

# Physiological and pathological population dynamics of circulating human red blood cells

John M. Higgins<sup>a,b,1</sup> and L. Mahadevan<sup>b,c,1</sup>

<sup>a</sup>Center for Systems Biology and Department of Pathology, Massachusetts General Hospital, Boston, MA 02114; <sup>b</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115; and <sup>c</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Edited by Leo P. Kadanoff, The James Franck Institute, Chicago, IL, and approved October 6, 2010 (received for review August 29, 2010)

The systems controlling the number, size, and hemoglobin concentrations of populations of human red blood cells (RBCs) and their dysregulation in anemia, are poorly understood. After release from the bone marrow, RBCs undergo reduction in both volume and total hemoglobin content by an unknown mechanism [Lew VL, et al. (1995) *Blood* 86:334–341; Waugh RE, et al. (1992) *Blood* 79:1351–1358]; after ~120 d, responding to an unknown trigger, they are removed. We used theory from statistical physics and data from the hospital clinical laboratory [d’Onofrio G, et al. (1995) *Blood* 85:818–823] to develop a master equation model for RBC maturation and clearance. The model accurately identifies patients with anemia and distinguishes thalassemia-trait anemia from iron-deficiency anemia. Strikingly, it also identifies many pre-anemic patients several weeks before anemia becomes clinically detectable. More generally we illustrate how clinical laboratory data can be used to develop and to test a dynamic model of human pathophysiology with potential clinical utility.

hematology | mathematical modeling | medical sciences | stochastic differential equations | systems biology

In healthy human adults,  $\sim 2.5 \times 10^{11}$  new red blood cells (RBCs) are released from the bone marrow into the peripheral circulation per day, and about the same number are cleared. The cells composing the circulating population are thus continuously changing, but in healthy individuals (and patients with mild disease) the characteristics of the population are very stable. In the clinic, population characteristics such as the volume fraction of cells in the blood (hematocrit), the average RBC volume (MCV), the coefficient of variation in RBC volume (RDW), and the mean intracellular hemoglobin mass (MCH) are routinely measured in complete blood counts (CBCs) (1). Recently, it has become possible to identify and characterize very young (hours or days old) circulating RBCs (reticulocytes) (2). RBCs undergo a rapid reduction in volume and hemoglobin in the few days after release from the bone marrow (3). This rapid phase is followed by a much longer period of slower reduction (4–7) during which volume and hemoglobin are coregulated (8) (Fig. 1A). A comparison of the probability distributions of reticulocytes and of all circulating RBCs (Fig. 1A) shows that the correlation between volume and hemoglobin content increases as the cells mature, from an initial correlation coefficient of  $\sim 0.40$  in the reticulocyte population to  $\sim 0.85$  in the full population. Thus, whereas many of the molecular mechanisms involved are unknown, it is clear that the average RBC matures in such a way that its hemoglobin concentration tends toward the population mean corpuscular hemoglobin concentration (MCHC), shown as an iso-concentration line in Fig. 1A and B: the variation in hemoglobin concentration is lower than that for volume and hemoglobin content (8).

## Results

The volume and hemoglobin regulation of an individual RBC in vivo during the course of its lifetime is extremely complex and difficult to understand. Understanding the average behavior of a large population of RBCs may be more tractable. To gain insight into this population-level behavior, we developed a model of RBC maturation and clearance that describes the dynamics of an RBC population. Our model applies only to the typical behavior

of RBCs in healthy individuals or those with mild anemia and may not apply to RBCs in more severe disease or even to extreme behavior of RBCs in healthy individuals. Our model decomposes the volume ( $v$ ) and hemoglobin ( $h$ ) dynamics of an average RBC over time ( $t$ ) into deterministic reductions ( $\mathbf{f}$ ) and random fluctuations ( $\zeta$ ) as shown in Eq. 1, where  $v$  and  $h$  are scaled by their population means ( $\bar{v}, \bar{h}$ ), and  $t$  is scaled by the average cell age ( $\bar{\tau}$ ). On the basis of data from prior reports (3–5, 7), we introduced two parameters into the deterministic component: a fast change ( $\beta$ ), whose effect dominates until the RBC is close to the MCHC line, and a slow change ( $\alpha$ ). The random fluctuation is modeled as a Gaussian random variable with mean zero and variance given by a diffusion tensor  $2\mathbf{D}$ , as shown in Fig. 1B and Eq. 1:

$$\begin{bmatrix} \frac{dv}{dt} \\ \frac{dh}{dt} \end{bmatrix} = \mathbf{f} + \zeta \quad [1]$$

$$\mathbf{f} = \begin{cases} \alpha \cdot e^{\beta v} (v-h) \\ \alpha \cdot e^{\beta h} (h-v) \end{cases}$$

$$\zeta = \begin{cases} N(0, 2D_v) \\ N(0, 2D_h) \end{cases}$$

As with inverse problems in general and human pathophysiology in particular (9), this problem is ill posed in the sense that similar in vivo dynamics can result from different functional forms of  $\mathbf{f}$ . We find that the precise functional form of  $\mathbf{f}$  is less important to the behavior of this model than the qualitative combination of fast and slow deterministic dynamics and random fluctuations. See *SI Text (Functional Forms, Figs. S1–S4, and Table S1)* for details.

In our model, the random fluctuation and deterministic dissipation or reduction of volume and hemoglobin content for a typical individual cell are described by a Langevin equation, commonly used to model Brownian motion in a potential (10). The dynamics of the entire circulating population of RBCs may then be described by a master equation for the time-dependent joint volume–hemoglobin probability distribution ( $P(v, h, t)$ ) which can be approximated by a Fokker–Planck equation (10, 11). Eq. 2 describes the drift ( $\mathbf{f}$ ), diffusion ( $\mathbf{D}$ ), birth ( $b$ ), and death ( $d$ ) of probability density for this joint volume–hemoglobin distribution:

$$\frac{\partial P}{\partial t} = -\nabla \cdot (\mathbf{P}\mathbf{f}) + \nabla \cdot (\mathbf{D} \cdot \nabla P) + b(v, h, t) - d(v, h, t)P \quad [2]$$

Author contributions: J.M.H. and L.M. designed research; J.M.H. performed research; J.M.H. and L.M. analyzed data; and J.M.H. and L.M. wrote the paper.

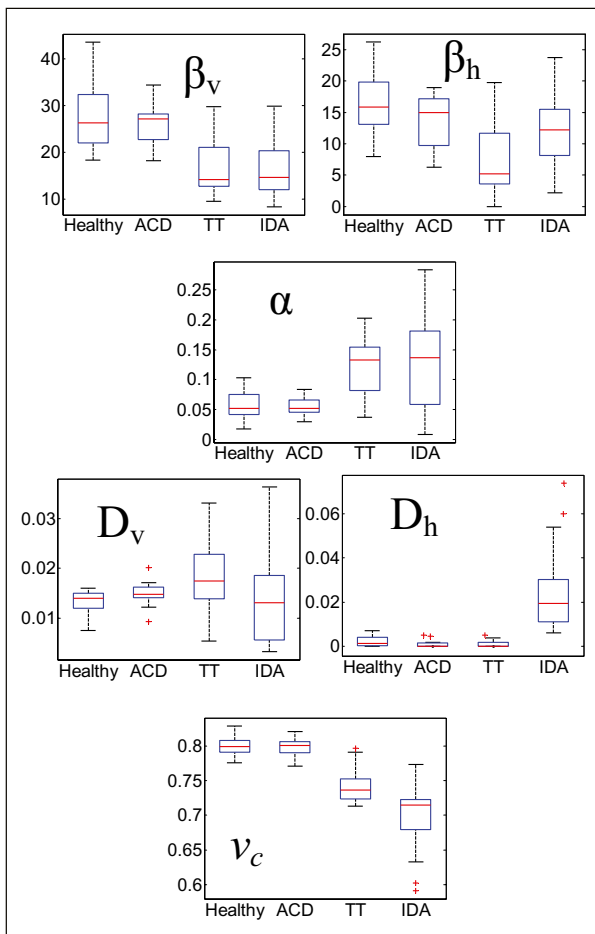
Conflict of interest statement: J.M.H. and L.M. are listed as inventors on a patent application related to this manuscript submitted by their institutions.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence may be addressed. E-mail: john\_higgins@hms.harvard.edu or lm@seas.harvard.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012747107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012747107/-DCSupplemental).





**Fig. 2.** Boxplots of model parameters for 20 healthy individuals and patients with three forms of mild anemia: 11 with anemia of chronic disease (ACD), 33 with thalassemia trait (TT), and 27 with iron deficiency anemia (IDA). The upper and lower edges of each box are located at the 75th and 25th percentiles. The median is indicated by a horizontal red line. Vertical lines extend to data points whose distance from the box is <1.5 times the interquartile distance. More extreme data points are shown as red plus (+) symbols. The fast dynamics are characterized by  $\beta$ , the slow by  $\alpha$ , random fluctuations by  $D$ , and the clearance threshold by  $v_c$ .

( $P_{0.85}$ ). Fig. 3 A–C shows the evolution of one patient’s CBC from normal to latent IDA and ultimately to IDA. The CBC shown in Fig. 3B was clinically unremarkable, but the  $P_{0.85}$  was abnormal, predicting anemia that did not come to medical attention for 51 d. Fig. 3D shows values of  $P_{0.85}$  for 20 normal CBCs from patients who remained healthy and 20 normal CBCs from those who developed IDA between 30 and 90 d later. The value of  $P_{0.85}$  predicted IDA with a sensitivity of 75% and a specificity of 100%. See *Predicting Iron Deficiency Anemia in Materials and Methods* for details. Common existing approaches have a sensitivity of 0% because all CBCs were “normal.” This

model-based prediction relies on only a single CBC measurement at one point in time, in contrast to statistical regression approaches that often rely on the integration of multiple measurements and types of information from different sources and different time points (15). IDA is often the initial presentation in serious conditions such as colon cancer (16) and childhood malnutrition (17). Earlier detection of IDA via this method would enable a faster response to such conditions. If a reduced  $v_c$  represents an adaptive physiologic response to offset the anemia, then perhaps one would expect the normal  $v_c$  seen here for ACD, where the anemia itself may represent an adaptive physiologic response (18).

Fig. 2 shows that  $D_h$  differentiates IDA and TT, the two most common causes of microcytic anemia. TT is one of the most commonly screened conditions in the world, but existing diagnostic methods either are too expensive or have unacceptably low diagnostic accuracy, with false positive rates of up to 30% (19, 20). Our model provides a possibly more accurate way to distinguish these conditions. We first established a threshold value for  $D_h$  of 0.0045 by analyzing 10 training samples. We then analyzed 50 independent patient samples where diagnosis of either mild IDA or TT could be confidently established and calculated  $D_h$ . Fig. 4 shows that this  $D_h$  threshold had a diagnostic accuracy of 98%, correctly identifying 22/22 cases of IDA and 27/28 cases of TT and outperforming other published approaches by between 6% and 41% (20). See *Diagnosis of Microcytic Anemia in Materials and Methods* for details.

Our mathematical model, although simple in concept, captures important features of the complex pathophysiology of circulating RBCs and allows us to identify quantitative differences in the dynamics of health and disease, using readily available clinical measurements. This model also provides a useful framework for thinking about the dynamical processes governing physiological homeostasis and shows how an understanding of these dynamics motivates diagnostic tests and generates pathophysiological hypotheses.

## Materials and Methods

**Blood Samples.** Blood samples and CBC results were obtained from the clinical laboratory of a tertiary care adult hospital under a research protocol approved by the Partners Healthcare Institutional Review Board. Reticulocyte and CBC measurements were made within 6 h of collection (21) on a Siemens Advia 2120 automated hemanalyzer.

IDA was defined as mild reduction in hematocrit to no more than 20% below the lower limit of normal, low MCV, low ferritin, and historical evidence of normal MCV with normal hematocrit. Patients with acute illness, acute bleeding, transfusion in the prior 6 mo, concurrent hospitalization, chronic inflammatory illness, or hemoglobinopathy were excluded.

TT was defined as either a high hemoglobin A2 or heterozygosity for the presence of an  $\alpha$ -globin gene mutation, as well as a reduction in hematocrit to no more than 20% below the lower limit of normal, low MCV, and normal ferritin. Patients with acute illness, acute bleeding, transfusion in the prior 6 mo, concurrent hospitalization, chronic inflammatory illness, or additional hemoglobinopathy were excluded.

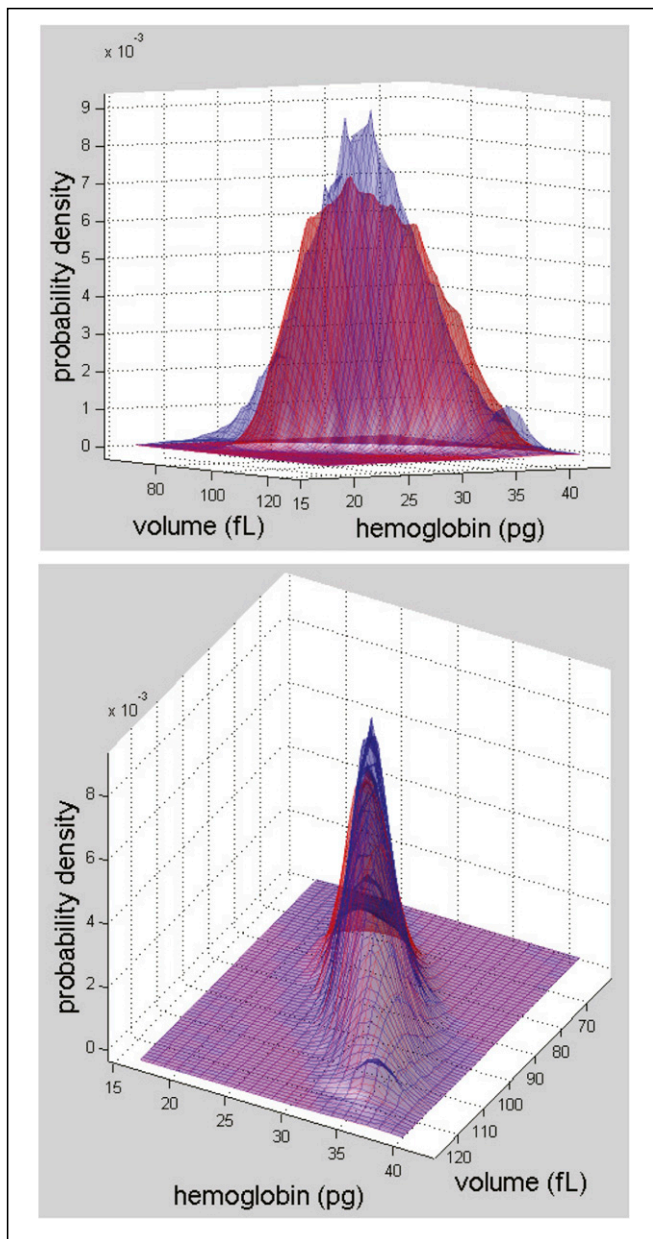
ACD was defined as a reduction in hematocrit to no more than 20% below the lower limit of normal, normal or high ferritin, and a low or normal total iron binding capacity. Patients with acute illness, acute bleeding, transfusion in the prior 6 mo, concurrent hospitalization, or hemoglobinopathy were excluded.

**Table 1.** Median values of nondimensional and dimensional fitted parameters (where appropriate)

	Normal	ACD	TT	IDA
$\beta_v$	26	27	14	15
$\beta_h$	16	15	5	12
$\alpha$	0.05 (0.09 fL/d and 0.03 pg/d)	0.05 (0.09 fL/d and 0.03 pg/d)	0.13 (0.20 fL/d and 0.07 pg/d)	0.13 (0.20 fL/d and 0.07 pg/d)
$D_v$	0.014 (2.3 fL <sup>2</sup> /d)	0.015 (2.4 fL <sup>2</sup> /d)	0.017 (2.2 fL <sup>2</sup> /d)	0.013 (1.7 fL <sup>2</sup> /d)
$D_h$	0.0014 (0.025 pg <sup>2</sup> /d)	$2.7 \times 10^{-5}$ ( $4.9 \times 10^{-4}$ pg <sup>2</sup> /d)	$2.7 \times 10^{-15}$ ( $3.6 \times 10^{-14}$ pg <sup>2</sup> /d)	0.019 (0.34 pg <sup>2</sup> /d)
$v_c$	0.80 (72 fL)	0.80 (72 fL)	0.74 (59 fL)	0.71 (56 fL)







**Fig. 6.** Comparison of fitted (red) and empirical (blue) steady-state joint volume–hemoglobin content probability distributions for a healthy individual. (Upper) The view projected on a vertical plane through the MCHC line (Fig. 1). (Lower) Ninety degree rotated view looking toward the origin.

empirically measured reticulocyte distribution. The volume–hemoglobin content plane was discretized with a constant mesh width and represented as a vector ( $\mathbf{P}$ ) of variables equal to the probability density in each mesh cell. Simulation results reported here were executed with a mesh width of 1.8 fL along the volume axis and 1.8 pg along the hemoglobin content axis. This mesh width was comparable to the analytic resolution of the empirical volume and hemoglobin content measurements. All results were confirmed using a smaller mesh width (1.2 fL and 1.2 pg). The convection contribution ( $\mathbf{f}$ ) was modeled numerically using an upwind finite difference approximation of the spatial derivative. The resulting linear system of ordinary differential equations was integrated using the MATLAB ode15s integrator and iterated until a steady state ( $\mathbf{P}_\infty$ ) was reached.

The steady-state distribution for the numerical problem ( $\mathbf{P}_\infty$ ) was also determined analytically by noting that the numerical approximations of the evolution ( $\mathbf{J} + \mathbf{L}$ ) and clearance ( $\mathbf{d}$ ) terms are linear operators and that the integral scaling the birth process is a constant equal to the reciprocal of twice the

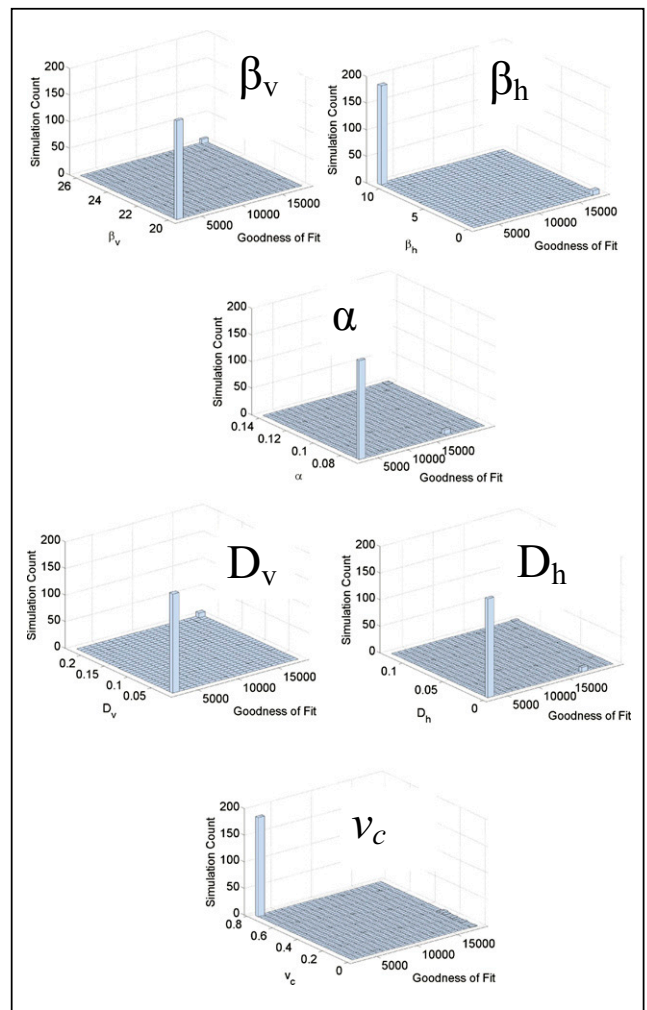
mean age ( $1/2\tau$ ). The linear operators can be inverted, yielding a family of steady-state distributions indexed by the mean cell age, as shown in Eq. 4:

$$\begin{aligned} \frac{d\mathbf{P}_\infty}{dt} = 0 &= -\mathbf{J} \cdot \mathbf{P}_\infty + \mathbf{L} \cdot \mathbf{P}_\infty + \mathbf{d} \cdot \mathbf{P}_\infty + \mathbf{P}_0 \int_h \int_v d(v, h) \\ &= (-\mathbf{J} + \mathbf{L} + \mathbf{d}) \cdot \mathbf{P}_\infty + \mathbf{P}_0 \frac{1}{2\tau} \Rightarrow \mathbf{P}_\infty = -(-\mathbf{J} + \mathbf{L} + \mathbf{d})^{-1} \mathbf{P}_0 \frac{1}{2\tau}. \end{aligned} \quad [4]$$

We found negligible differences between the steady-state distributions obtained by each approach.

**Parameter Fits.** We identified optimal neighborhoods in parameter space for each patient, using gradient and nongradient optimization methods. We started with a patient’s empirically measured reticulocyte distribution and an initial randomly chosen parameter set. We calculated the resulting steady-state RBC distribution using Eq. 4. This calculated distribution ( $\mathbf{P}_\infty$ ) was then compared with the empirical distribution ( $\mathbf{P}_{\text{CBC}}$ ). The quality of the parameter estimates was quantified by computing an objective function equal to the sum of the normalized squared residuals for the discretized distributions, as shown in Eq. 5, where  $i$  and  $j$  represent indices of cells in the discretized volume–hemoglobin plane:

$$C(\mathbf{P}_{\text{CBC}}^{i,j}, \mathbf{P}_\infty^{i,j}) = \sum_{i,j} \frac{(P_{\text{CBC}}^{i,j} - P_\infty^{i,j})^2}{P_{\text{CBC}}^{i,j}}. \quad [5]$$



**Fig. 7.** Histograms of optimized parameter values from >200 separate simulations and the goodness of fit as determined by a sum of squared residuals objective function. Smaller values of the objective function signify a better fit. All parameters have well-defined optimal neighborhoods. See Eq. 5 for objective function.

New parameter values were then chosen to reduce this objective function. We used gradient-based (lsqnonlin function) and non-gradient-based (fminsearch and patternsearch functions) optimization algorithms in MATLAB to search for optimal parameters. We constrained all parameters to be nonnegative and defined a uniform initial parameter space to exclude mean cell ages  $>1,000$  or  $<5$  d. We then picked initial parameters from this space, using latin hypersquare sampling. We imposed one optimization constraint, limiting  $\alpha$  to be small enough that the mean cell age would be  $>5$  d and large enough that the mean cell age would  $<1,000$  d. Fig. 6 shows that the model provides a faithful reproduction of  $P_{CBC}$  for this healthy patient by comparing the simulated and measured steady-state probability distributions. When projected along the MCHC line, the empirical distribution for this patient has slightly higher density near the mode and slightly lower density in the regions to either side of the mode. See [Movie S2](#) for more detail.

Fig. 7 shows the smallest local minima for parameters obtained from  $>200$  optimizations for a single patient. Some results had local minima above the range of the axes. The best fits among all simulations form a small neighborhood of values for all parameters, demonstrating that the parameter estimation process reached a well-defined optimal neighborhood for this patient's blood sample.

**Predicting Iron Deficiency Anemia.** We tested the hypothesis that compensated or latent IDA can be predicted up to 90 d earlier than is currently possible on the basis of an expanding population of cells that project along the MCHC line closer to the origin than the mean. The projection operation is pictured in Fig. 5. We first determined the projected position ( $u$ ) of each cell with volume ( $v$ ) and hemoglobin ( $h$ ) along the MCHC line:

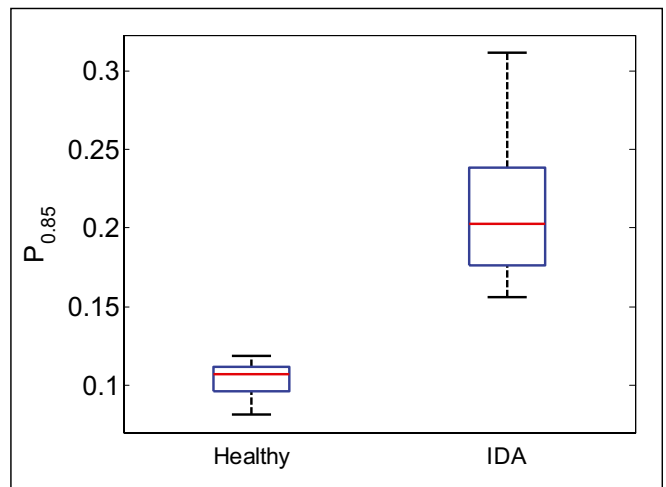
$$\begin{aligned} u &= v \cdot \cos\theta - h \cdot \sin\theta \\ \theta &= -\tan^{-1}\left(\frac{h}{v}\right). \end{aligned}$$

We then calculated the fraction of projected cells located between the origin and 85% of  $\bar{u}$ , where  $f_{MCHC}$  is the probability density of the projected cells as a function of location on the MCHC line:

$$P_{0.85} = \int_0^{0.85\bar{u}} f_{MCHC}(u) du.$$

The 85% threshold was chosen by comparing the discrimination efficiency of different thresholds for the steady-state CBCs used in Fig. 2. The 85% threshold shown in Fig. 8 provided the greatest separation when compared with other thresholds (70, 75, 80, 90, and 100%). We selected a threshold value for  $P_{0.85}$  of 0.121 on the basis of this training set.

We then identified 40 new and independent patient CBCs, all of which were normal. Twenty of these normal CBCs came from individuals who had additional normal CBCs 30–90 d later, and 20 of these normal CBCs came from individuals who presented with IDA no more than 90 d later. Patients with acute bleeding or any iron supplementation between the two CBCs were excluded. See *Blood Samples* in this section for definition of IDA. Fig. 3 shows that the threshold of 0.121 for  $P_{0.85}$  is able to predict IDA with a sensitivity of 75% and a specificity of 100% in this independent test group.



**Fig. 8.** Identifying a threshold for latent IDA. This boxplot shows the distributions of  $P_{0.85}$  for the steady-state healthy and IDA patients shown in Fig. 2. On the basis of these results, we picked a threshold value for  $P_{0.85}$  of 0.121 to use in a test of an independent set of patients shown in Fig. 3.

**Diagnosis of Microcytic Anemia.** To assess the diagnostic accuracy of  $D_h$  in differentiating the two most common causes of microcytic anemia, we fit parameters for 10 training cases: 5 with IDA and 5 with TT, as shown in Fig. 4. See *Blood Samples* for case definitions. The TT training set had  $D_h$  ranging from  $1.7 \times 10^{-15}$  to  $2.3 \times 10^{-15}$ . The IDA training set had  $D_h$  ranging from 0.009 to 0.043. We picked a threshold value equal to 0.0045, the average of the lowest  $D_h$  among the IDA training set and the highest of the TT training set. We then identified 50 new and independent cases, 22 with IDA and 28 with TT. All 22 IDA cases were correctly classified, and 27/28 TT cases were correctly classified for an overall diagnostic accuracy of 98%. We find this accuracy to be superior to that of four commonly cited discriminant functions (20): The Green and King formula yielded an accuracy of 92%; Micro/Hypo, 84%; Mentzer, 68%; and England and Fraser, 57%. See ref. 20 for further details on these other discriminant functions.

**ACKNOWLEDGMENTS.** We thank Alicia Soriano, Svetla Patel, Rosy Jaromin, Hasmukh Patel, David Dorfman, Paul Charest, the Harvard Crimson Tissue Bank, and the Partners Research Patient Data Repository for help acquiring samples and data. We thank Ricardo Paxson, Madhusudhan Venkadesan, and Shreyas Mandre for useful suggestions on the analysis; Frank Bunn, Carlo Brugnara, and Steve Sloan for helpful discussions regarding the findings; and Rebecca Ward for critical reading of the manuscript. The simulations in this paper were primarily run on the Orchestra cluster supported by the Harvard Medical School Research Information Technology Group. J.M.H. acknowledges support from National Institute of Diabetes and Digestive and Kidney Diseases Grant DK083242. L.M. acknowledges support from National Heart, Lung, and Blood Institute Grant HL091331.

- Fauci AS (2008) *Harrison's Principles of Internal Medicine*, eds Fauci AS, et al. (McGraw-Hill Medical, New York), 17th Ed.
- d'Onofrio G, et al. (1995) Simultaneous measurement of reticulocyte and red blood cell indices in healthy subjects and patients with microcytic and macrocytic anemia. *Blood* 85:818–823.
- Gifford SC, Derganc J, Shevkoplyas SS, Yoshida T, Bitensky MW (2006) A detailed study of time-dependent changes in human red blood cells: From reticulocyte maturation to erythrocyte senescence. *Br J Haematol* 135:395–404.
- Waugh RE, et al. (1992) Rheologic properties of senescent erythrocytes: Loss of surface area and volume with red blood cell age. *Blood* 79:1351–1358.
- Willekens FL, et al. (2003) Hemoglobin loss from erythrocytes in vivo results from spleen-facilitated vesiculation. *Blood* 101:747–751.
- Sens P, Gov N (2007) Force balance and membrane shedding at the red-blood-cell surface. *Phys Rev Lett* 98:018102.
- Willekens FL, et al. (2005) Liver Kupffer cells rapidly remove red blood cell-derived vesicles from the circulation by scavenger receptors. *Blood* 105:2141–2145.
- Lew VL, Raffos JE, Sorette M, Bookchin RM, Mohandas N (1995) Generation of normal human red cell volume, hemoglobin content, and membrane area distributions by "birth" or regulation? *Blood* 86:334–341.
- Zenker S, Rubin J, Clermont G (2007) From inverse problems in mathematical physiology to quantitative differential diagnoses. *PLoS Comput Biol* 3:e204.
- Zwanzig R (2001) *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, UK).
- Kampen NGv (1992) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
- Lang KS, et al. (2005) Mechanisms of suicidal erythrocyte death. *Cell Physiol Biochem* 15:195–202.
- Garner C, et al. (2000) Genetic influences on F cells and other hematologic variables: A twin heritability study. *Blood* 95:342–346.
- Robbins SL, Kumar V, Cotran RS (2010) *Robbins and Cotran Pathologic Basis of Disease* (Saunders/Elsevier, Philadelphia), 8th Ed.
- Milbrandt EB, et al. (2006) Predicting late anemia in critical illness. *Crit Care* 10:R39.
- Rockey DC, Cello JP (1993) Evaluation of the gastrointestinal tract in patients with iron-deficiency anemia. *N Engl J Med* 329:1691–1695.
- Lozoff B, Jimenez E, Wolf AW (1991) Long-term developmental outcome of infants with iron deficiency. *N Engl J Med* 325:687–694.
- Zarychanski R, Houston DS (2008) Anemia of chronic disease: A harmful disorder or an adaptive, beneficial response? *CMAJ* 179:333–337.
- Jopang YP, Thinkhamrop B, Puangpruk R, Netnee P (2009) False positive rates of thalassemia screening in rural clinical setting: 10-year experience in Thailand. *Southeast Asian J Trop Med Public Health* 40:576–580.
- Ntaios G, et al. (2007) Discrimination indices as screening tests for beta-thalassemic trait. *Ann Hematol* 86:487–491.
- Lippi G, Salvagno GL, Solero GP, Franchini M, Guidi GC (2005) Stability of blood cell counts, hematologic parameters and reticulocytes indexes on the Advia A120 hematologic analyzer. *J Lab Clin Med* 146:333–340.