Title: Computer vision quantitation of erythrocyte shape abnormalities provides diagnostic, prognostic, and mechanistic insight

Supplemental Materials 1.0
Supplemental Figures and Supplemental Tables

Brody H. Foy[1,2,3,+], Jonathan A. Stefely[1,4+], Pavan K. Bendapudi[4,5], Robert P. Hasserjian[1], Hanny Al-Samkari[5], Abner Louissaint Jr.[1], Megan Fitzpatrick[1], Bailey Hutchison[1], Christopher Mow[1,6], Julia Collins[1], Hasmukh R. Patel[1,2], Chhaya H. Patel[1,2], Nikita Patel[1,2], Samantha N. Ho[1,2], Richard M. Kaufman[7], Walter H. Dzik[1,4], John M. Higgins[1,2,3,#], Robert S. Makar[1,4,#]

[1]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, USA
[2]Center for Systems Biology, Massachusetts General Hospital, Harvard Medical School, Boston, USA
[3]Systems Biology Department, Harvard Medical School, Boston, USA
[4]Blood Transfusion Service, Massachusetts General Hospital, Harvard Medical School, Boston, USA
[5]Division of Hematology, Massachusetts General Hospital, Harvard Medical School, Boston, USA
[6]Partners Healthcare Enterprise Research Information Systems, Boston, USA
[7]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

+ These co-first authors contributed equally: Brody H. Foy, Jonathan A. Stefely

These co-corresponding authors jointly supervised this work: John M. Higgins (higgins.john@mgh.harvard.edu), and Robert S. Makar (rmakar@mgh.harvard.edu).

Contents

Supplemental Methods

Clinical smear collection guidelines. At Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH), peripheral blood smears are regularly collected as part of routine clinical care. Smears can be manually ordered by a clinician, but are most often automatically ordered by reflex, following detection of abnormalities in complete blood count or white cell differential markers, or in the flow cytometry distributions underlying these tests. The full list of reflex triggers are typically long, and vary by make and model of the hematology analyzer. The five most common triggers across the cohorts in this study were an elevated presence of immature granulocytes, a detected left-shift in the neutrophil population, monocytosis, elevated presence of blasts or abnormal lymphocytes, or suspected anemia.

Morphology grading flag reporting guidelines. In the MGH clinical hematology laboratory, guidelines in 2022 for semi-quantitative reporting are as follows: (i) For schistocytes and teardrop cells: 1+, 1–5 per high power (100X, oil immersion) field (HPF); 2+, 6–10 per HPF; 3+, >10 per HPF. (ii) Sickle cells, elliptocytes, and spiculated cells (acanthocytes or echinocytes) are recorded as "Present" *(recorded in the dataset as "1+")*. Of note, the laboratory guidelines also indicate that helmet cells, triangulocytes, and horn cells should be counted and reported as schistocytes (these are considered as schistocyte subtypes). In the BWH clinical hematology laboratory, guidelines for semi-quantitative reporting of sickle cells, schistocytes, teardrops, elliptocytes/ovalocytes, and spiculated cells (acanthocytes or burr cells/echinocytes) are as follows: 1+, 1–4 per HPF (defined as a 100x oil immersion field); 2+, 5–10 per HPF; 3+ >10 per HPF. Of note, in practice, there is reported variability between technicians in how HPFs are used. Some technicians use a 100x oil immersion field as an HPF, while others use one-quarter of a 50x oil immersion field as an HPF.

Collection of clinical data. Using the Partners Healthcare Electronic Data Warehouse and Research Patient Data Registry under an approved Institutional Review Board protocol, demographics, hospital admissions, diagnosis history, medications, procedures, laboratory test results, and mortality data were collected for all patients with available blood smears. Comorbidities were defined as any presence of relevant ICD9 or ICD10 codes in the year prior to the blood smear.

Clinical cohort studies. Differentials were analyzed across the following cohorts:

- *Elliptocytosis v. spherocytosis (**Figure 2A**).* All MGH patients with a listed diagnosis of hereditary spherocytosis (n = 17) or elliptocytosis (n = 6) were included. The first available smear for each patient was analyzed.
- *Pre- and post-liver transplant (**Figure 2B**).* All patients who underwent a liver transplant at MGH, who had an available blood smear in the 6 months prior to transplant, and between 1–6 months post-transplant (n = 46). If multiple smears were available, the smears closest to date of surgery (pre-op) and closest to 4 months post-op were taken.
- *Pre- and post-RBC exchange in sickle cell patients (**Figure 2C**).* All BWH patients with a diagnosis of sickle cell disease, who had blood smears available in the 48 hrs pre- and post-RBC exchange. Patients with a pre-exchange sickle cell count below 0.1% were excluded, and repeat measurements were used if a patient had multiple RBC exchanges, provided they were >1 month apart (n = 9).
- *Pre- and post-iron supplementation (**Figure 2D**)*. All MGH patients with low ferritin (<20 ng/mL) that normalized (>50 ng/mL) within 1–12 months, where both ferritin measurements were

recorded within 72 hours of a blood smear (n = 53). Patients were included if they had received intravenous iron supplementation at any point between the two ferritin tests (n = 30).

- *Pre- and post-splenectomy (**Figure 2E**)*. All MGH patients who underwent a splenectomy, and who had an available blood smear in the year prior to surgery, and between 1–12 months post-surgery (n = 22). If multiple smears were available, the smear closest to date of surgery (for pre-op) and closest to 4 months post-op were taken.

Characteristics of the clinical cohort studies are given in **Supplemental Table 2**.

Thrombotic microangiopathy (TMA) cohort details. For the derivation TMA cohort, the hospital site was MGH, and the date range was 31 March 2017 – 30 November 2020. Cases with clinical concern for TMA were identified by orders for ADAMTS13 enzyme activity. To capture a new cohort, patients already included in a previously published TMA cohort[47] were excluded. No clinical or laboratory exclusion criteria were applied. Comprehensive chart reviews were conducted essentially as described[47]. All cases with peripheral smear data available within five days of the ADAMTS13 test were carried forward in the analysis. The most likely clinical diagnoses and TMA diagnostic subgroups (e.g., iTTP) were determined by comprehensive chart review conducted by a physician, considering clinical and laboratory data. Clinical information was gathered by reviewing clinical notes with an emphasis on the admission history and physical note, hematology consult notes, and discharge notes. Immune thrombotic thrombocytopenic purpura (iTTP) was defined by an ADAMTS13 enzyme activity level $\leq$ 10% or $\leq$ 25% with an inhibitor of >1.0 BU. Outpatient cases of Upshaw-Schulman Syndrome were not defined as iTTP cases. Hemolytic uremic syndrome (HUS) was defined by TMA and ADAMTS13 activity > 10% plus at least one of the following: (i) serum creatinine > 2 mg/dL, (ii) stool culture positive for a shiga toxin-producing organism, or (iii) mutations identified in complement regulatory proteins.

All recorded parameters were required to be within five days of the ADAMTS13 test and related to the current presentation. Cases were classified as: (i) “presentation” if the ADAMTS13 test was sent on hospital day 0 (the day of admission), 1, or 2, (ii) “inpatient” if the ADAMTS13 test was sent on hospital day 3 or later, or (iii) “outpatient” if the ADAMTS13 test was sent for an outpatient. For presentation cases, the admission (earliest drawn) labs were preferentially recorded. For inpatient cases, laboratory data from the day of the ADAMTS13 test were preferentially recorded (using the earliest morning lab if multiple labs were drawn on the same date). Hospital length of stay was based on the time since the ADAMTS13 order, not the admission date. A history of transplant was recorded for either solid organ transplants or hematopoietic stem cell transplants.

Disseminated intravascular coagulation (DIC) was defined by laboratory evidence of overt DIC (e.g., coagulation parameters, D-dimer, fibrinogen) with a clear precipitant, including sepsis and multiorgan failure, cancer, or pancreatitis. Drug-associated TMA was defined by TMA and a temporal association with a drug previously reported to be associated with TMA (e.g., tacrolimus) without another clear cause of the TMA. Transplant-related TMA was defined by TMA in the setting of solid organ or hematopoietic stem cell transplant without another clear cause of the TMA. Pregnancy-associated TMA was defined by TMA in the setting of pregnancy without another clear cause of the TMA. Autoimmune TMA was defined as TMA in a patient with a known autoimmune disease without another clear cause of the TMA. Cancer-associated TMA was defined as TMA in a patient with active cancer (either solid organ or hematopoietic) without another clear cause of the TMA. The “other” subgroup included cases that did not fit into the above categories and did not have enough numbers to be an informative sub-group

in this study (e.g., a single case of hypertensive emergency, a single case of heparin-induced thrombocytopenia, etc.).

Characteristics of the TMA cohort are given in **Supplemental Table 3.**

Matched cohort analysis. Relationships between RBC-diff counts and all-cause mortality were estimated through matched cohort analysis. Using each patient's first available smear, for a given cell type, each patient with a corresponding count < 0.5% was matched to a patient of the same gender, race, comorbidity profile (anemia, cancer, liver disease, renal disease, sepsis, and transplant history; see clinical data collection section for comorbidity definitions), age (<5-year gap), hematocrit (<10% gap), similar morphology counts (<0.5% gap), and morphology grading flags, but with the given cell type count either between 0.5–1% or >1%. Patients were allowed to match to only one other patient, and if multiple matches were available, matching was chosen to minimize the distance between cell counts, ages, and hematocrit, with age and hematocrit divided by 10 to weight the two contributions more evenly. For schistocytes, results were also generated after matching on RDW levels (<2%) (**Supplemental Figure 9**). For spiculated cells, patients were also matched on delay time (<2hr gap) to account for potential storage effects (**Supplemental Figure 14**). Patients without a valid match were excluded. Mortality differences were analyzed using Kaplan-Meier curves, and the log-rank test. Mortality data was censored at either 1-year post smear or Nov 15th, 2021 (for patients whose first smear was later than Nov 15th, 2020). An equivalent approach was implemented at BWH, though only the flag of the relevant type was matched due to the smaller cohort size. This analysis was not performed for sickle cells, due to the substantially lower prevalence of positive flags. For the low and high schistocyte MGH cohorts, primary causes of death were evaluated through clinical chart review of 100 patients randomly selected from each group. Cohort characteristics are given in **Supplemental Tables 4–6**.

RBC-diff algorithm details. The RBC-diff is designed to calculate the relative abundances of 9 types of RBC morphologies (normal RBCs, elliptocytes, microcytes, macrocytes, schistocytes, sickle cells, spiculated cells, teardrop cells, and other abnormal RBCs) using a peripheral blood smear image. Algorithm development was primarily performed using images generated by the CellaVision™ system, consisting of ~3000x3000 pixel images (1 pixel ~0.2 micrometers width and length), containing ~1000-3000 cells. The algorithm was also evaluated using manually photographed images (**Supplemental Figure 4**) at a different resolution (~25% larger). The algorithm continued to operate normally with images that were 3x smaller and 3x larger than the typical resolution (after image resizing). However, robust testing across wider resolution ranges was not undertaken. It should be expected that there is a minimum image resolution at which red cell features can meaningfully be distinguished, either by human eye or algorithm.

Once collected, the peripheral smear image is processed as follows. The image is first converted to grayscale, and then binarized using a global black-white threshold calculated using Otsu's method. Following this, all polygonal objects within the image are identified using black-white boundary detection (via MATLAB function *bwboundaries*). For each potential RBC, 10 geometric features are then calculated and used to classify potential RBCs.

- Area: The size (pixels^2) enclosed by the shape boundary.
- Perimeter: The length (in pixels) of the shape boundary.

- Circularity: The normalized ratio of area to perimeter. $Circularity = \frac{2\sqrt{\pi \cdot Area}}{Perimeter}$. Circularity is always between 0 and 1, with 1 indicating a perfect circle, and 0 a line.
- Spiculation: The ratio of the shape's perimeter to the perimeter of the corresponding convex hull. A high value indicates the shape has a rough boundary. Convex hulls were calculated using MATLAB function *convhull*.
- Thickness: The minimum distance between any pair of opposing points on the shape boundary. In an ordered list of boundary points, the opposing point to a point *i* was defined as $opp_i = mod(i + \frac{n}{2}, n)$ where *n* is the total number of boundary points.
- Curvature: This property was specifically designed to assist with detection of image artefacts, such as platelet adhesions to an RBC. Curvature is the minimum ratio between the Euclidean distance between two boundary points, and their list distance (how many points are along the boundary between them). An extremely low curvature value means that some points that are extremely close physically, are quite far away from each other along the boundary, suggesting a finely joined protrusion. $Curvature = \min_{i,j \in n \;\&\; i \neq j} \frac{\|\overrightarrow{x_i} - \overrightarrow{x_j}\|}{\min\left(mod(i-j,N), mod(j-i,N)\right)}$.
- Convexity: The ratio of the shape's area to the area of the convex hull.
- Convex circularity: The circularity of the shape's convex hull.
- Elliptic error: The root mean squared error (in pixels) from the shape boundary to the least-squares fit of a general rotated ellipse of form: $\left(\frac{x\cos\theta + y\sin\theta}{a}\right)^2 + \left(\frac{x\sin\theta - y\sin\theta}{b}\right)^2 = 1$. The rotated ellipse is fit as the least-squares minimizer of a general conic equation ($Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$), with non-elliptic conic solutions resulting in an elliptic error of NaN (leading to the object being classified as a non-RBC object).
- Convex elliptic error: The root mean squared error from the convex hull boundary to the least-squares fit of a general rotated ellipse.

Using these 10 geometric features, potential RBC shapes were classified using the following procedure:

1. Small (< 400 pixels^2) and large (>1700 pixels^2) objects were discarded due to likely being white blood cells, RBC agglutinations, platelets, or other small smear artefacts.
2. Moderately large (>1200 pixels^2) objects with convexity < 0.95 were discarded due to likely being RBC agglutinations or WBCs.
3. Objects with both convexity > 0.94 and circularity > 0.74 were classified as normal RBCs.
4. Objects with extreme staining (mean red channel > 220 and mean green value outside of 100–200) were discarded due to likely being WBCs or artefacts. Typically, <2–3 objects were discarded this way per smear.
5. All remaining objects were classified as either normal RBCs, elliptocytes, schistocytes, sickle cells, spiculated cells, teardrop cells, other RBCs, or not RBCs, using a multi-class support-vector machine (SVM) classifier.
6. Taking all normal RBCs, any objects with an area below 770 pixels^2 (approximately 6.2-micron diameter, 75fL volume) were classified as microcytes, and any objects with an area above 1450 pixels^2 (approximately 8.5-micron diameter, 140 fL volume) were classified as macrocytes.

Note that for computational efficiency, geometric properties not used in steps 1–3 were only calculated after these steps were completed. Thresholds in steps 1–3 were manually calibrated using a set of 10

random smears from the MGH database (though distinct from SVM testing/training sets). 2000 single-object images were manually reviewed to estimate accuracy. Of 1000 objects excluded in steps 1-2, 13/1000 (1.3%) were determined to actually be RBCs; and of 1000 objects classified as normal RBCs in step 3, 3/1000 (0.3%) were found to be abnormal (2 schistocytes, 1 teardrop).

The SVM model training set comprised sets of 20 randomly chosen MGH smears with 1+ flags for elliptocytes, schistocytes, sickle cells, spiculated cells, or teardrop cells. For each of these 100 images, steps 1–3 of the above procedure were first run (to enrich the data set for abnormal RBCs), and 50 individual cells were chosen randomly from the remaining unclassified cells. This set of 5000 cells were then independently classified by two research fellows in MGH's pathology department (without knowledge of the associated morphology grading flags). Inter-classifier agreement was high (Cohen's $\kappa^1 = 0.77$), and discordant classifications were jointly reviewed and reclassified by consensus. A series of 7 individual single-class SVM models (for the 8 classification groups, excluding "other RBCs") was then trained using 3667 of the 5000 images, holding the other 1333 images as a test set. The models were trained using all 10 geometric features, after Z-normalization using the mean and standard deviation of each feature across the training set, and with hyperparameters optimized using MATLAB function *fitcsvm*. Model operating points were set at 90% specificity (estimated from the training set). For final classifications, each object was assigned to the class with the highest SVM score, provided at least 1 score was above the operating point. If no scores were above the operating point, the object was classified as an "other RBC". To generate the RBC-diff counts, every object in the smear image was classified, and relative abundances calculated. Absolute cell counts were calculated by multiplying relative counts by the total RBC count from the associated complete blood count panel.

The choice to use multiple binary classification SVM models instead of a multi-class SVM was made to allow for easier manipulation of sensitivity and specificity thresholds for each class (a potentially desirable trait in clinical applications). All models in this manuscript have used operating points at 90% specificity (in binary encoding), but we felt that retaining this easy adjustability was a desirable feature.

Updated training set. After initial model training, a set of another 5000 single-cell images were collected and manually labelled. These images were drawn randomly from the abnormal cells (those passing steps 1-3 of the algorithm) of 500 patients (10 cells per patients). The 500 patients comprised five groups of 100 patients with 1+ flags for the major abnormal cell types. Manual labels were generated by the same two research fellows as the initial image set, following the same procedure as described previously. The RBC-diff was then retrained using the initial training set, and 3000 images from the new training set (holding out 2000 for testing), with each patient only appearing in either the training or testing set. Comparative to the original model, the increased training set generated only modest accuracy improvements (AUC increases of 0.01, 0.03, 0.03, 0.01, and 0 for elliptocytes, schistocytes, sickle, spiculated, and teardrops respectively), and did not lead to improved performance at the cell population level. Given this, results in the manuscript are generated solely using the original training set. However, as a resource, we provide both image training sets in **Supplemental Data 1**.

Comparison to manually collected images. To test performance separate from the CellaVision™ system, 20 randomly chosen blood smear slides were also imaged manually through a 50x microscope, using an iPhone 12 with a 12 megapixel camera. For each slide 10 images of different high-powered fields (each containing ~500 red cells) were collected, and average cell densities estimates were compared to

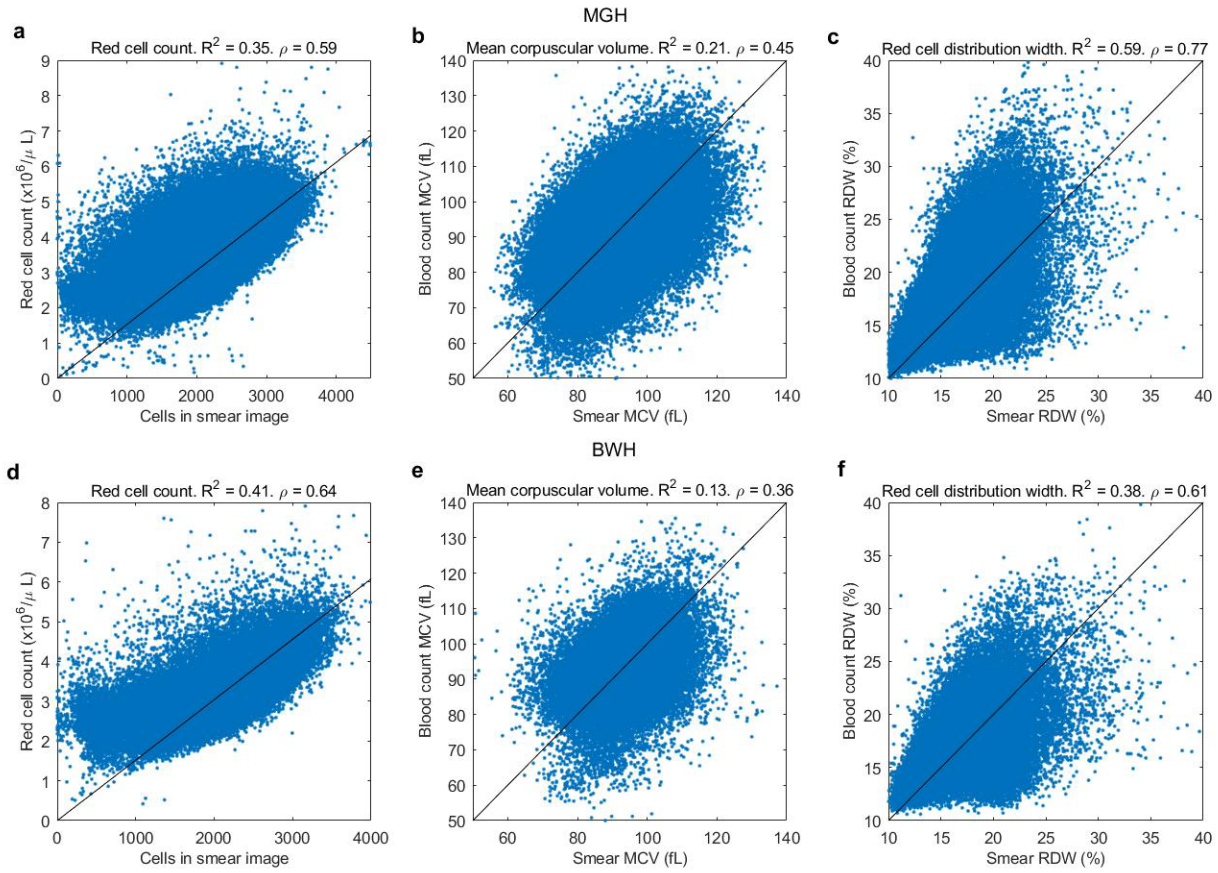
estimates using the corresponding CellaVision™ image. The manual collected images were taken at a resolution approximately 25% greater than the CellaVision™ images, and so were resized accordingly.

Sensitivity to dataset shift. To test sensitivity of our approach to data set shift[2], all 5000 images in the first training and testing sets were reclassified following a 3-fold increase in image saturation (to reflect potential inter-hospital differences in lighting and smear preparation protocols). The geometric approach showed strong robustness to these changes, with the mean classification probability for the correct class exhibiting an average decrease of only 4%. Examples are given in **Supplemental Figure 5**. The approach was fast, with a ~2500 cell differential taking an average of 0.72s to run (averaged over 100 runs, using a Dell XPS 17 laptop, 64Gb RAM, Intel I7 processor).

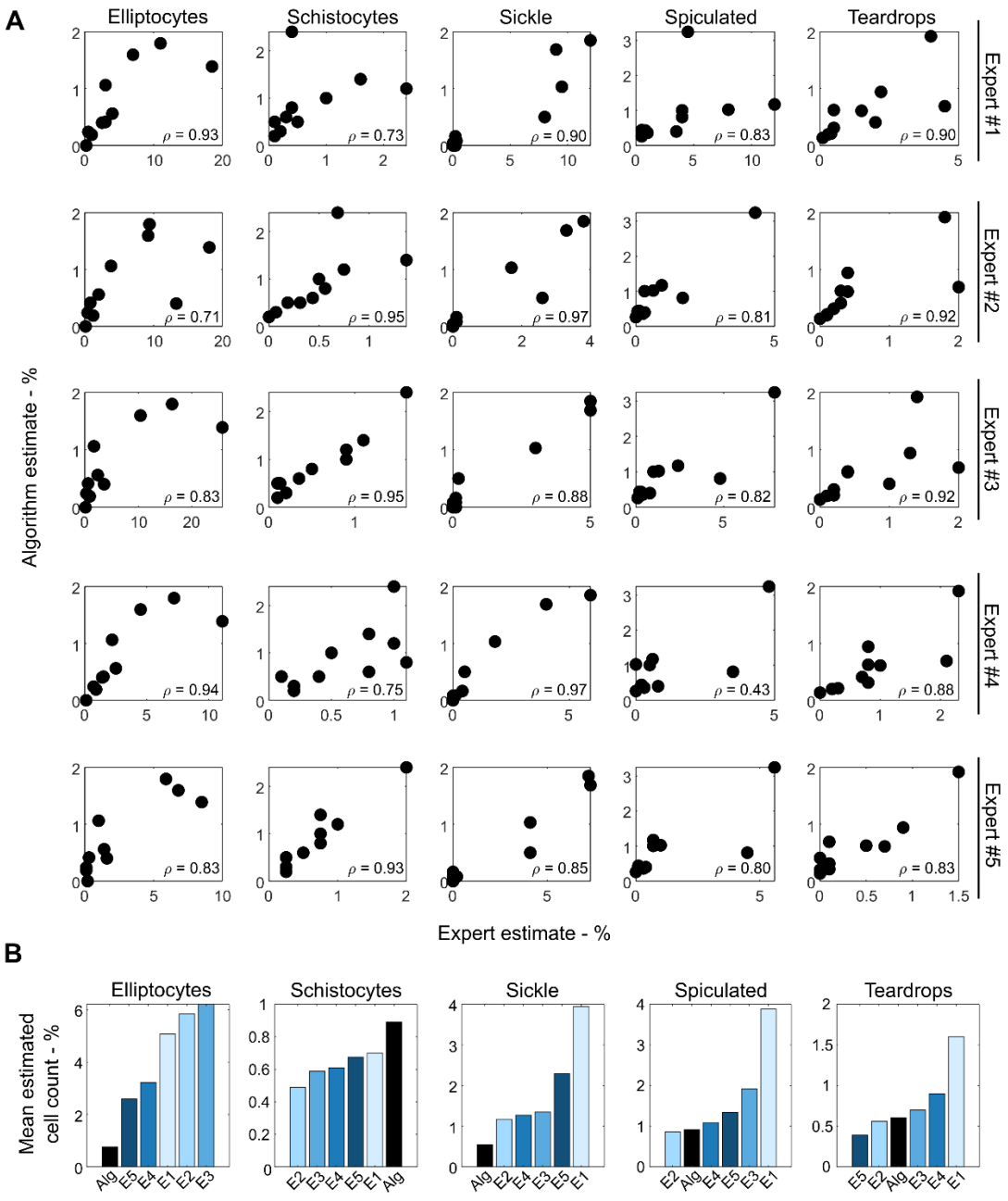
Intra-sample variability. Intra-sample variability of RBC-diff counts was estimated using 8459 cases where two separate blood smear slides were created from a single blood sample, due to simultaneous automatic and manual orders of peripheral blood smear creation. Variability was estimated as the median absolute difference in RBC-diff counts for each cell type. This difference was compared to the expected variability in estimates assuming each blood smear is randomly sampling from the underlying cell population. This variability was estimated by comparison of 10,000 pairs of binomial random variables assuming 2000 events (approximately the number of blood cells per smear image) and using the mean cell prevalence of each morphology type as the binomial event rate. Intra-sample variability is given in **Figure 1C** and **Supplemental Table 7.**

Effect of sample preparation delay on RBC-diff counts. To estimate the effect of delays between blood draw and smear preparation on RBC-diff counts, median counts of each cell type were calculated for each 30 min window of delay time from 0–24 hours. Calculations were performed for all smears in the MGH (n = 281,745) and BWH (n = 56,832) data sets. This analysis highlighted that prolonged storage (>3 hours) significantly increased spiculated cell counts but did not substantially alter the counts of other morphologies (**Supplemental Figure 14**). Thus, prolonged storage may lead to false identification of acanthocytosis and potentially mask other underlying morphologic signals. Across both data sets 93% of smears had storage delays under 3hrs, and thus are likely free from this potential bias.

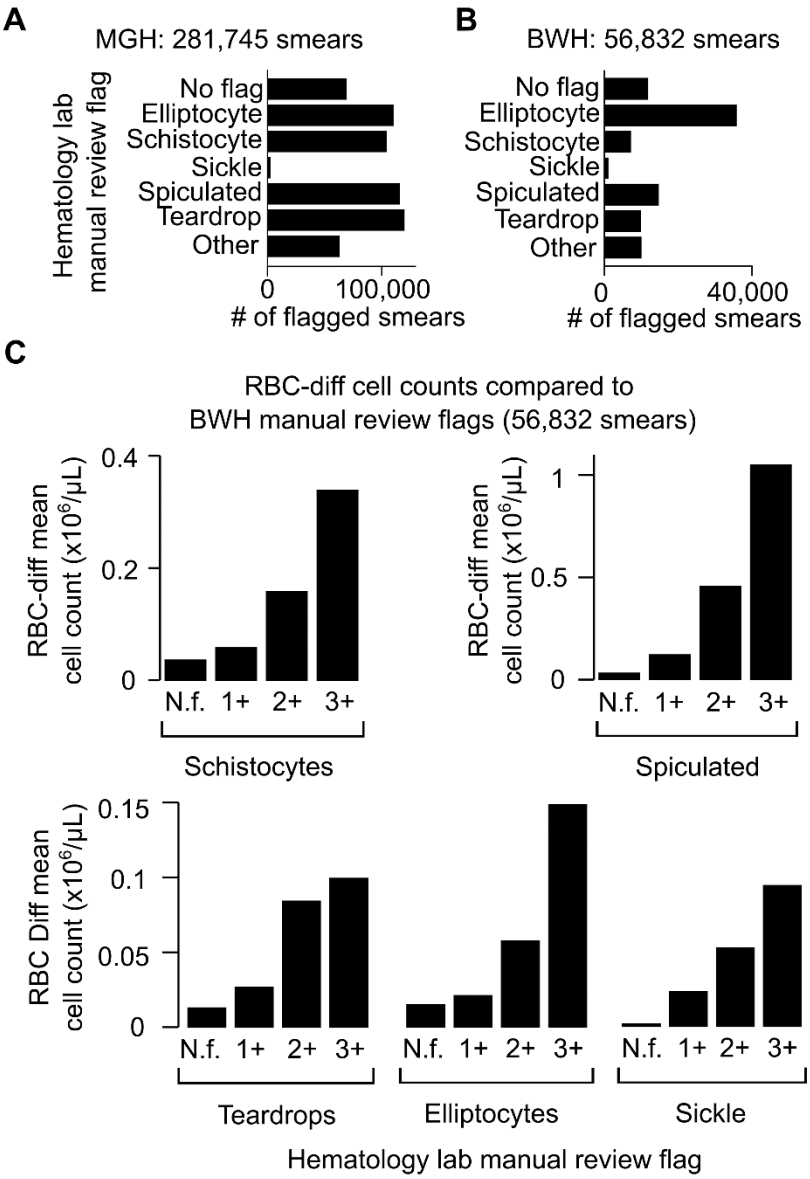
Approximate normal cohort. Due in part to COVID-19 pandemic-related research restrictions, we did not have access to healthy subject smears to define a reference range for RBC-diff counts. Instead, we approximated reference ranges using a patient subset enriched for health. We started with the first smear of each adult MGH patient and excluded anyone with positive morphology grading flags, major comorbidities, an elevated (>48 hours) smear-associated hospital stay, or over 65 years old. Any patient with any recorded blood count abnormalities (see **Supplemental Table 8**), or with more than 5 blood counts in the past 5 years was also excluded, leaving 412 patients: 43% male, 63% white, mean(std) age 37(20) yrs. Chart review of 20 randomly selected patients identified that the most common indication for the blood smear was mild lymphocytosis or white cell left shift due to mild viral infections in otherwise healthy patients. The approximate normal range for absolute and relative cell counts is given in **Supplemental Table 9**.



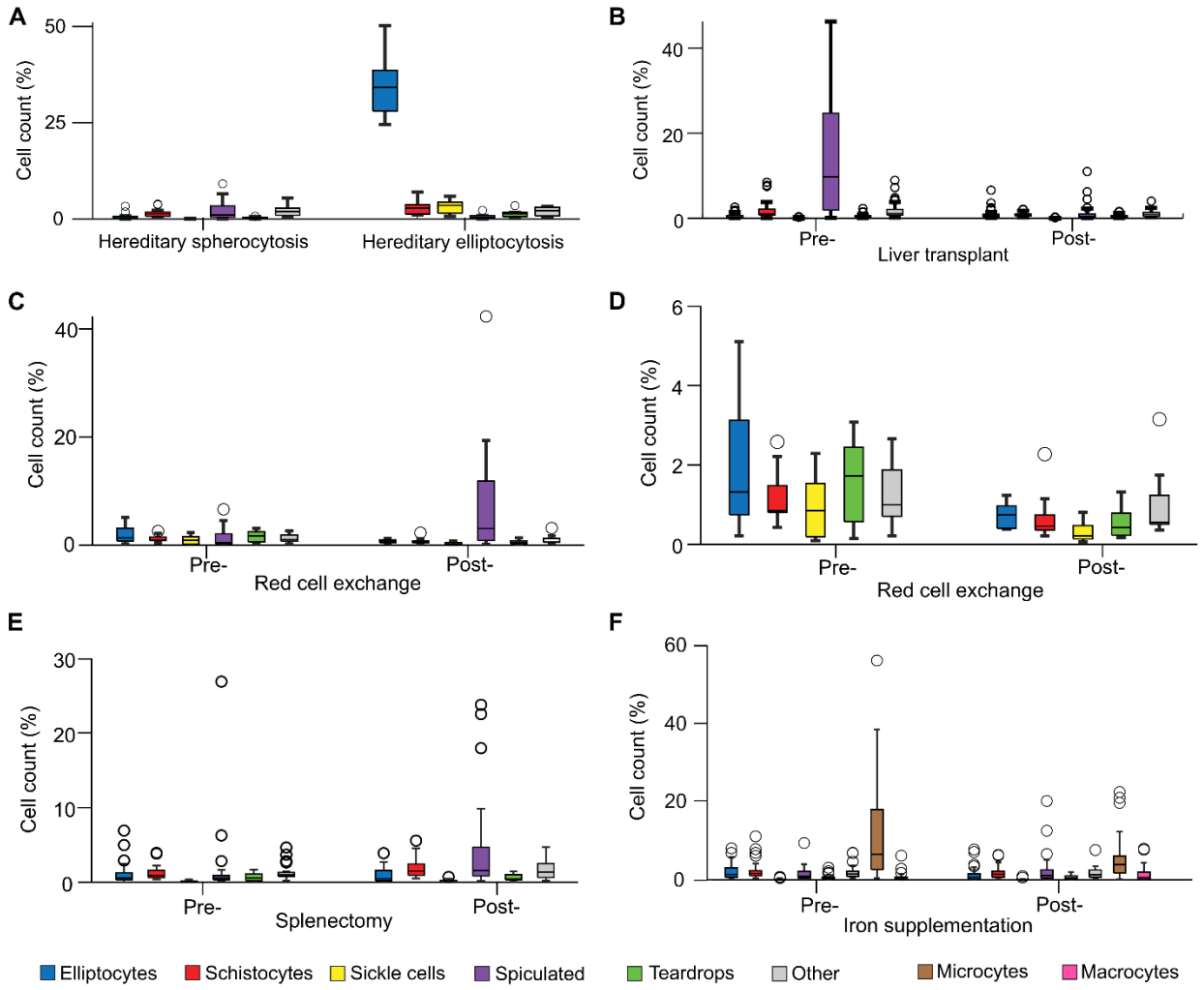
Supplemental Figure 1. Accuracy of smear-estimated blood count indices (Complete blood count [CBC] vs RBC-Diff). The association between red blood cell count (RBC), mean corpuscular volume (MCV), and red cell distribution width (RDW) and smear cell count, smear estimated MCV, and RDW for MGH (**A–C**) and BWH (**D–F**) blood smears. For MCV and RDW the line of unity is presented (black). Given uncertainty in the exact blood volume used per smear, calibration of RBC and smear cell count is not possible, and a linear fit with zero intercept is plotted. The linear fit parameter for RBC and smear cell count at BWH (0.00152) is very similar to that at MGH (0.00153). Concordance is strong even though smears contain much smaller sample sizes (~2000 cells) than CBCs (~50,000 cells), and smear estimations of cell volume are lack of information in the vertical axis.



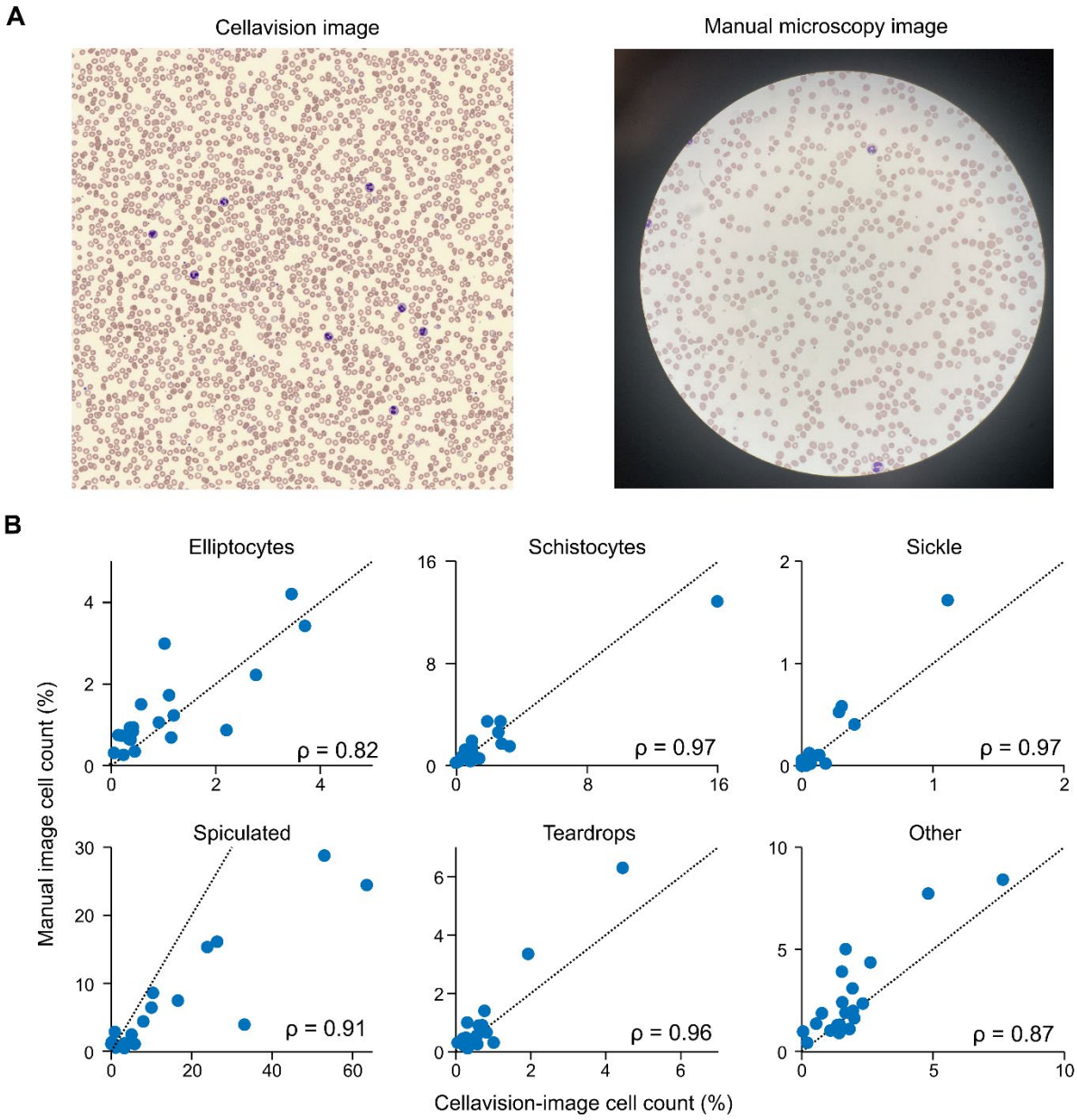
Supplemental Figure 2. Validation of RBC-diff counts by expert hematopathologists and hematologists. (**A**) Comparisons between expert and RBC-diff counts for 5 cell types, across 5 reviewers. RBC-diff estimates showed strong concordance with expert estimates, consistently having strong positive correlations (Spearman's correlation coefficient). (**B**) Despite strong concordance, there were significant differences in the mean count estimate from each expert, reflecting miscalibration. The relative size of RBC-diff mean count varied by cell type, being more conservative than the experts for elliptocytes and sickle cells, more aggressive than experts for schistocytes, and between experts for spiculated and teardrop cells.



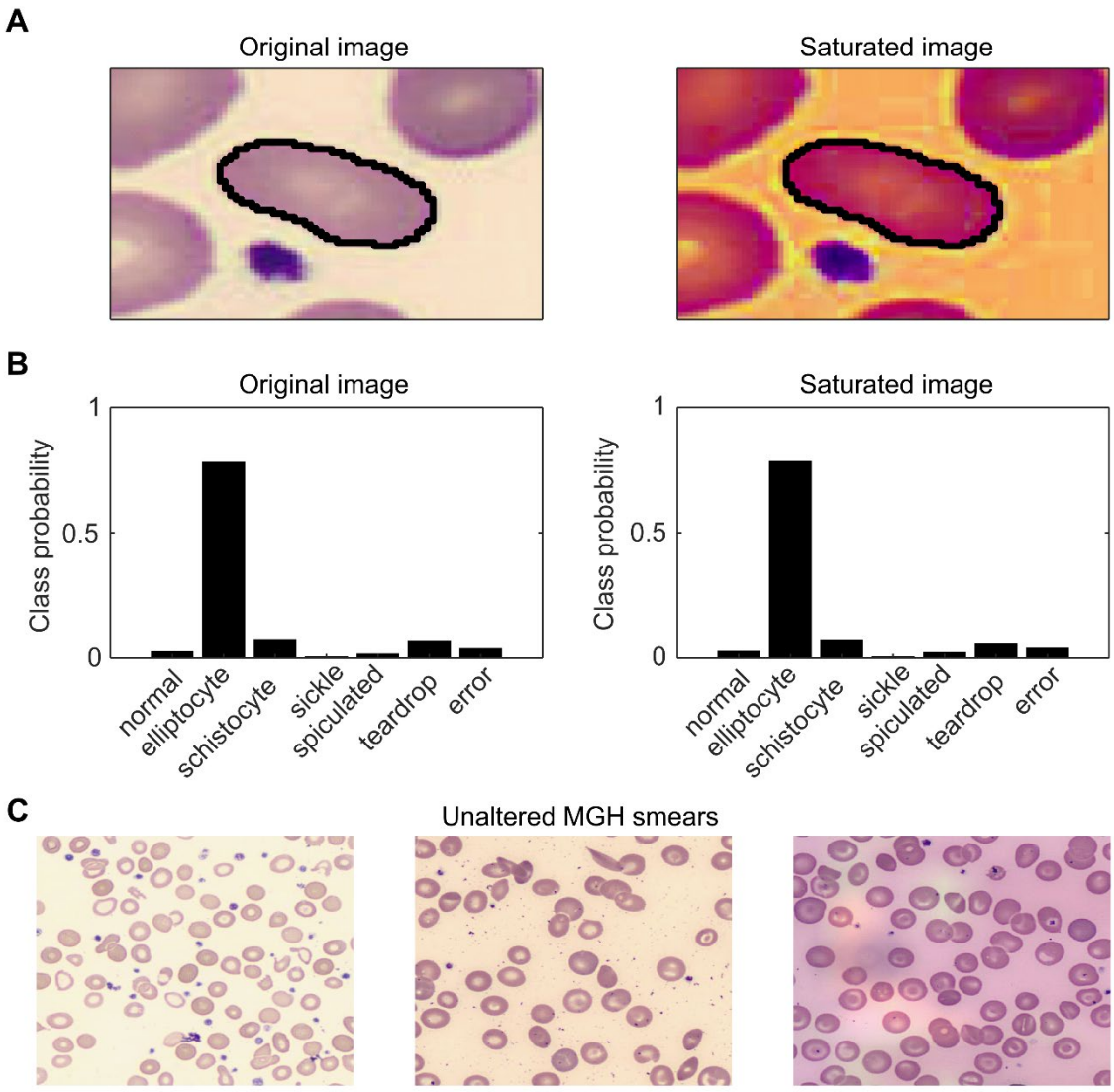
Supplemental Figure 3. Association of clinical lab morphology grading flags with RBC-diff counts. The distributions of flags (from manual review of smears by the hematology lab) across the MGH and BWH smear datasets are presented (**A–B**), showing significant inter-hospital differences in relative frequencies of morphologies. For each morphology, RBC-diff counts were consistent with flags at BWH (**C**) and MGH (see **Figure 1**). A flagged smear refers to a smear having any positive flag ("1+", "2+", "3+", etc.). A smear can have flags for multiple morphologies, and as such the numbers in panels **B** and **C** will add up to more than the total smears at each hospital. N.f., No flag.



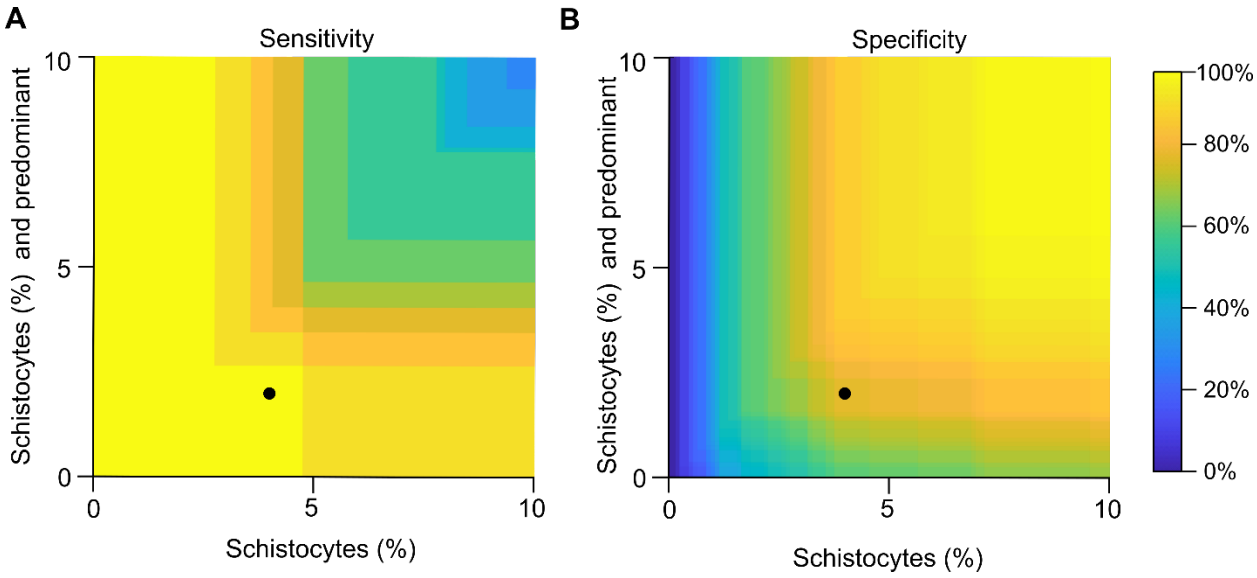
Supplemental Figure 4. Full RBC-diff counts for clinical cohort studies. The full distribution of RBC morphologies is given for the clinical cohorts in **Figure 2**: (**A**) hereditary spherocytosis vs elliptocytosis, (**B**) pre and post liver transplant, (**C-D**) pre and post red cell exchange in sickle cell patients (with and without spiculated cell counts to improve visual clarity), (**E**) pre and post splenectomy, and (**F**) pre and post intravenous iron therapy in iron deficient patients. Most of the exhibited changes in **Figure 2** occur against relatively stable background in other morphologies. In (**C**) the drastic increase in spiculated cells may be due to storage artefacts in the transfused blood. In (**E**) spiculated cells also rise following splenectomy which is also likely a sign of the lack of splenic clearance of RBCs with abnormal morphology.



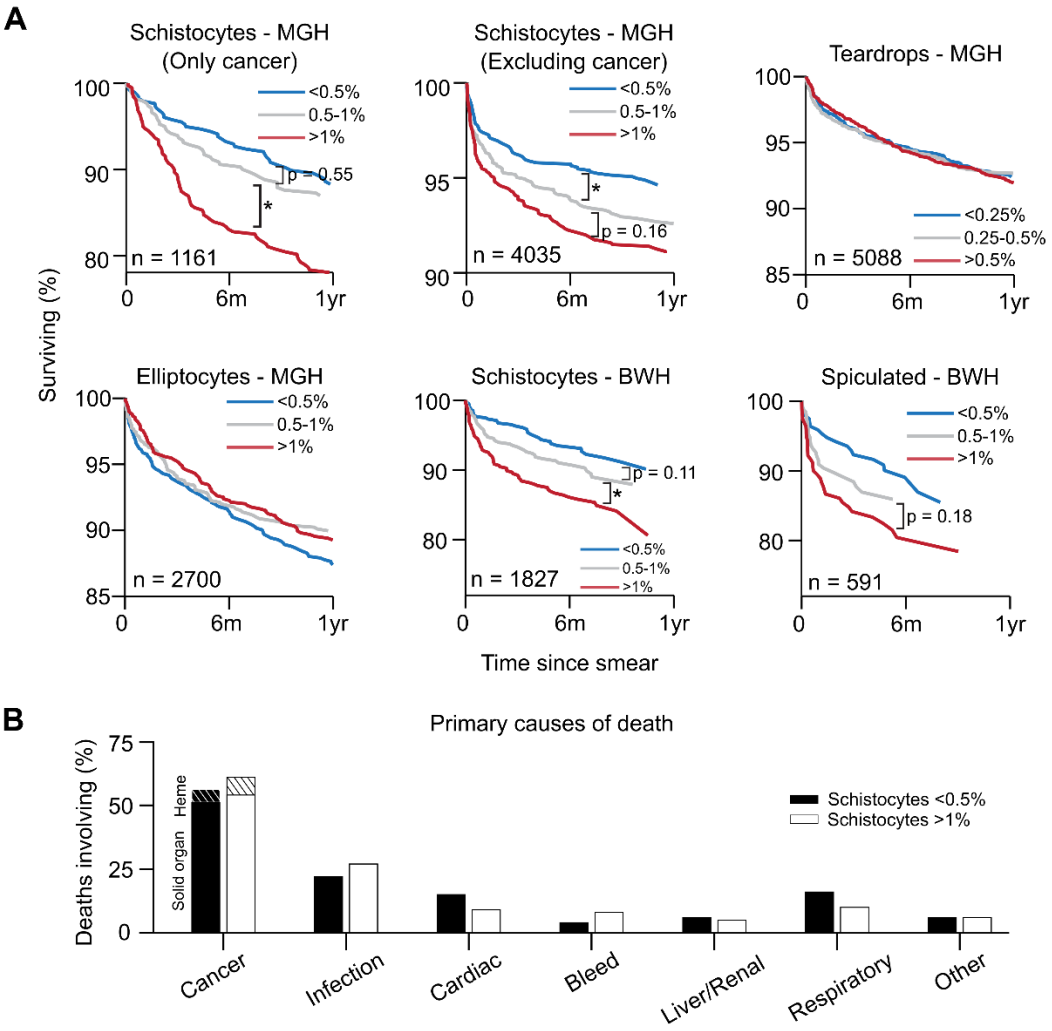
Supplemental Figure 5. Concordance between RBC-diff counts with CellaVision™ images and manually photographed smears. (**A**) Examples of the same blood smear imaged by CellaVision™ and manually through a 50x microscope using a 12 megapixel iPhone 12 camera. (**B**) Concordance between RBC-diff counts across 20 blood smears using CellaVision™ images and manually photographed images. Due to manual microscopy capturing lower density fields than CellaVision™, RBC-diff counts were averaged over photographs of 5-10 microscopy fields. Black space in the manually photographed image in (**A**) corresponds to the microscope casing. Manual images had this black space removed to allow for better contrast detection of RBC boundaries. In (**B**) black lines denote unity, and Pearson correlation coefficients are given. Spiculated cells show lower concordance than other types as spiculated counts are potentially influenced by the length of time since the blood draw which may have varied for CellaVision™ and manual images (see **Supplemental Figure 14**).



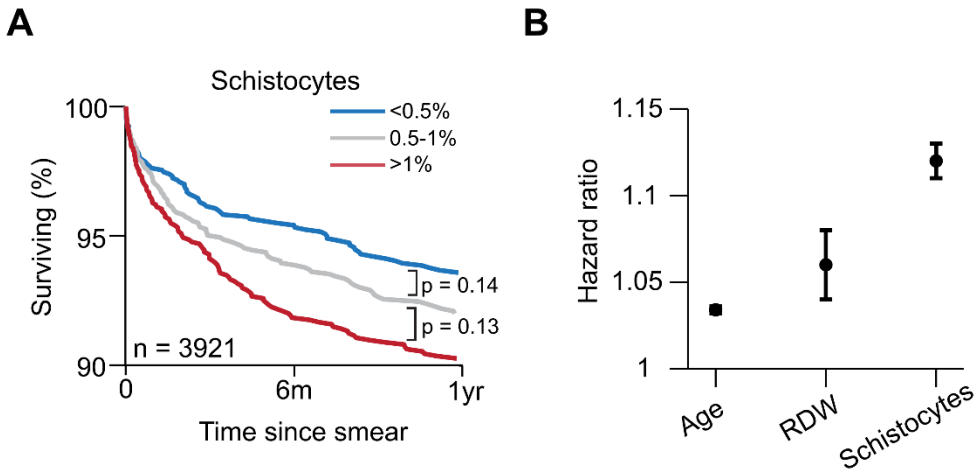
Supplemental Figure 6. Robustness of RBC-diff classifications to changes in image color. Class probabilities for a given elliptocyte were calculated using the RBC-diff before and after increasing image saturation 3-fold (**A**). The RBC-diff's class probabilities (**B**) are seen to be stable across the two images. Across 5000 cell images, image saturation led to a decrease in the correct class probability of 4% for the RBC-diff. Changes to image color of this magnitude were regularly seen across smears at MGH (**C**) and may be even more extreme in other settings due to any variation in smear preparation protocols and imaging devices.



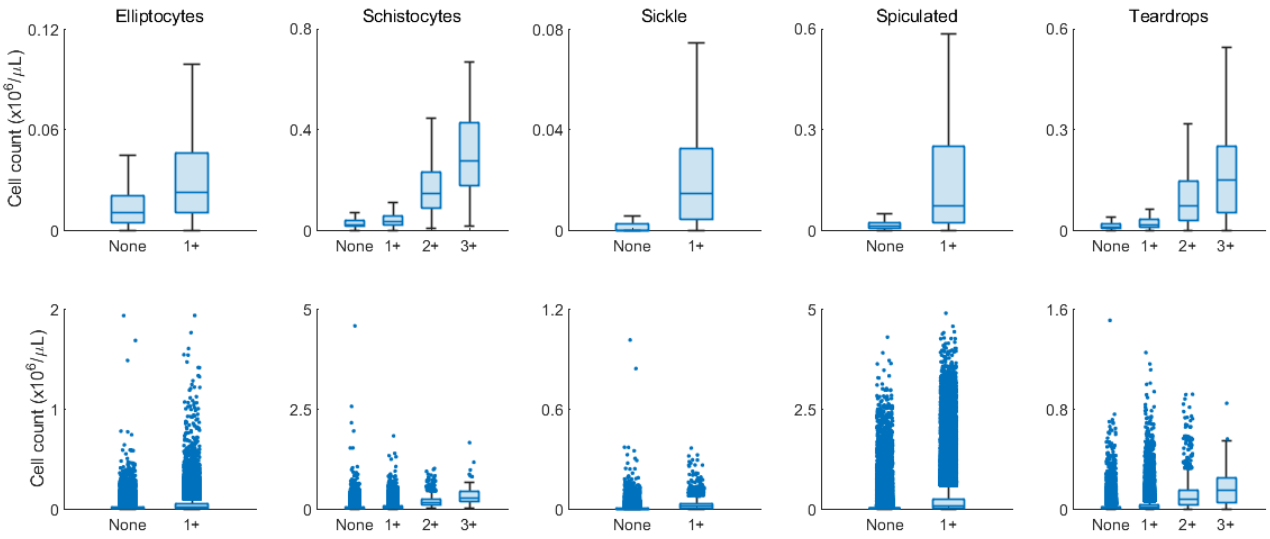
Supplemental Figure 7. Sensitivity and specificity for diagnosis of iTTP/HUS using variable schistocyte quantitation thresholds. In **Figure 3** we illustrated potential diagnostic value of elevated schistocytes (with or without predominance) for differential diagnosis of iTTP/HUS compared to other TMAs. To use both types of information, we investigated sensitivity and specificity of diagnosis using the joint criteria of "schistocyte (%) > x or schistocytes-predominant and schistocyte (%) > y", for all values of x, y from 0–10%. The heatmaps above illustrate the corresponding sensitivity (**A**) and specificity (**B**) in the TMA derivation cohort. An optimal value of 2% for schistocytes + predominant, and 4% for schistocytes was selected to maximize specificity, while retaining high sensitivity. Thresholds were rounded to reflect precision of the algorithm and to adhere to standard clinical practices. Following this analysis, the joint threshold was validated using the validation cohort (**Figure 3F**).



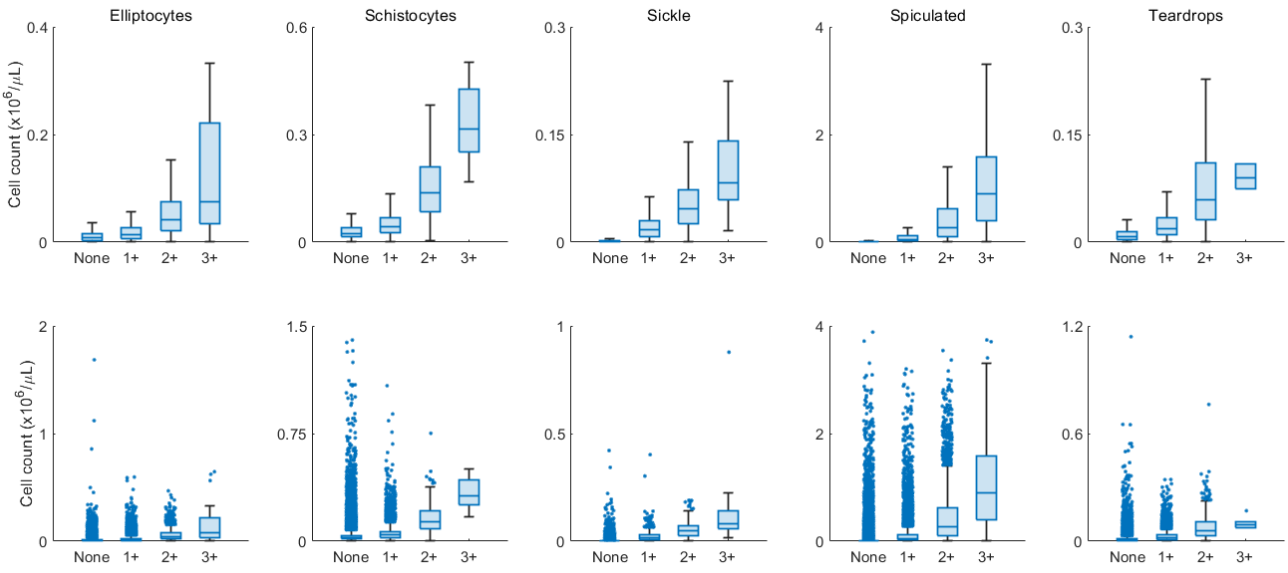
Supplemental Figure 8. Extended all-cause mortality curves and cause of death analysis. (**A**) All-cause mortality curves, as in **Figure 4B**, are given for schistocytes at MGH excluding any patients with a cancer diagnosis in their comorbidity profile, and only those with a cancer diagnosis (see **Supplemental Methods** for definition of comorbidities); elliptocytes and teardrop cells at MGH; and for schistocytes, and spiculated cells at BWH. All cohorts used the same matching criteria as in **Figure 4**, though due to smaller sample sizes, BWH cohorts were not matched on race. No significant associations can be seen for teardrop cells or elliptocytes. Each schistocyte cohort shows a significant risk elevation. Some pairwise schistocyte group comparisons did not reach statistical significance and may reflect reduced power from the smaller sample size than analyzed in **Figure 4B**. (**B**) Chart review of 100 randomly selected deceased patients from the low and high schistocyte groups in **Figure 4B** revealed no significant differences (2-sided Chi-squared test, $\chi^2 = 4.9$, p = 0.56, df = 6) in primary causes of death between the two groups, with cause of death classified subjectively as shown in the figure. As with the MGH cohort, elevated schistocytes (>1%) were associated with significantly increased all-cause mortality at BWH. Qualitatively similar stratifications to those in **Figure 4B** were seen between the BWH low and mid schistocyte groups, and between the mid and high spiculated groups, but results were not statistically significant, potentially due to smaller sample sizes.



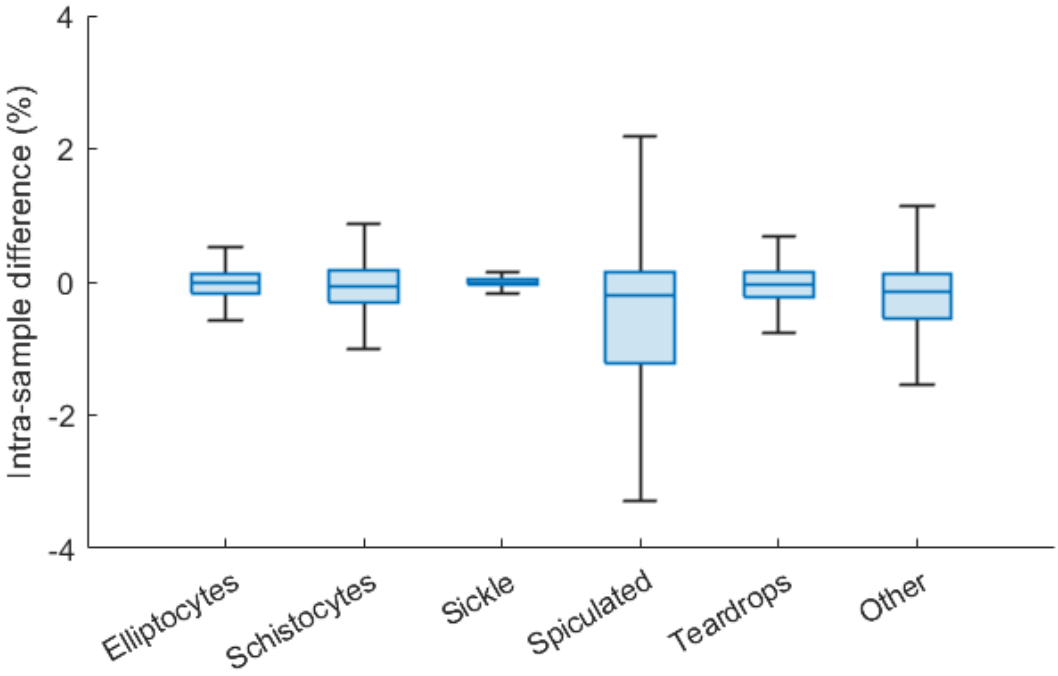
Supplemental Figure 9. Schistocyte-mortality associations after controlling for red cell distribution width (RDW). (**A)** After expanding patient matching criteria for **Figure 4B** to include red cell distribution width (<2% difference), the schistocyte-mortality association persisted (though significance was reduced due to the smaller sample size). (**B**) Hazard ratios from a multivariate Cox proportional hazards using age, RDW and schistocyte count as predictors, and 30d all-cause mortality as the outcome, for all initial smears in the MGH cohort. Hazard ratios were adjusted to be the relative risk change per 1yr for age, and per 1% absolute increase for RDW and schistocytes.



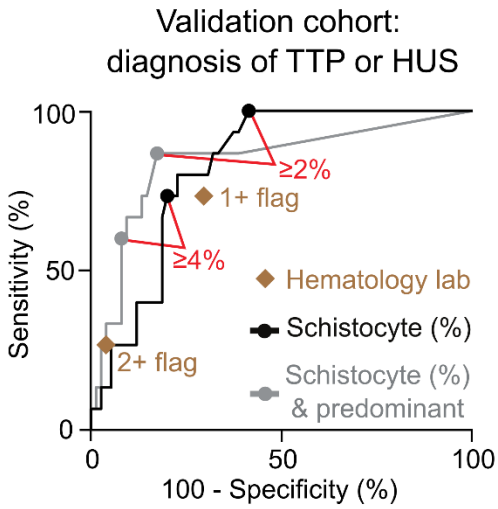
Supplemental Figure 10. Concordance between clinical morphology grades and RBC-diff counts at MGH. Distributions showing RBC-diff counts for all 281,745 smears at MGH stratified by clinical morphology grades (underlying the summary data in **Figure 1D**). Given the presence of some extreme outliers, plots are given with (bottom) and without (top) outliers included. Manual review of a small number of cases with no morphology grading flag but extremely high cell counts regularly revealed cases that appeared to have high densities of the cells of interest, and probably warranted flags, but this analysis was not performed systematically.



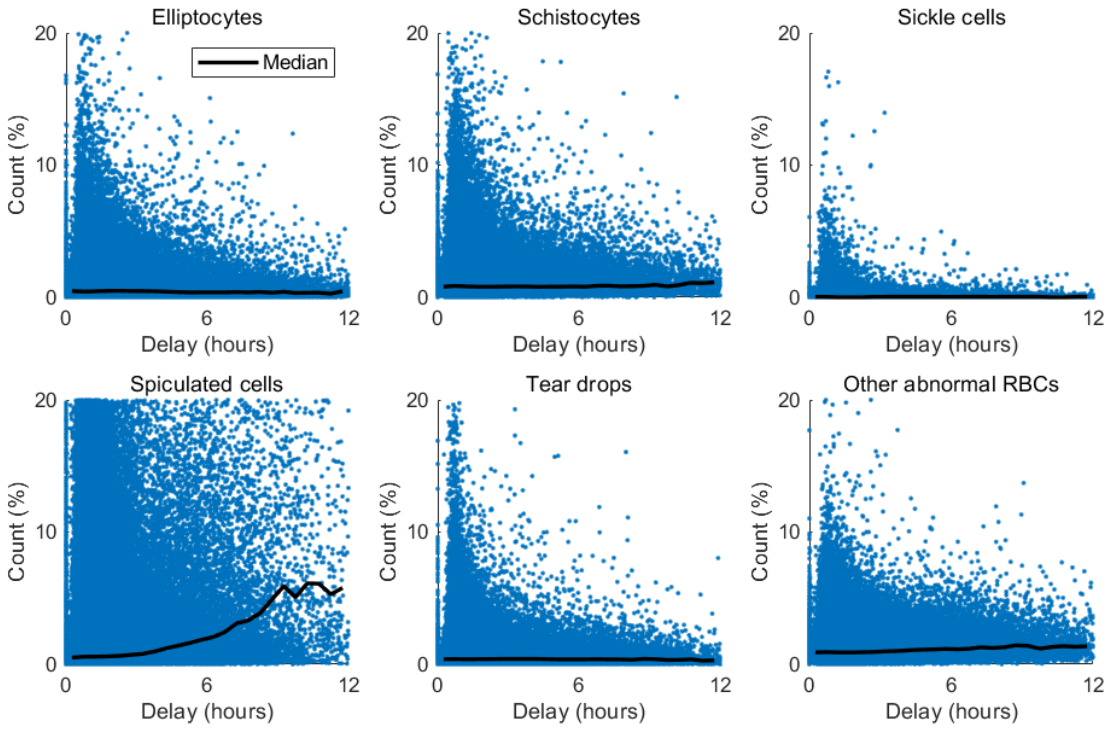
Supplemental Figure 11. Concordance between clinical morphology grades and RBC-diff counts at BWH. Distributions showing RBC-diff counts for all 56,832 smears at BWH stratified by clinical morphology grades (data underlying **Supplemental Figure 3C**). Given the presence of some extreme outliers, plots are given with (bottom) and without (top) outliers included. Manual review of the small number of cases with no morphology grading flag but extremely high cell counts regularly revealed cases that appeared to have high densities of the cells of interest, and probably warranted flags, but this analysis was not performed systematically.



Supplemental Figure 12. Distribution of intra-sample differences in RBC-diff counts, for 8459 pairs of blood smears. The distribution underlying summary data in **Figure 1C** is given. Note that spiculated cell differences are much larger than those for other cell types, and likely reflects the fact that spiculation increases with the delay time between blood draw and smear creation (**Supplemental Figure 14**). Paired smears were created from the same blood sample but often many hours apart.



Supplemental Figure 13. ROC curves for diagnosis of iTTP or HUS using schistocyte counts in the TMA validation cohort. In the TMA validation cohort (90 patients, 10 iTTP, 5 HUS), we present sensitivity-specificity curves for diagnosis of iTTP or HUS, using either schistocyte morphology grading flags (diamonds), schistocytes (black), and schistocytes when they are the predominant morphologies (gray). Specific operating points at 2% and 4% are noted in red. Curve characteristics are similar to those for the derivation cohort (**Figure 3E**), with the RBC-diff criteria providing better performance and more granularity than the morphology grading flags.



Supplemental Figure 14. Association between smear preparation delay and RBC-diff counts. For each morphology, the median cell count is presented for every 30 min increment of delay time from 0–24 hours (black line). Most morphologies are not significantly affected by increased delay time between blood draw and smear preparation. However, while spiculated cell counts are reasonably stable over the first 3 hours of delay, additional delay is associated with a ~0.8% increase in median cell count per hour. 93% of smears in the dataset had a delay time of less than 3 hours, and this effect is not expected to alter the key findings of this study.

Supplemental Tables

Supplemental Table 1. Characteristics of MGH and BWH cohort

Demographics	MGH	BWH
Total patients - N	48196	9894
Smears per patient - N	5.7(13.4), 2	5.7(10.7), 2
Age - years	52.6(24.4), 58.3	59.1(18.1), 62
Sex - % male	51.40%	47.70%
Race -% white/caucasian	73%	69.80%
Comorbidities & diagnoses		
Anemia - %	12.1%	15.4%
Cancer (hematologic) - %	18.4%	18.5%
Cancer (solid organ) - %	29.6%	32.4%
Liver disease - %	5.2%	4.2%
Organ transplant - %	8.3%	10.1%
Pregnancy - %	0.6%	1.3%
Renal disease - %	9.1%	12.3%
Sickle cell disease - %	0.6%	2.9%
Sepsis - %	4.8%	5.5%
Thalassemia - %	0.1%	0.2%
Thrombocytopenia - %	4.7%	6.0%
Morphology grading flags		
No RBC flags - %	42.3%	44.10%
Elliptocytes - % +ve	23%	43.20%
Schistocytes - % +ve	18.30%	7.30%
Sickle cells - % +ve	0.30%	0.60%
Spiculated cells - % +ve	35.30%	22%
Teardrop cells - % +ve	20.30%	11.20%
Relative RBC-diff counts		
Elliptocytes - %	0.7(1.2), 0.3	0.6(1.1), 0.3
Schistocytes - %	1.0(1.2), 0.7	1.2(1.5), 0.8
Sickle cells - %	0.06(0.2), 0.03	0.06(0.3), 0
Spiculated cells - %	4.1(9.5), 0.8	2.5(7.8), 0.3
Teardrop cells - %	0.5(0.8), 0.3	0.4(0.7), 0.2
Other RBCs - %	1.3(1.2), 0.9	1.1(1.9), 0.6
Absolute RBC-diff counts		
Elliptocytes - $x10^6/\mu L$	0.03(0.05), 0.01	0.02(0.04), 0.01
Schistocytes - $x10^6/\mu L$	0.04(0.05), 0.03	0.04(0.05), 0.03
Sickle cells - $x10^6/\mu L$	0.002(0.007), 0.001	0.002(0.008), 0
Spiculated cells - $x10^6/\mu L$	0.16(0.37), 0.03	0.09(0.28), 0.01
Teardrop cells - $x10^6/\mu L$	0.02(0.03), 0.01	0.01(0.03), 0.01
Other RBCs - $x10^6/\mu L$	0.05(0.05), 0.03	0.04(0.07), 0.02
Outcomes		
30-day mortality - %	6.30%	8%
1 year mortality - %	12%	15.50%

Legend: Data is presented for each patient's earliest available smear. Continuous variables are presented as mean(std), median, and binary variables as percentages.

Supplemental Table 2. Characteristics of clinical cohort studies

	Differential diagnosis		Liver transplant		Red cell exchange		Iron supplementation		Splenectomy	
Demographics	Spherocytosis	Elliptocytosis	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Total patients - N	17	6	46	46	9	9	30	30	21	21
Age - years	31.4(25.5) 31.9	56.8(26.8) 63.7	52.8(12.7) 55.4	53.2(12.7) 55.8	39.8(9.9) 41.2	39.8(9.9) 41.2	49.5(22.7) 55	49.9(22.7) 55.5	49.5(20.2) 54.3	50(20.3) 54.7
Sex - % male	59%	33%	48%	48%	33%	33%	30%	30%	62%	62%
Race -% white/Caucasian	82%	50%	80%	80%	0%	0%	80%	80%	71%	71%
Morphology grading flags										
No RBC flags - %	41%	17%	26%	48%	33%	22%	37%	43%	38%	10%
Elliptocytes - % +ve	18%	83%	26%	33%	44%	56%	40%	40%	48%	38%
Schistocytes - % +ve	6%	50%	52%	22%	44%	22%	37%	37%	29%	48%
Sickle cells - % +ve	0%	0%	0%	0%	67%	56%	0%	0%	5%	10%
Spiculated cells - % +ve	35%	0%	65%	22%	33%	22%	20%	27%	24%	38%
Teardrop cells - % +ve	6%	17%	28%	24%	22%	22%	27%	23%	33%	10%
Relative RBC-diff counts										
Elliptocytes - %	0.6(0.8) 0.3	34.9(9.1) 34.2	0.5(0.5) 0.3	0.9(1.1) 0.6	2(1.7) 1.3	0.7(0.4) 0.7	2.1(2) 1.4	1.5(1.9) 0.8	1.3(1.8) 0.5	1(1.1) 0.5
Schistocytes - %	1.6(1.1) 1.3	3.1(2.2) 2.8	1.9(2.1) 1.1	0.8(0.5) 0.7	1.2(0.7) 0.9	0.7(0.6) 0.5	2.4(2.5) 1.7	1.9(1.6) 1.5	1.2(1) 0.9	2(1.5) 1.5
Sickle cells - %	0(0) 0	3.3(1.9) 3.6	0(0.1) 0	0(0.1) 0	0.9(0.8) 0.9	0.3(0.3) 0.2	0.2(0.2) 0.1	0.1(0.2) 0.1	0.1(0.1) 0	0.2(0.2) 0.1
Spiculated cells - %	2.1(2.6) 1	0.7(0.8) 0.5	13.2(13.1) 9.8	1(1.9) 0.4	1.6(2.3) 0.4	9(13.9) 3	1.6(1.8) 0.9	2.6(4.2) 1.2	2.2(5.8) 0.6	4.9(7.3) 1.5
Teardrop cells - %	0.28(0.19) 0.24	1.45(1.13) 1.34	0.43(0.44) 0.29	0.47(0.39) 0.4	1.58(1.05) 1.72	0.54(0.4) 0.42	0.8(0.71) 0.56	0.71(0.56) 0.56	0.67(0.54) 0.58	0.6(0.43) 0.41
Other RBCs - %	2.1(1.4) 1.9	2(1.2) 2.2	1.8(1.8) 1.1	1(0.7) 0.9	1.2(0.8) 1	1(0.9) 0.6	2(1.7) 1.5	1.8(1.4) 1.3	1.3(1.1) 1	1.6(1.2) 1.4
Absolute RBC-diff counts										
Elliptocytes - $x10^6/\mu L$	0.02(0.03) 0.01	1.34(0.49) 1.2	0.01(0.02) 0.01	0.03(0.04) 0.02	0.05(0.03) 0.04	0.02(0.01) 0.02	0.08(0.08) 0.05	0.07(0.09) 0.03	0.05(0.07) 0.02	0.04(0.05) 0.02
Schistocytes - $x10^6/\mu L$	0.06(0.05) 0.04	0.11(0.09) 0.09	0.05(0.08) 0.03	0.03(0.02) 0.02	0.03(0.03) 0.02	0.02(0.02) 0.01	0.09(0.1) 0.06	0.08(0.08) 0.05	0.04(0.04) 0.03	0.07(0.04) 0.06
Sickle cells - $x10^6/\mu L$	0(0) 0	0.12(0.06) 0.13	0(0) 0	0(0) 0	0.02(0.02) 0.01	0.01(0.01) 0.01	0.01(0.01) 0	0.01(0.01) 0	0(0) 0	0.01(0.01) 0
Spiculated cells - $x10^6/\mu L$	0.08(0.11) 0.03	0.03(0.03) 0.02	0.35(0.37) 0.2	0.04(0.07) 0.01	0.05(0.08) 0.01	0.29(0.45) 0.08	0.06(0.07) 0.03	0.11(0.19) 0.05	0.09(0.28) 0.02	0.19(0.31) 0.05
Teardrop cells - $x10^6/\mu L$	0.01(0.06) 0.09	0.05(0.03) 0.05	0.01(0.02) 0.01	0.02(0.01) 0.01	0.04(0.02) 0.05	0.02(0.01) 0.02	0.03(0.03) 0.02	0.03(0.03) 0.02	0.02(0.02) 0.02	0.02(0.02) 0.02
Other RBCs - $x10^6/\mu L$	0.07(0.05) 0.07	0.07(0.04) 0.08	0.05(0.07) 0.03	0.03(0.03) 0.03	0.03(0.02) 0.02	0.03(0.03) 0.02	0.08(0.07) 0.06	0.08(0.07) 0.05	0.05(0.05) 0.03	0.06(0.04) 0.04

Legend: Continuous variables are presented as mean(std), median, and binary variables as percentages.

Supplemental Table 3. Mean RBC-diff counts for the TMA derivation cohort

	Elliptocytes (%)	Schistocytes (%)	Sickle (%)	Spiculated (%)	Teardrops (%)	Other (%)
iTTP (n = 15)	0.23 (0.14)	6.29 (3.67)	0.05 (0.07)	2.51 (4)	1.54 (0.98)	2.08 (0.86)
HUS (n = 3)	0.41 (0.14)	8.04 (4.65)	0.04 (0.03)	0.95 (0.51)	1.08 (0.3)	1.82 (0.38)
Transplant (n = 10)	1.19 (1.33)	2.36 (1.64)	0.09 (0.12)	2.17 (1.99)	0.94 (0.84)	1.66 (1.01)
DIC (n = 25)	0.53 (0.58)	2.58 (2.21)	0.11 (0.16)	15.92 (22.3)	1.01 (1.7)	2.1 (1.39)
Cancer (n = 19)	1 (1.2)	2.69 (3.59)	0.09 (0.13)	6.51 (11.51)	0.56 (0.29)	1.65 (1.32)
Other (n = 15)	0.75 (0.91)	2.73 (2.95)	0.08 (0.17)	3.31 (9.49)	0.81 (0.9)	1.43 (1.14)
Pregnancy (n = 3)	0.17 (0.24)	1.67 (1.94)	0.02 (0.04)	4.46 (3.69)	0.21 (0.36)	0.92 (0.89)
Autoimmune (n = 5)	0.5 (0.35)	1.52 (1.41)	0.12 (0.08)	5.77 (4.84)	1.38 (1.25)	1.67 (1.02)
Drug (n = 11)	0.39 (0.31)	0.64 (0.41)	0.02 (0.03)	3.97 (8.28)	0.41 (0.31)	0.8 (0.45)

Legend: Continuous variables are presented as mean (std)

Supplemental Table 4. Characteristics of MGH matched survival analysis cohorts, schistocytes and spiculated cells

	Schistocytes			Spiculated		
Demographics	<0.5%	0.5–1%	>1%	<0.5%	0.5–1%	>1%
Total patients - N	1732	1732	1732	2142	2142	2142
Age - years	54.1 (22.2), 59.7	54.0(22.2), 59.8	53.7(22.5), 59.3	50.6(24.5), 57.8	50.3 (24.8), 57.5	50.2(24.9), 57.6
Sex - % male	51%	51%	51%	53%	53%	53%
Race -% white/caucasian	83%	83%	83%	83%	83%	83%
Cancer - %	22%	22%	22%	15%	15%	15%
Morphology grading flags						
No RBC flags - %	75%	74%	70%	75%	76%	75%
Elliptocytes - % +ve	6%	6%	6%	7%	7%	7%
Schistocytes - % +ve	4%	4%	4%	2%	2%	2%
Sickle cells - % +ve	0%	0%	0%	0%	0%	0%
Spiculated cells - % +ve	9%	9%	9%	15%	15%	15%
Teardrop cells - % +ve	7%	7%	7%	5%	5%	5%
Relative RBC-diff counts						
Elliptocytes - %	0.3(0.3) 0.2	0.3(0.3) 0.2	0.3(0.3) 0.2	0.3(0.3) 0.2	0.3(0.3) 0.2	0.3(0.3) 0.2
Schistocytes - %	0.3(0.1) 0.3	0.7(0.1) 0.7	1.4(0.5) 1.3	0.6(0.3) 0.5	0.6(0.3) 0.5	0.6(0.3) 0.5
Sickle cells - %	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - %	0.6(0.9) 0.4	0.6(0.9) 0.4	0.6(0.9) 0.4	0.3(0.1) 0.3	0.7(0.1) 0.7	3.1(6) 1.6
Teardrop cells - %	0.3(0.2) 0.2	0.3(0.2) 0.2	0.3(0.2) 0.3	0.3(0.2) 0.2	0.3(0.2) 0.2	0.2(0.2) 0.2
Other RBCs - %	0.7(0.3) 0.6	0.7(0.4) 0.7	0.9(0.3) 0.9	0.7(0.4) 0.6	0.8(0.4) 0.7	0.9(0.4) 0.8
Absolute RBC-diff counts						
Elliptocytes - $x10^6/\mu L$	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01
Schistocytes - $x10^6/\mu L$	0.01(0) 0.01	0.03(0.01) 0.03	0.06(0.02) 0.05	0.02(0.01) 0.02	0.02(0.02) 0.02	0.02(0.01) 0.02
Sickle cells - $x10^6/\mu L$	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - $x10^6/\mu L$	0.03(0.04) 0.02	0.03(0.04) 0.02	0.03(0.04) 0.02	0.01(0.01) 0.01	0.03(0.01) 0.03	0.12(0.22) 0.07
Teardrop cells - $x10^6/\mu L$	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01
Other RBCs - $x10^6/\mu L$	0.03(0.02) 0.02	0.03(0.02) 0.03	0.04(0.02) 0.03	0.03(0.02) 0.02	0.03(0.02) 0.03	0.04(0.02) 0.03
Outcomes						
6-month mortality	4.7%	6.6%	9.5%	5.3%	4.7%	7.7%
1 year mortality	6.4%	8.2%	11.3%	6.9%	6.8%	8.6%

Legend: Continuous variables are presented as mean(std), median, and binary variables as percentages.

Supplemental Table 5. Characteristics of MGH matched survival analysis cohorts, elliptocytes and teardrop cells

	Elliptocytes			Teardrops		
Demographics	<0.5%	0.5–1%	>1%	<0.25%	0.25-0.5%	>0.5%
Total patients - N	900	900	900	1696	1696	1696
Age - years	63.1(17.9), 66.7	63.0(18.0), 66.5	62.7(18.1), 66.7	53.7(20.9), 59.1	53.4(21.2), 59.1	53.1(21.2), 58.5
Sex - % male	41%	41%	41%	51%	51%	51%
Race -% white/caucasian	93%	93%	93%	88%	88%	88%
Cancer - %	39%	39%	39%	31%	31%	31%
Morphology grading flags						
No RBC flags - %	60%	61%	61%	76%	76%	76%
Elliptocytes - % +ve	26%	26%	26%	6%	6%	6%
Schistocytes - % +ve	6%	6%	6%	3%	3%	3%
Sickle cells - % +ve	0%	0%	0%	0%	0%	0%
Spiculated cells - % +ve	10%	10%	10%	5%	5%	5%
Teardrop cells - % +ve	10%	10%	10%	12%	12%	12%
Relative RBC-diff counts						
Elliptocytes - %	0.2(0.1) 0.2	0.7(0.1) 0.7	1.6(1.7) 1.3	0.3(0.3) 0.2	0.4(0.3) 0.3	0.4(0.3) 0.3
Schistocytes - %	0.6(0.4) 0.6	0.7(0.3) 0.6	0.7(0.3) 0.7	0.6(0.4) 0.5	0.6(0.3) 0.6	0.7(0.3) 0.6
Sickle cells - %	0(0) 0	0(0) 0	0.1(0.1) 0	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - %	0.6(0.7) 0.4	0.6(0.7) 0.4	0.6(0.7) 0.4	0.5(0.7) 0.4	0.5(0.7) 0.4	0.5(0.8) 0.3
Teardrop cells - %	0.3(0.2) 0.2	0.3(0.2) 0.3	0.4(0.2) 0.3	0.1(0.1) 0.1	0.4(0.1) 0.3	0.9(0.5) 0.7
Other RBCs - %	0.8(0.4) 0.7	0.8(0.4) 0.7	0.9(0.4) 0.9	0.7(0.4) 0.6	0.7(0.4) 0.7	0.9(0.4) 0.8
Absolute RBC-diff counts						
Elliptocytes - $x10^6/\mu L$	0.01(0.01) 0.01	0.03(0.01) 0.03	0.06(0.06) 0.05	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01
Schistocytes - $x10^6/\mu L$	0.02(0.01) 0.02	0.03(0.01) 0.02	0.03(0.01) 0.02	0.02(0.02) 0.02	0.03(0.02) 0.02	0.03(0.02) 0.02
Sickle cells - $x10^6/\mu L$	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - $x10^6/\mu L$	0.02(0.03) 0.01	0.02(0.03) 0.02	0.02(0.03) 0.02	0.02(0.03) 0.01	0.02(0.03) 0.01	0.02(0.03) 0.01
Teardrop cells - $x10^6/\mu L$	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0) 0.01	0.01(0) 0.01	0.03(0.02) 0.03
Other RBCs - $x10^6/\mu L$	0.03(0.02) 0.03	0.03(0.02) 0.03	0.04(0.02) 0.03	0.03(0.02) 0.02	0.03(0.02) 0.03	0.04(0.02) 0.03
Outcomes						
6-month mortality	8.3%	7.9%	7.3%	5.3%	5.4%	5.5%
1 year mortality	12.0%	9.6%	10.0%	7.3%	6.9%	7.4%

Legend: Continuous variables are presented as mean(std), median, and binary variables as percentages.

Supplemental Table 6. Characteristics of BWH matched survival analysis cohorts, schistocytes

	Schistocytes		
Demographics	<0.5%	0.5–1%	>1%
Total patients - N	609	609	609
Age - years	55.4(17.2), 57.4	55.2(17.1), 57.6	55.0(17.4), 57.9
Sex - % male	45%	45%	45%
Race -% white/caucasian	67%	67%	67%
	23%	23%	23%
Morphology grading flags			
No RBC flags - %	69%	66%	65%
Elliptocytes - % +ve	28%	28%	28%
Schistocytes - % +ve	1%	1%	1%
Sickle cells - % +ve	0%	0%	0%
Spiculated cells - % +ve	2%	2%	2%
Teardrop cells - % +ve	2%	2%	2%
Relative RBC-diff counts			
Elliptocytes - %	0.2(0.2) 0.2	0.2(0.3) 0.2	0.3(0.3) 0.2
Schistocytes - %	0.3(0.1) 0.3	0.7(0.1) 0.7	1.4(0.4) 1.3
Sickle cells - %	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - %	0.2(0.4) 0.1	0.3(0.4) 0.2	0.3(0.4) 0.2
Teardrop cells - %	0.1(0.1) 0.1	0.2(0.2) 0.1	0.2(0.2) 0.2
Other RBCs - %	0.4(0.2) 0.3	0.5(0.2) 0.4	0.7(0.2) 0.6
Absolute RBC-diff counts			
Elliptocytes - $x10^6/\mu L$	0.01(0.01) 0.01	0.01(0.01) 0.01	0.01(0.01) 0.01
Schistocytes - $x10^6/\mu L$	0.01(0) 0.01	0.03(0.01) 0.03	0.06(0.02) 0.05
Sickle cells - $x10^6/\mu L$	0(0) 0	0(0) 0	0(0) 0
Spiculated cells - $x10^6/\mu L$	0.01(0.01) 0	0.01(0.01) 0.01	0.01(0.02) 0.01
Teardrop cells - $x10^6/\mu L$	0.01(0) 0	0.01(0.01) 0.01	0.01(0.01) 0.01
Other RBCs - $x10^6/\mu L$	0.01(0.01) 0.01	0.02(0.01) 0.02	0.03(0.01) 0.03
Outcomes			
6-month mortality	5.25%	7.55%	11.99%
1 year mortality	6.24%	8.54%	12.81%

Legend: Continuous variables are presented as mean(std), median, and binary variables as percentages.

Supplemental Table 7. Variability in RBC-diff counts across 8459 pairs of biologically equivalent blood smears.

	Mean count	Median absolute error (MAE)	Expected sampling MAE	MAE above sampling error
Elliptocytes	0.76%	0.14%	0.20%	0%
Schistocytes	1.23%	0.24%	0.25%	0%
Sickle	0.07%	0.04%	0.05%	0%
Spiculated	5.40%	0.60%	0.50%	0.10%
Teardrops	0.67%	0.18%	0.15%	0.03%
Other	1.40%	0.35%	0.23%	0.12%

*Expected sampling MAE was calculated assuming a binomial distribution (see **Methods** for details)*.

Supplemental Table 8. Reference intervals for complete blood count components at Massachusetts General Hospital

Complete blood count	Abbreviation	Units	Reference interval Male	Reference interval Female
Hematocrit	HCT	%	41–53	36–46
Hemoglobin	HGB	g/dL	13.5–17.5	12.0–16.0
Mean corpuscular hemoglobin	MCH	pg	26–34	26–34
Mean corpuscular hemoglobin concentration	MCHC	g/dL	31–37	31–37
Mean corpuscular volume	MCV	fL	80–100	80–100
Platelet count	PLT	$10^3/\mu L$	150–400	150–400
Red blood cell count	RBC	$10^6/\mu L$	4.5–5.9	4.0–5.2
Red cell distribution width	RDW	%	11.5–14.5	11.5–14.5
White blood cell count	WBC	$10^3/\mu L$	4.5–11.0	4.5–11.0

Supplemental Table 9. Reference ranges for RBC-diff counts based on an "approximate" normal cohort

Cell count	Percentile				
Relative (%)	10th	25th	50th	75th	90th
Normal	94.3%	96.6%	97.9%	98.6%	99.1%
Elliptocytes	<0.01%	0.05%	0.13%	0.26%	0.51%
Schistocytes	0.22%	0.31%	0.47%	0.79%	1.32%
Sickle	0%	0%	<0.01%	0.04%	0.09%
Spiculated	0.10%	0.22%	0.42%	0.86%	1.82%
Teardrops	0.04%	0.08%	0.15%	0.38%	1.00%
Other	0.25%	0.37%	0.59%	1.02%	1.73%
Absolute ($x10^6/\mu L$)	Presented as male value, female value				
Normal	4.47, 4.08	4.66, 4.22	4.9, 4.37	5.12, 4.61	5.31, 4.81
Elliptocytes	0.002, 0	0.003, 0.003	0.007, 0.006	0.016, 0.012	0.03, 0.023
Schistocytes	0.014, 0.009	0.02, 0.013	0.031, 0.019	0.052, 0.03	0.087, 0.042
Sickle	0, 0	0, 0	0.002, 0	0.003, 0.002	0.005, 0.003
Spiculated	0.005, 0.005	0.013, 0.008	0.027, 0.018	0.062, 0.031	0.087, 0.054
Teardrops	0.002, 0.002	0.005, 0.003	0.009, 0.007	0.023, 0.015	0.058, 0.032
Other	0.014, 0.01	0.022, 0.015	0.036, 0.022	0.058, 0.037	0.098, 0.059