

Physiological and pathological population dynamics of circulating human red blood cells

John M. Higgins^{a,b,1} and L. Mahadevan^{b,c,1}

^aCenter for Systems Biology and Department of Pathology, Massachusetts General Hospital, Boston, MA 02114; ^bDepartment of Systems Biology, Harvard Medical School, Boston, MA 02115; and ^cSchool 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 ~ 0.40 in the reticulocyte population to ~ 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 (ζ) 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 (β), whose effect dominates until the RBC is close to the MCHC line, and a slow change (α). 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.

¹To 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.

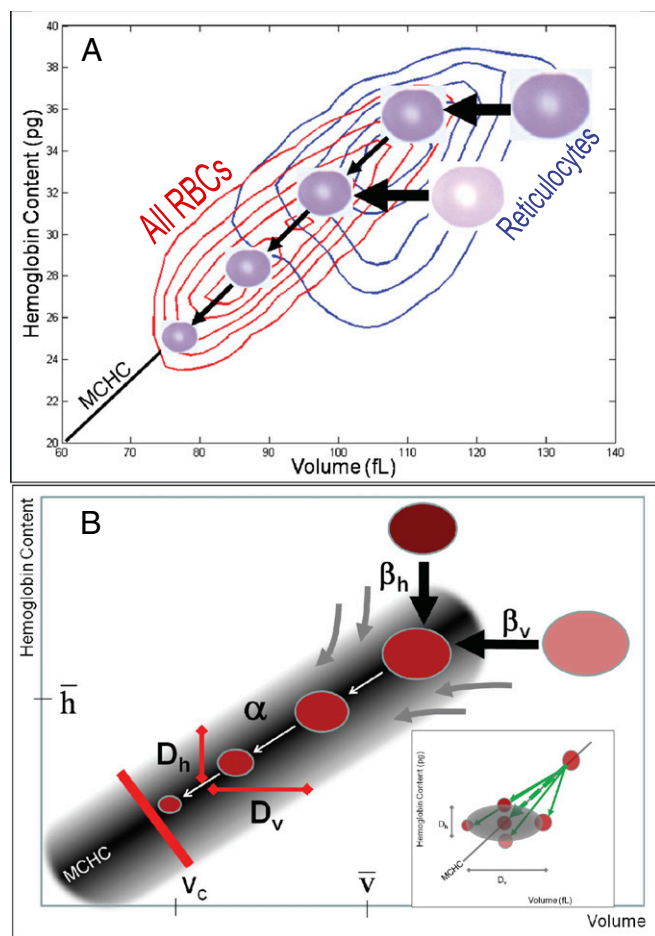


Fig. 1. Empirical measurement (A) and dynamic model (B) of coregulation of volume and hemoglobin of an average RBC in the peripheral circulation. The reticulocyte distribution is shown as blue iso-probability density contours and the population of all RBCs as red. The diagonal line projecting to the origin in A and B represents the average intracellular hemoglobin concentration (MCHC) in the population. An RBC located anywhere on this line will have an intracellular hemoglobin concentration equal to the MCHC. Fast dynamics (β) first reduce volume and hemoglobin for the typical large immature reticulocytes shown in the upper right of A and B. Slow dynamics (α) then reduce volume and hemoglobin along the MCHC line. Because biological processes are inherently noisy, we suggest that small random variations during the events required for reduction of volume and hemoglobin cause individual cellular hemoglobin concentrations to drift about the MCHC line, fluctuating with magnitude (D) around the MCHC line as shown in the *Inset* in B until reaching a critical volume (v_c in B) when cells are removed.

$$\mathbf{D} = \begin{bmatrix} D_v & 0 \\ 0 & D_h \end{bmatrix}.$$

The birth and death processes account for the RBCs that are constantly added to and removed from the population. In states of health and mild illness, the total number of cells added equals the total number removed: $\int \int d(v, h) P dv dh = \int \int b(v, h) dv dh$. The precise trigger and mechanism for RBC removal are not fully understood (12), but empirical measurements such as those shown in Fig. 1A suggest that there is a threshold (v_c) along the MCHC line beyond which most RBCs have been cleared. See *Clearance Function* in *Materials and Methods* for details. CBC measurements vary from person to person but do not change significantly for a healthy individual (13), indicating that these dynamic processes reach a steady state in vivo, P_∞ , where $\lim_{t \rightarrow \infty} P(v, h, t) \rightarrow P_\infty(v, h)$, i.e., $\partial P_\infty / \partial t = 0$. See *Movie S1* for an

example of a numerical solution to Eq. 2. With appropriate choice of parameters (α , β , \mathbf{D} , and v_c), our model faithfully reproduces the observed distribution of RBC populations in healthy individuals. See *Movie S2* for a comparison of a fitted and an observed distribution.

To test whether the model can distinguish the dynamics of RBC populations in healthy individuals from those in anemic individuals, we obtained CBC and reticulocyte measurements for individuals with three common forms of anemia with different underlying etiologies: anemia of chronic disease (ACD), an inflammatory condition; thalassemia trait (TT), a genetic disorder; and iron deficiency anemia (IDA), a nutritional condition (14). We selected mild cases of each anemia where RBC population characteristics appeared stable and a quasi-steady-state assumption was reasonable; we also selected several apparently healthy controls. See *Blood Samples in Materials and Methods* for details on sample collection, data collection, and diagnosis. For each patient sample, we identified an optimal parameter set (α , β , \mathbf{D} , and v_c) that reproduces the steady state observed for that patient. We used a least-squares fit between the simulated steady-state distribution and the measured CBC distribution to identify the best fit. Where repeat tests were available for the same individual, we found that any variation in fitted parameters was explained by analytic variation in the CBC measurement. See *Parameter Fits in Materials and Methods* and *Movie S2* for details. Fitted parameters for healthy and anemic individuals are shown in Fig. 2 and median fitted parameters are listed in Table 1.

We find clear differences between the best-fit parameters derived for healthy individuals and those with anemia, and we also find that the different anemic conditions have different characteristic parameter sets. For example, healthy individuals and ACD patients have high β_v and β_h and low α ; i.e., they lose relatively more of their volume and hemoglobin during the fast phase than they do during the slow phase. In contrast, patients with TT and IDA lose relatively more volume and hemoglobin during the slow phase than in the fast phase. Patients with ACD show slightly elevated D_v and slightly reduced D_h relative to healthy individuals, whereas TT is associated with a larger increase in D_v along with a substantially reduced D_h . IDA patients have a D_v similar to that of healthy individuals and show dramatic elevation in D_h with most individuals >10 times higher than normal. The normalized critical volume, v_c , in healthy individuals and those with ACD is $\approx 80\%$ of \bar{v} , or ~ 72 fL. Most patients with TT or IDA typically have a reduced \bar{v} and reduced \bar{h} . Fig. 2 shows that in addition to absolute reductions in \bar{v} and \bar{h} , the v_c for these patients is further reduced and shows much greater variability across different individuals.

Discussion

These parameters show that RBCs in patients with TT and IDA remain in the periphery with much smaller volumes and lower hemoglobin contents, in both absolute and relative terms, than they would under normal conditions. Their persistence may reflect a compensatory delay in clearance in response to the less efficient erythropoiesis of these anemias. Thus, mechanisms must exist that can alter the behavior of the trigger for RBC clearance. Comparing RBC clearance in TT and IDA with that of healthy individuals or ACD patients may provide a route to identifying the trigger. The variation in v_c is much smaller than the variation in \bar{v} for healthy individuals (14), suggesting that the trigger is highly correlated with position on the MCHC line.

Our model offers a potential way to identify patients with latent or compensated IDA before it leads to clinical anemia, by looking for signs of clearance delay. We tested this possibility in an independent set of patients who had normal CBCs followed at least 30 and no more than 90 d later by either another normal CBC or clinical IDA. For each patient sample, we projected the (v, h) coordinates of all cells onto the MCHC line and integrated the probability density along this line below 85% of the mean

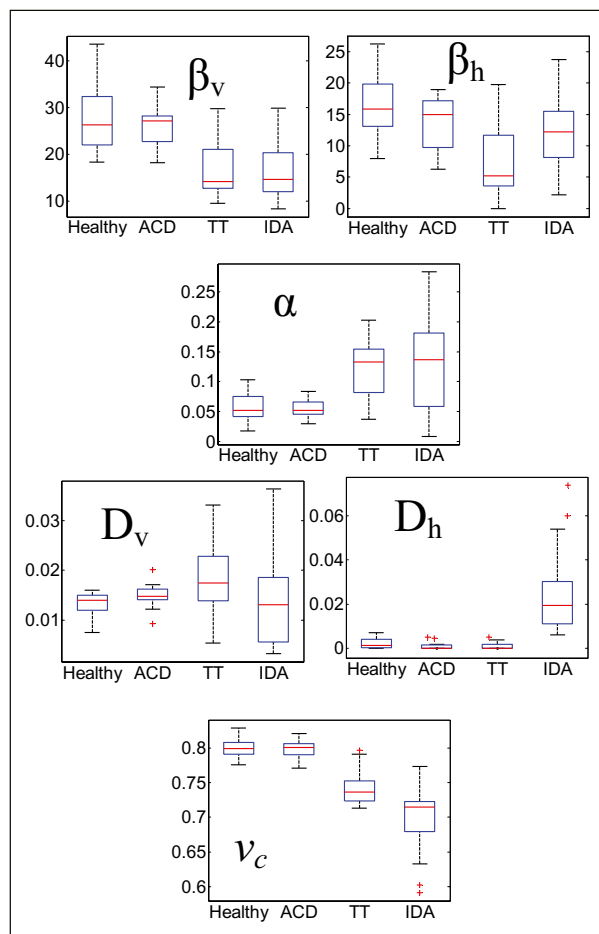


Fig. 2. Boxplots of model parameters for 20 healthy individuals and patients with three forms of mild anemia: 11 with anemia of chronic disease (ACD), 33 with thalassemia trait (TT), and 27 with iron deficiency anemia (IDA). The upper and lower edges of each box are located at the 75th and 25th percentiles. 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 β , the slow by α , 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 α -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
β_v	26	27	14	15
β_h	16	15	5	12
α	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 ² /d)	0.015 (2.4 fL ² /d)	0.017 (2.2 fL ² /d)	0.013 (1.7 fL ² /d)
D_h	0.0014 (0.025 pg ² /d)	2.7×10^{-5} (4.9×10^{-4} pg ² /d)	2.7×10^{-15} (3.6×10^{-14} pg ² /d)	0.019 (0.34 pg ² /d)
v_c	0.80 (72 fL)	0.80 (72 fL)	0.74 (59 fL)	0.71 (56 fL)

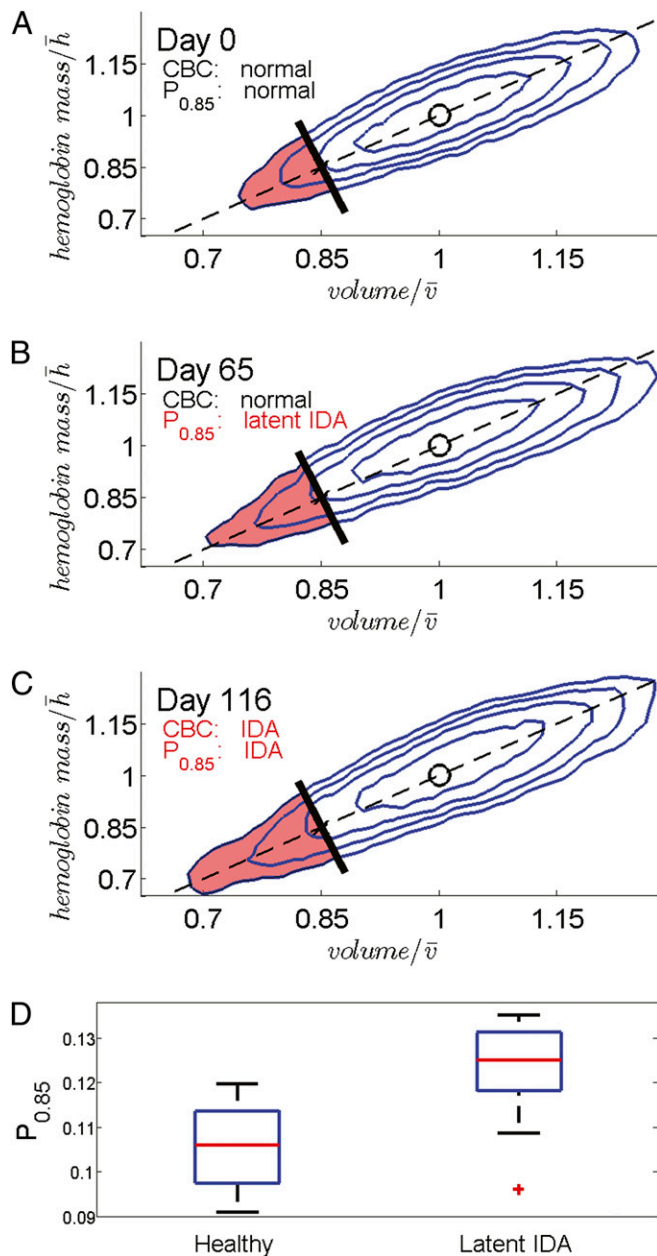


Fig. 3. Contour plots (A–C) of CBCs for a patient developing IDA after 4 mo. Each plot shows contours enclosing 35, 60, 75, and 85% of the probability density. The dashed line from the origin represents the MCHC. The short solid line perpendicular to the dashed line marks the position along the line corresponding to 85% of the mean projected cell. The circle shows the mean projection. *A* shows a normal CBC measured 116 d before the patient's presentation with iron deficiency anemia. The calculated $P_{0.85}$ (red area) is normal. *B* shows the normal CBC measured 65 d later and 51 d before detection of iron deficiency anemia. $P_{0.85}$ is abnormal even though the CBC is normal. *C* shows the CBC at the time IDA was diagnosed. *D* shows boxplots of $P_{0.85}$ for 20 normal CBCs from patients with a second normal CBC within 90 d and 20 normal CBCs from patients who were diagnosed with IDA up to 90 d later. $P_{0.85}$ successfully predicts IDA up to 90 d earlier than is currently possible with a sensitivity of 75% and a specificity of 100%. See main text and *Predicting Iron Deficiency Anemia in Materials and Methods* for more detail.

Clearance Function. On the basis of observations of empirical RBC distributions, we model probability of RBC clearance as a function of the RBC volume and hemoglobin content. Each RBC's position in the volume-hemoglobin content plane is projected onto the MCHC line, and the probability of clearance (d) is defined as either a sigmoid or a step function of the

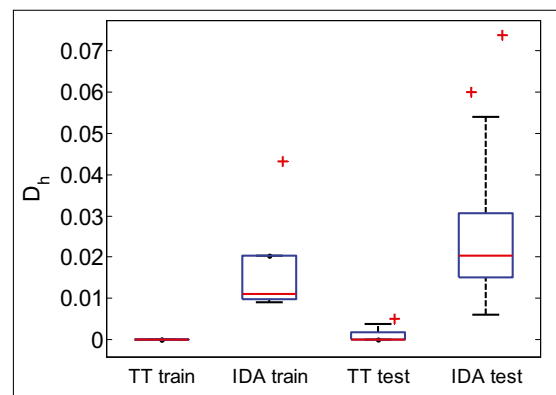


Fig. 4. Differentiating TT and IDA as causes of microcytic anemia. This boxplot shows the distributions of D_h for 5 training cases with TT, 5 training cases with IDA, 28 test cases with TT, and 22 test cases with IDA.

distance from this projected point to a threshold, v_c , on this line (Figs. 1 and 5; Fig. S2). Eq. 3 quantifies this relationship:

$$d(v, h) = \frac{1}{1 + e^\Delta} \quad [3]$$

or

$$d(v, h) = \begin{cases} 1 & \Delta \leq 0 \\ 0 & \Delta > 0 \end{cases}$$

$$\Delta(v, h) = 100 \cdot \frac{\cos(\theta) \sqrt{(v\bar{v})^2 + (h\bar{h})^2} - v_c \sqrt{\bar{h}^2 + \bar{v}^2}}{v_c \sqrt{\bar{h}^2 + \bar{v}^2}}$$

$$\theta = \tan^{-1}\left(\frac{\bar{h}}{\bar{v}}\right) - \tan^{-1}\left(\frac{h\bar{h}}{v\bar{v}}\right).$$

Description of Simulation and Explicit Numerical Solution. For a given set of parameters, Eq. 2 was solved numerically using a finite difference approximation for first-order ($\mathbf{J} = \Delta_k[\mathbf{f} \cdot \mathbf{P}](v)/k + \Delta_k[\mathbf{f} \cdot \mathbf{P}](h)/k$) and second-order ($\mathbf{L} = D_v(s_k^2[P](v)/k^2) + D_h(s_k^2[P](h)/k^2)$) spatial derivatives with boundary conditions of vanishing probability at volumes and hemoglobin contents outside the pathophysiological range and initial conditions equal to the

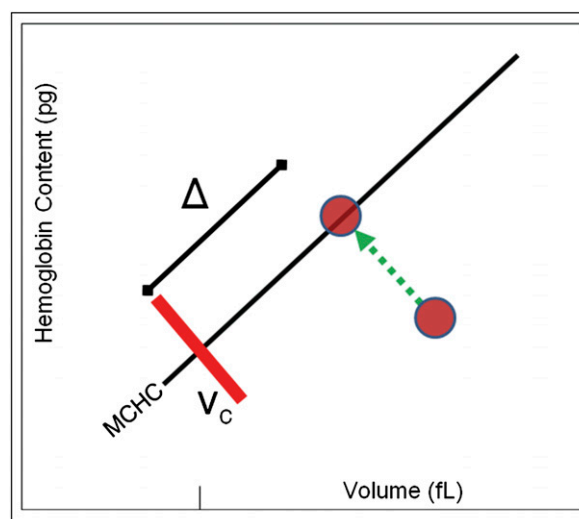


Fig. 5. Schematic of the projected distance (Δ) used to calculate the probability of clearance as described in Eq. 3. The cell is projected onto the MCHC line, and the probability of clearance is a function of the distance from the projected point to a threshold (v_c) along this line.

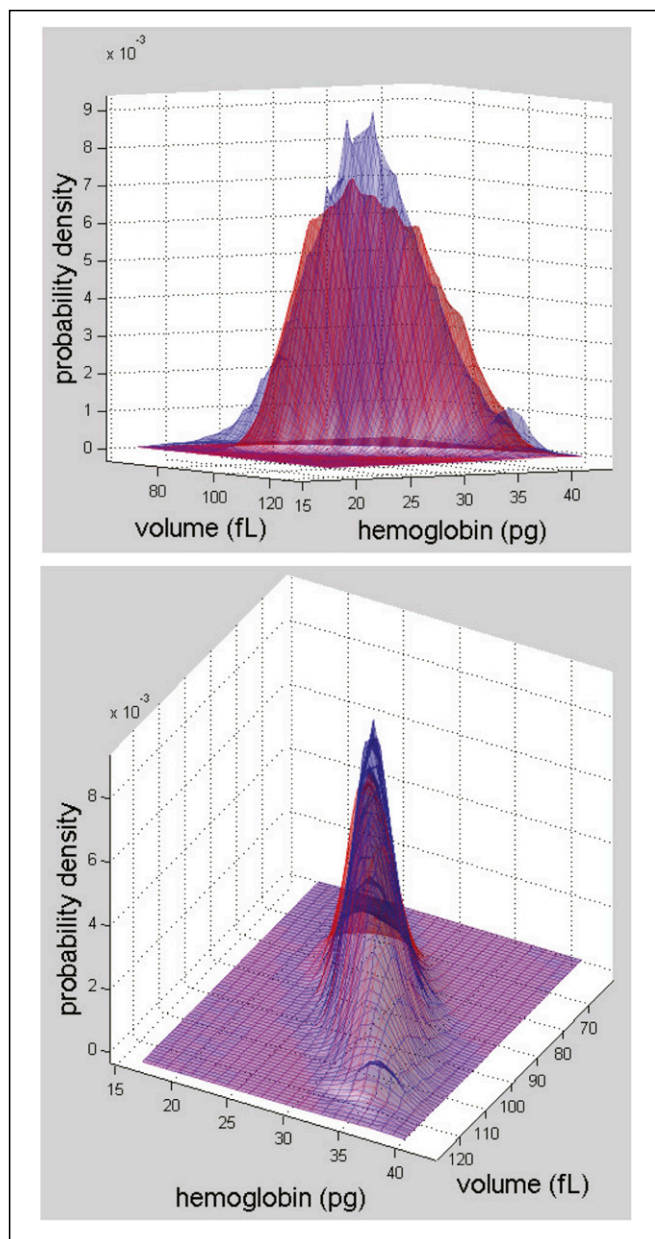


Fig. 6. Comparison of fitted (red) and empirical (blue) steady-state joint volume-hemoglobin content probability distributions for a healthy individual. (*Upper*) The view projected on a vertical plane through the MCHC line (Fig. 1). (*Lower*) Ninety degree rotated view looking toward the origin.

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}_{∞}) was reached.

The steady-state distribution for the numerical problem (\mathbf{P}_∞) 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 ($1/2\bar{\tau}$). The linear operators can be inverted, yielding a family of steady-state distributions indexed by the mean cell age, as show in Eq. 4:

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

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}_{∞}) was then compared with the empirical distribution (\mathbf{P}_{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:

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

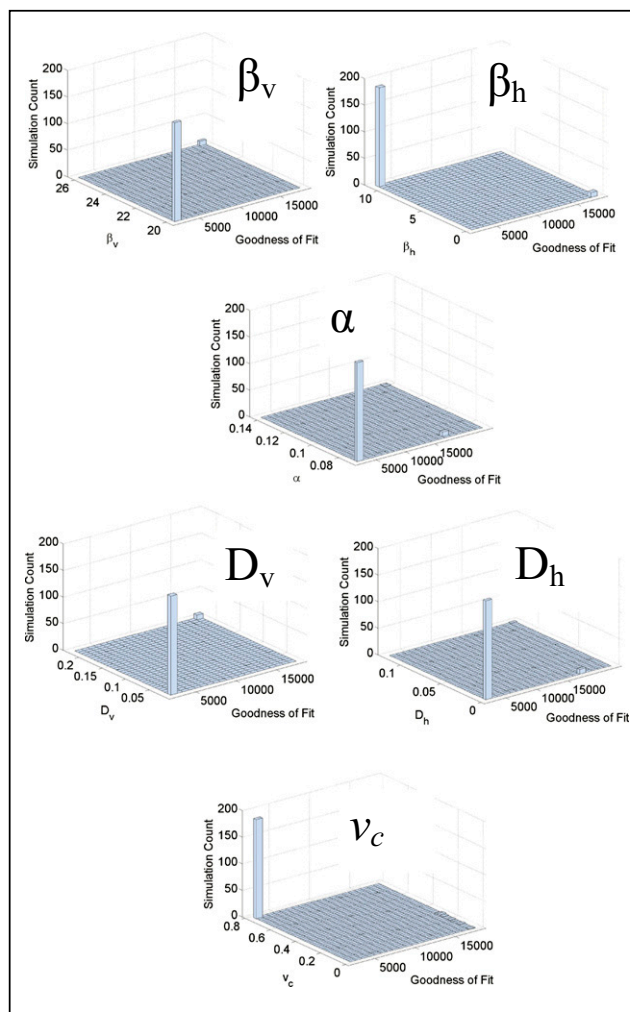


Fig. 7. Histograms of optimized parameter values from >200 separate simulations and the goodness of fit as determined by a sum of squared residuals objective function. Smaller values of the objective function signify a better fit. All parameters have well-defined optimal neighborhoods. See Eq. 5 for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

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

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.

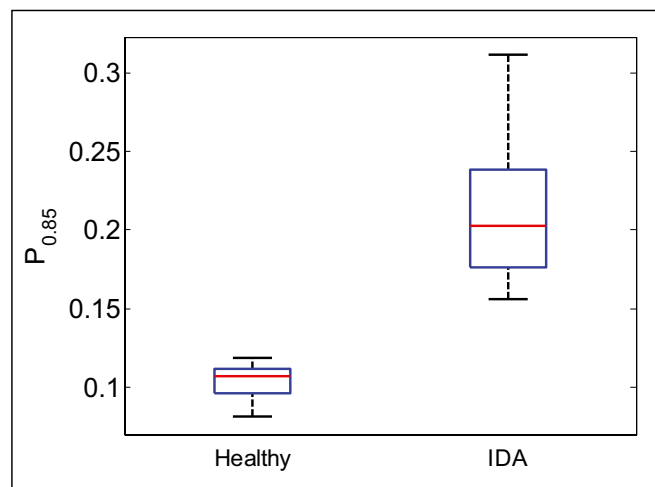


Fig. 8. Identifying a threshold for latent IDA. This boxplot shows the distributions of $P_{0.85}$ for the steady-state healthy and IDA patients shown in Fig. 2. On the basis of these results, we picked a threshold value for $P_{0.85}$ of 0.121 to use in a test of an independent set of patients shown 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 1.7×10^{-15} to 2.3×10^{-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, Shevokoplyas 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, Raftos 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.

CLINICAL IMPLICATIONS OF BASIC RESEARCH

Systems Biology and Red Cells

David J. Weatherall, M.D.

The remarkable advances in molecular and cell biology over the past half century have revealed layer upon layer of biologic complexity. It has even been suggested that we are reaching a stage at which further progress may be beyond the scope of human imagination, that these branches of science — like physics in its endless search for a unifying theory for the basis of the universe — may be at an impasse.¹ A more optimistic view suggests that if we apply the mathematical approaches of systems biology, many of these

problems may be solvable, although Sydney Brenner has criticized this approach, suggesting that we first try to understand the organization and regulatory biology of individual cells.²

The relevance of systems biology to the interplay between the basic and clinical medical sciences was recently highlighted in a report by Higgins and Mahadevan.³ They used a systems-biology approach to analyze the physiologic and pathologic population dynamics of circulating human red cells. The mechanisms that regulate the number, size, and hemoglobin concentration of normal red cells in circulation — and how these go awry in anemia — are not well understood. It has been established, however, that after their release from the bone marrow red cells undergo a reduction in their volume and total hemoglobin content. To approach this problem, Higgins and Mahadevan used theory from statistical physics together with standard red-cell indexes derived from electronic cell counters to develop a master equation for the maturation and clearance of red cells. Their mathematical model implies that the total number of red cells added to the circulation equals the number removed and suggests that there is a threshold for the mean cellular hemoglobin concentration below which most red cells are cleared from the circulation (Fig. 1).

Quite remarkably, this model appears to clearly distinguish the dynamics of red-cell populations in normal persons from those in persons with anemia of chronic disease, iron deficiency, or α - or β -thalassemia trait. Higgins and Mahadevan suggest that in persons with iron deficiency or thalassemia trait, the persistence of red cells with smaller volumes and lower hemoglobin content, which would not normally be retained, may reflect a compensatory delay in clearance as a result of the less efficient red-cell production characteristic of these anemias. This mechanism does not, however, explain the dif-

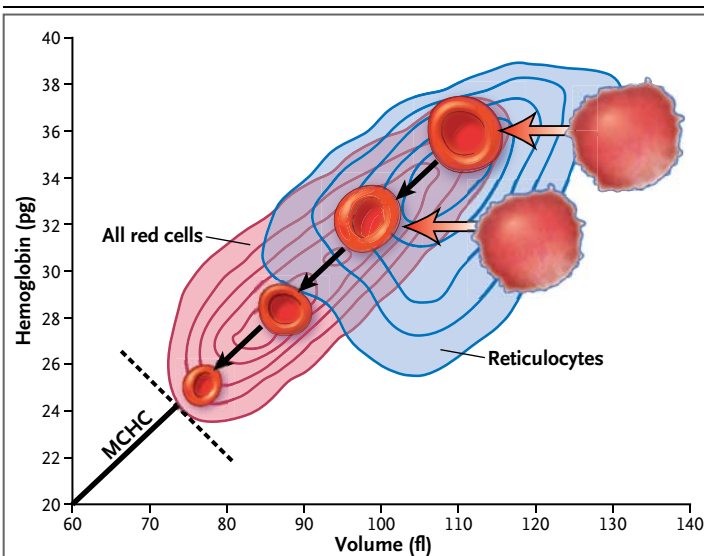


Figure 1. Dynamics of Red-Cell Volume and Hemoglobin Content.

Higgins and Mahadevan recently described in mathematical terms the mechanisms that regulate the volume and hemoglobin content of an average red cell in the peripheral circulation. The reticulocyte distribution is shown in blue and the isoprobability density contours and population of all red cells in red. The diagonal line projecting to the origin represents the average intracellular hemoglobin concentration in the population. A red cell located anywhere on this line will have a mean corpuscular hemoglobin concentration (MCHC) equal to the average concentration. The dotted black line indicates the point at which red cells are removed from the circulation in the physiologic state. Higgins and Mahadevan went on to describe how this threshold is altered in patients with iron deficiency or thalassemia trait, and how related measures may serve to differentiate these two causes of anemia. Adapted from Higgins and Mahadevan.³

ferences in red-cell abnormalities between the two conditions, nor does it explain the differences in abnormalities evident in the various genetic forms of α -thalassemia, an issue that Higgins and Mahadevan do not address.³ Although both the thalassemias and iron-deficiency anemia are characterized by small hypochromic red cells, the red-cell counts in persons with thalassemia trait tend to be normal or raised, whereas in persons with iron deficiency they tend to be reduced. Higgins and Mahadevan found that the degree of variability in the hemoglobin content of red cells is much greater in persons with iron-deficiency anemia than in those with thalassemia trait, which allows differentiation of the two conditions. The mechanisms for inefficient red-cell production in these two conditions are quite different, and there are also differences in terms of red-cell survival and turnover.⁴

Remarkably, the model described by Higgins and Mahadevan appears to be able to predict which patients are likely to become anemic because of iron deficiency in the near future, even though at present their hematologic profile is normal. This observation is also based on evidence of clearance delay.

What are the implications of these findings? They tell us nothing of the intracellular regulatory mechanisms governing the metabolism of red cells, or of the structural changes in hemoglobin or the membrane of red cells as they age during their 120-day sojourn in this unfavorable circulatory environment. However, the model developed by Higgins and Mahadevan certainly raises questions about these mechanisms. As for the potential clinical application of the model, the ability to distinguish between iron-deficiency anemia and thalassemia trait is absolutely cen-

tral to screening programs and to the process of micromapping the traits for different forms of thalassemia in large populations. A survey⁵ of several different discrimination indexes designed for this purpose in screening tests for thalassemia trait has suggested that none of the indexes are entirely satisfactory and certainly do not reach the level of discrimination described in this recent report. However, it remains to be seen whether this new approach will be as robust for use in screening in developing countries, where it is really needed and where the results are complicated by many other issues, including associated infection or other nutritional deficiencies. It is also noteworthy that these studies were carried out in adults. Would the model be equally successful if applied in early childhood, during the normal developmental changes that occur in red-cell indexes?

In short, although the study by Higgins and Mahadevan supports the view that the regulation of dynamic systems will be understood only by means of detailed analysis of the regulatory mechanisms in individual cells, it also shows how knowledge of the dynamics of biologic systems may turn out to have clinical application.

From the Weatherall Institute of Molecular Medicine, University of Oxford, and John Radcliffe Hospital, Headington, Oxford, United Kingdom.

1. Horgan J. The end of science: the limits of knowledge in the twilight of the scientific age. New York: Broadway Books, 1996.
2. Brenner S. Sequences and consequences. *Philos Trans R Soc Lond B Biol Sci* 2010;365:207-12.
3. Higgins JM, Mahadevan L. Physiological and pathological population dynamics of circulating human red blood cells. *Proc Natl Acad Sci U S A* 2010;107:20587-92.
4. Weatherall DJ, Clegg JB. The thalassaemia syndromes. 4th ed. Oxford, England: Blackwell Science, 2001:345-50.
5. Ntaios G, Chatzinikolaou A, Saouli Z, et al. Discrimination indices as screening tests for beta-thalassemic trait. *Ann Hematol* 2007;86:487-91.

Copyright © 2011 Massachusetts Medical Society.