Prussian blue stain.

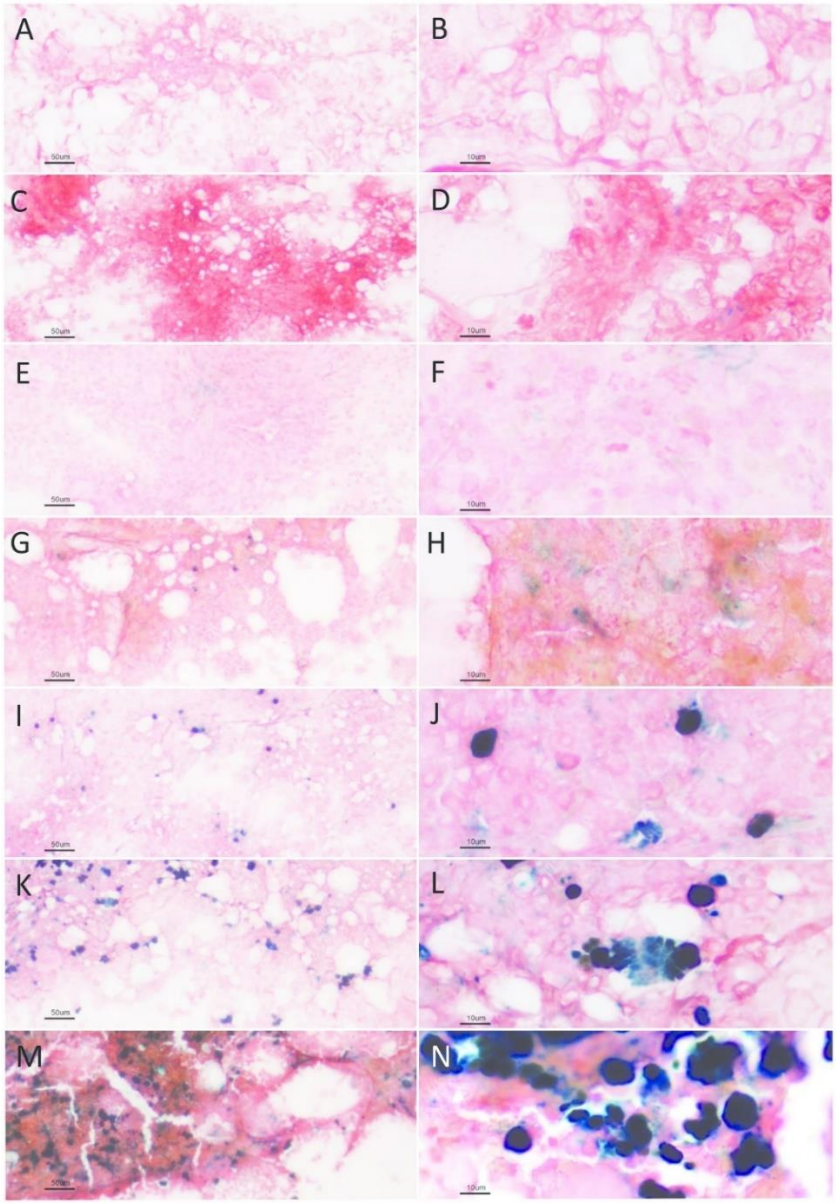

Figure 2 Gale scoring images

Gale marrow scoring images. A-B) Score 0. C-D) Score 1. E-F) Score 2. G-H) Score 3, example one. I-J) Score 3, example two. K-L) Score 4. M-N) Score 5. Gale scores of 0-2 required a high power scan at the 100x objective. An appropriate example of a grade 6 iron score was not available within the available archived samples of canine marrow aspirates. The images in the left column are low-power views using a 10x objective. The images in the right column are high-power views using a 50x objective. Prussian blue stain.

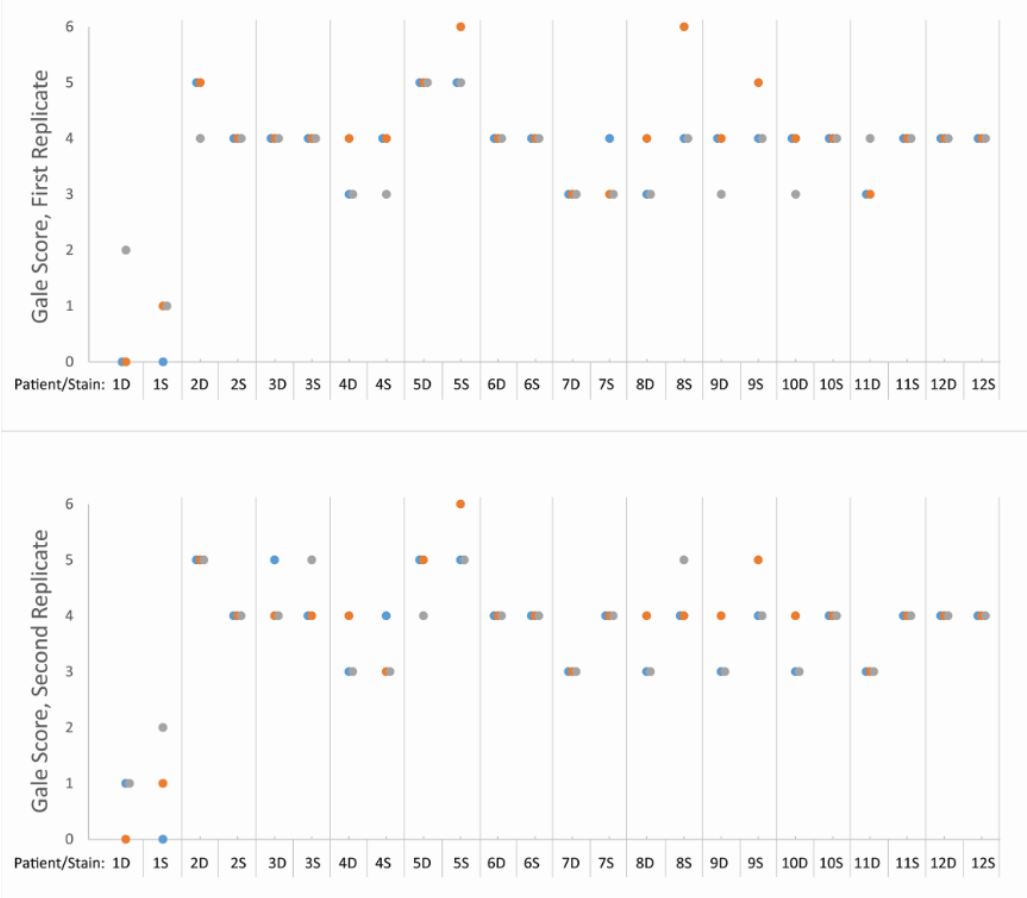

Figure 3 Gale variation

Interobserver variation in replicate Gale iron scores on marrow aspirate smears from 12 dogs. Two smears were included for each dog, one that was previously stained with Wright stain, then decolorized before Prussian blue staining (D, destained) and one that was stained directly by Prussian blue (N, non-destained). Tall vertical lines were placed between patients for ease of viewing separate patient results. Different colored markers indicate results for each of three different observers.

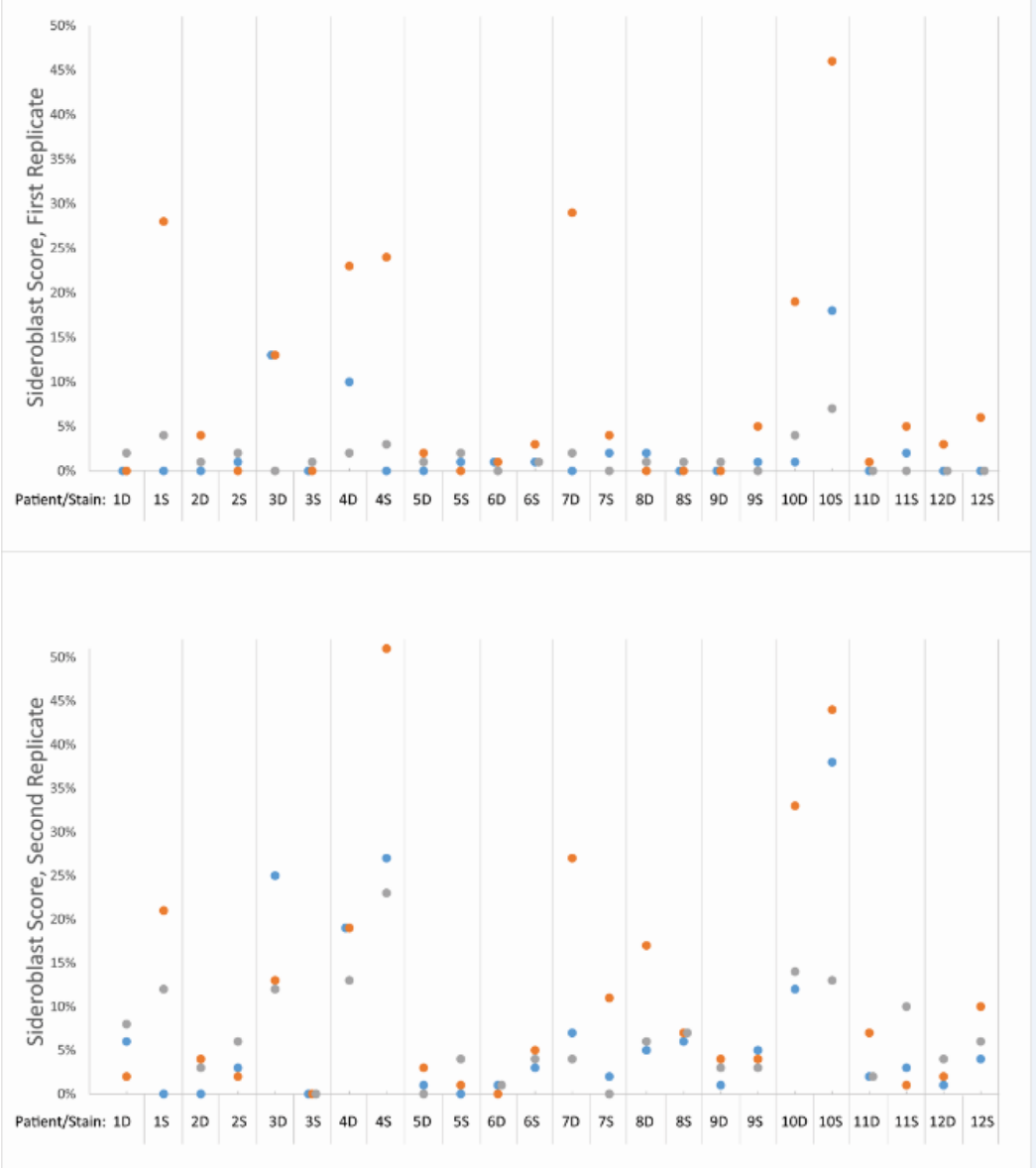

Figure 4 Sideroblast variation

Interobserver variation in replicate sideroblast iron scores on marrow aspirate smears from 12 dogs. Two smears were included for each dog, one that was previously stained with Wright stain, then decolorized before Prussian blue staining (D, destained) and one that was stained directly by Prussian blue (N, non-destained). Tall vertical lines were placed between patients for ease of

viewing separate patient results. Different colored markers indicate results for each of three different observers.

VITA

Grace Anne Pawsat is from Kentucky. She graduated from the College of Veterinary Medicine at Tuskegee University, Alabama in 2013. She went on to practice for five years in general practice in Kentucky. She returned to Tuskegee University for an internship in veterinary clinical pathology. From there she accepted a veterinary clinical pathology residency at University of Tennessee. In October 2021, she received Diplomate status from the American College of Veterinary Pathologists. She wrote and defended her thesis “Investigation of semiquantitative marrow iron scoring and reticulocyte hemoglobin content (CHr) in healthy and ill canines with and without clinical evidence of blood loss.”