# White blood cell population dynamics for risk stratification of acute coronary syndrome

Anwesha Chaudhury[a,b,c], Lorette Noiret[a,b,c], and John M. Higgins[a,b,c,1]

[a]Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114; [b]Department of Pathology, Massachusetts General Hospital, Boston, MA 02114; and [c]Department of Systems Biology, Harvard Medical School, Boston, MA 02115

Edited by Shu Chien, University of California, San Diego, La Jolla, CA, and approved September 28, 2017 (received for review June 3, 2017)

**The complete blood count (CBC) provides a high-level assessment of a patient's immunologic state and guides the diagnosis and treatment of almost all diseases. Hematology analyzers evaluate CBCs by making high-dimensional single-cell measurements of size and cytoplasmic and nuclear morphology in high throughput, but only the final cell counts are commonly used for clinical decisions. Here, we utilize the underlying single-cell measurements from conventional clinical instruments to develop a mathematical model guided by cellular mechanisms that quantifies the population dynamics of neutrophil, lymphocyte, and monocyte characteristics. The dynamic model tracks the evolution of the morphology of WBC subpopulations as a patient transitions from a healthy to a diseased state. We show how healthy individuals and hospitalized patients with similar WBC counts can be robustly classified based on their WBC population dynamics. We combine the model with supervised learning techniques to risk-stratify patients under evaluation for acute coronary syndrome. In particular, the model can identify more than 70% of patients in our study population with initially negative screening tests who will be diagnosed with acute coronary syndrome in the subsequent 48 hours. More generally, our study shows how mechanistic modeling of existing clinical data can help realize the vision of precision medicine.**

acute coronary syndrome | mathematical modeling | white blood cells | disease prognosis | population dynamics

Circulating blood cells continuously interrogate almost all tissues in high throughput, and their collective states of maturation, activation, proliferation, and senescence reflect current pathophysiologic conditions: healthy quiescence, acute response to pathology, chronic compensation for disease, and ultimately, decompensation. Routine complete blood counts (CBCs) involve measurements of single-cell characteristics for tens of thousands of blood cells and provide a high-level view of these pathophysiologic states. Each routine CBC measures high-dimensional single-cell information, but clinical decisions are currently based on only a few derived statistics. The vast potential of the complete set of CBC measurements has been well-appreciated, with previous efforts attempting early detection of infection by identifying immature granulocytes or prognosis for some malignancies by counting the number of WBCs with atypical characteristics (1–3). These efforts have had limited impact but hint at the potential for enhanced clinical decision support (4, 5).

Here, we develop mathematical models of the population dynamics of neutrophils, lymphocytes, and monocytes using these CBC measurements. To test the hypothesis that qualitative aspects of WBC population dynamics are altered in disease, we first compare WBC population dynamics in healthy individuals with those in patients with an acute illness requiring treatment in a hospital. We ensure that overall WBC counts are normal to investigate the effect of acute illness on WBC dynamics independent of its effects on absolute WBC count. We then focus on a common and important acute illness, acute coronary syndrome (ACS), one of the leading causes of death worldwide. ACS is caused by insufficient oxygen supply (severe ischemia) to the myocardium, resulting in either a myocardial infarction (MI) or unstable angina (6–9). MI is diagnosed by measuring cardiac troponin (cTn), a protein found in the cardiac myocytes that leaks into the blood stream after myocyte death. cTn appears in the blood only after a few hours or even days after the onset of symptoms, thus complicating rapid intervention for some patients (8). We hypothesize that the severe ischemia and developing infarction trigger an immediate and substantial inflammatory response that can be detected in terms of its perturbations to WBC dynamics, while patients experiencing unstable angina or other pathology not associated with MI will show WBC dynamics that are perturbed in a different manner or not at all.

## Model Derivation

We begin with a biological interpretation of the single-cell optical scatter properties whose multivariate distribution is measured in a typical CBC (Fig. 1). These measurements generally reflect single-cell size, cytoplasmic granularity, nuclear morphology, and other similar characteristics. For instance, the Abbott Cell-DYN Sapphire estimates cell size by measuring the axial light loss (ALL) for each cell, cytoplasmic and nuclear complexity by measuring intermediate angle scatter intensity (IAS), and nuclear lobularity by measuring polarized side scatter intensity (PSS) (10). These single-cell properties provide a great deal of information, however nonspecific, about pathophysiologic states, because they tend to change in particular ways as cells mature, become activated, proliferate, and senesce (4, 11).

## Significance

Medical doctors use blood counts to help diagnose and monitor almost all diseases. In the process of counting blood cells, most hematology analyzers actually measure features of thousands of individual blood cells, but this single-cell information is rarely utilized, and only the derived total cell counts are used to guide clinical care. This single-cell information helps further characterize each patient's inflammatory or immunologic state. Using a mathematical model, we show how this routinely available single-cell information can help distinguish healthy and sick patients in general and those with acute coronary syndrome in particular. More broadly, our study shows how mathematical modeling of existing routine clinical laboratory data can help realize the vision of precision medicine today.

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

Conflict of interest statement: A.C. and J.M.H. are listed as inventors on a patent application filed by their institution.

This article is a PNAS Direct Submission.



[1]To whom correspondence should be addressed. Email: john_higgins@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1709228114/-/DCSupplemental.

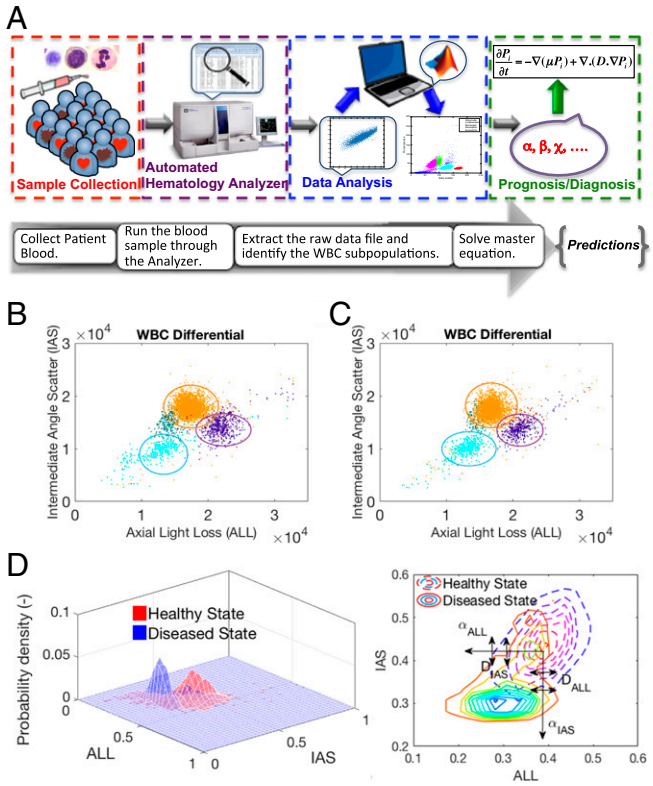


**Fig. 1.** Overview of the modeling approach. *A* shows a schematic of the analytic workflow, including blood sample collection, single-cell measurements on an automated hematology analyzer, data analysis and mathematical modeling, and patient classification. The IAS and ALL scatterplots for a typical patient's CBC are shown in a healthy state (*B*) and at the time of an elevated troponin level (*C*). Probability distributions for the lymphocyte population in healthy and acutely ill states are shown as surface and contour plots (*D*). The contour plots show how the model captures the trajectory of the distribution in terms of drift and diffusion as the blood cell populations are perturbed by or respond to pathologic conditions.

The size, nuclear morphology, and cytoplasmic complexity of individual cells can be utilized as a crude indicator of cell age and activation state, and the distribution of these characteristics can be correlated with many disease processes. There are many basic unanswered questions regarding the life cycle and response of WBCs in vivo in general and during heterogeneous pathologic conditions. We avoided making strong assumptions about the rates of production, proliferation, or turnover by not including explicit source terms (e.g., birth and death) in the model. To minimize assumptions and the number of model parameters, we lump these different functional processes into drift and diffusion parameters for each cell population, with the drift quantifying the average combined effects of proliferation, maturation, activation, and recycling on each individual cell's measurements and the diffusion quantifying the variation from one cell to the next and in the same cell over time.

We model the evolution over time of the probability density ($P$) of these attributes, decomposed into drift ($\mu$) and diffusion ($D$) processes for each cell population. The Fokker–Planck equation is a partial differential equation that can be used to describe this sort of temporal evolution of a probability density function (12):

$$\frac{\partial P_i}{\partial t} = -\nabla(P_i\mu) + \nabla(D\nabla P_i). \qquad [1]$$

The above equation tracks the evolution of the cellular characteristics as they deviate from or fluctuate around a healthy state (13). In Eq. **1**, $P_i$ is the probability density for the single-cell measurements of the $i$th WBC subpopulation ($i \in$ neutrophils, lymphocytes, monocytes); $\mu$ represents the drift function or velocity field, and $D$ is the diffusion coefficient. Fig. 1 shows a schematic of the modeling approach. These drift and diffusion parameters are estimated for each patient CBC and provide a personalized assessment of pathophysiologic state characterized by 13 parameters: ALL and IAS drift and diffusion parameters for neutrophils, lymphocytes, and monocytes and a PSS drift parameter for neutrophils. The model is solved using an initial condition provided by a normal CBC from the patient at an earlier, healthy time point and a final condition reflecting the integrated effects of healthy fluctuations and perturbations caused by any disease.

Lymphocyte population dynamics are modeled in terms of ALL and IAS. We describe the time-dependent evolution of the probability distribution of lymphocyte ALL and IAS ($P_{LYM}$) with a drift–diffusion or Fokker–Planck equation:

$$\frac{\partial P_{LYM}}{\partial t} = -\frac{\partial}{\partial x_{ALL}}\left(\mu_{ALL,L}P_{LYM}\right) - \frac{\partial}{\partial x_{IAS}}\left(\mu_{IAS,L}P_{LYM}\right) + D_{ALL,L}\frac{\partial^2 P_{LYM}}{\partial x_{ALL}^2} + D_{IAS,L}\frac{\partial^2 P_{LYM}}{\partial x_{IAS}^2}. \qquad [2]$$

The drift terms ($\partial/\partial x_{ALL}$ and $\partial/\partial x_{IAS}$) capture the changes in the mean with respect to a morphological attribute, while the diffusion terms ($\partial^2/\partial {x_{ALL}}^2$ and $\partial^2/\partial {x_{IAS}}^2$) track the changes in the variance of the distribution. A change in the central tendency or mean is generally hypothesized to be caused by a change in the distribution of ages or activation states in the circulating population. In Eq. **2**, $P_{LYM}$ is the 2D probability density of the lymphocyte population in the ALL–IAS plane, and $D_{ALL,L}$ and $D_{IAS,L}$ are the invariant diffusive coefficients with respect to the ALL and IAS dimensions.

The growth rate of individual lymphocytes as a function of lymphocyte age or size is unclear and difficult to measure. Previous investigators have found evidence that the growth rate of lymphoblasts is proportional to volume up to a point and then declines roughly linearly with increasing size (14). We represent ALL dynamics in a qualitatively consistent way by modeling the growth rate as a linearly increasing function of ALL up to the median of the final ALL distribution, after which it decreases. The growth rate at zero volume is zero and goes up to a maximum absolute value of $\alpha_{ALL}$. The drift, $\mu_{ALL}$, as a function of normalized ALL (Fig. S2) is

$$\mu_{ALL} = -\alpha_{ALL}\begin{cases} \dfrac{x_{ALL}}{\text{Median}_{P_{\text{end},ALL}}}, & \text{when } x_{ALL} < \text{Median}_{P_{\text{end},ALL}} \\ \dfrac{\text{Median}_{P_{\text{end},ALL}}}{x_{ALL}}, & \text{when } x_{ALL} \geq \text{Median}_{P_{\text{end},ALL}} \end{cases}. \qquad [3]$$

$\text{Median}_{P\text{end},IAS}$ and $\text{Median}_{P\text{end},ALL}$ represent the median IAS and ALL of the final probability density, respectively (target distribution used for the fitting). $\alpha_{ALL}$ and $\alpha_{IAS}$ are the fitted patient-specific drift parameters. We hypothesize that the cytoplasmic complexity and granularity of cells generically reflect cellular responses to activation or deactivation signals. Cellular signal transduction responses are often nonlinear, and we, therefore, model the typical response rate as initially very slow in terms of IAS and then increasing quickly after a signal before stabilizing at a poised level. Michaelis–Menten kinetics provide a standard mathematical description of this sort of sigmoidal

response, as shown in Eq. **4** and Fig. S2. The drift, $\mu_{IAS}$, as a function of normalized IAS is

$$\mu_{IAS} = -\alpha_{IAS}\begin{cases} 1 - 2\dfrac{\dfrac{x_{IAS}}{Median_{P_{end,IAS}}}}{1+\dfrac{x_{IAS}}{Median_{P_{end,IAS}}}}, & \text{when } x_{IAS} < Median_{P_{end,IAS}} \\ 2\dfrac{\dfrac{x_{IAS}}{Median_{P_{end,IAS}}}}{1+\dfrac{x_{IAS}}{Median_{P_{end,IAS}}}}, & \text{when } x_{IAS} \geq Median_{P_{end,IAS}} \end{cases} \quad [4]$$

The population dynamics of monocytes are modeled similarly:

$$\frac{\partial P_{MON}}{\partial t} = -\frac{\partial}{\partial x_{ALL}}\left(\mu_{ALL,M}P_{MON}\right) - \frac{\partial}{\partial x_{IAS}}\left(\mu_{IAS,M}P_{MON}\right) + D_{ALL,M}\frac{\partial^2 P_{MON}}{\partial x_{ALL}^2} + D_{IAS,M}\frac{\partial^2 P_{MON}}{\partial x_{IAS}^2}. \quad [5]$$

$P_{MON}$ indicates the 2D probability distributions of ALL and IAS for the monocyte subpopulation. The drift (15, 16) and diffusion terms are similar to those in the lymphocyte model (Eqs. **3** and **4**).

Neutrophil nuclear morphology changes during maturation, beginning with a simple mononuclear shape as a myeloblast and acquiring indentations as a "band" form, and eventually, multiple lobes as a fully mature polymorphonuclear leukocyte (17). Previous investigators have shown that neutrophil nuclear morphology is altered in response to disease; in particular, the fraction of neutrophils with band vs. segmented nuclear morphology has been shown to increase in states of inflammation, such as neonatal sepsis (18). The lobularity of the nucleus is reflected in the PSS measurement, and we, therefore, include this measurement only in our model of neutrophil population dynamics; we simplify this PSS model to assume no diffusion. Based on this interpretation of PSS, neutrophils with lower PSS are more likely to be immature. A simple way to model the drift along this dimension is to pick a threshold and track the fraction of neutrophils falling below the threshold. We calculated the 2.5th percentiles of several hundred normal CBCs and took the upper quartile of this distribution of quantiles as a candidate threshold to distinguish lower nuclear lobularity (as seen in band neutrophils) from higher lobularity (segmented neutrophils). The size of the fraction of neutrophils with PSS that is below this threshold ($K_{PSS,2.5}$) is hypothesized to reflect the degree of neutrophil immaturity, and a simple characterization of drift is the change in the size of this fraction over time:

$$f_{PSS,2.5}(t) = \int_0^{K_{PSS,2.5}} P_{PSS}(x,t)dx. \quad [6]$$

Modeling the drift of this fraction will capture the change in immaturity. We also want to account for the background level of immaturity on top of which these changes occur. For instance, an increase of 1% in the immature fraction from 1 to 2% may represent background fluctuations, while the same 1% fractional increase from 10 to 11% signifies that an already significantly perturbed state is even further perturbed. We nondimensionalize the drift by scaling by the fraction at the second time point. We found no persuasive biological rationale for adding the complexity of a diffusive term along the PSS dimension and therefore do not include it in our model. Because we are modeling a probability distribution across two grid points, the net drift from the first grid point must then equal the net drift into the second, yielding the following drift term for PSS:

$$\mu_{PSS} = -\begin{cases} \dfrac{\alpha_{PSS}x_{PSS}}{f_{PSS,\text{final}}}, & \text{when } x_{PSS} < K_{PSS,2.5} \\ -\dfrac{\alpha_{PSS}x_{PSS}}{f_{PSS,\text{final}}}, & \text{when } x_{PSS} \geq K_{PSS,2.5} \end{cases}. \quad [7]$$

Thus, our full model of neutrophil population dynamics is the following:

$$\frac{\partial P_{NEU}}{\partial t} = -\frac{\partial}{\partial x_{ALL}}\left(\mu_{ALL,N}P_{NEU}\right) - \frac{\partial}{\partial x_{IAS}}\left(\mu_{IAS,N}P_{NEU}\right) - \frac{\partial}{\partial x_{PSS}}\left(\mu_{PSS,N}P_{NEU}\right) + D_{ALL,N}\frac{\partial^2 P_{NEU}}{\partial x_{ALL}^2} + D_{IAS,N}\frac{\partial^2 P_{NEU}}{\partial x_{IAS}^2}. \quad [8]$$

For a given set of parameters, Eqs. **2**, **5** and **8** were solved numerically for variables normalized by the raw optical intensities: $[ALL, IAS] = [0,1] \times [0,1]$. Initial conditions $[P_{IC(t\,=\,0,\,X)} = P_1]$ were defined by the normalized measurements taken at a healthy time point for that patient, final conditions ($P_{FC} = P_2$) were defined by normalized measurements at the time of patient presentation, and boundary conditions were defined by zero probability outside the range of measured intensity values $[P_{(t,\,X\,\leq\,0)} = 0, P_{(t,\,X\,\geq\,1)} = 0]$. Here, $X$ represents the optical measurements ($X \in ALL, IAS, PSS$). *SI Results* has more details.

## Results

We first compare WBC population dynamics for healthy individuals with those for patients currently being treated for a significant acute illness that warranted measurement of cTn whether or not the test result was positive (abnormal). cTn is measured for patients undergoing emergent treatment for significant nonspecific chest discomfort (19) and often as well for severe and acute symptoms attributed to respiratory, gastrointestinal, musculoskeletal, and other systems. We chose cTn to focus on patients suffering from rapidly progressing disease processes as opposed to those with chronic diseases, which have effects on WBC dynamics that may have a longer timescale complicated by compensatory adaptation. We expected that the dynamics for the healthy individuals would typically correspond to fluctuations around a healthy steady state, while the dynamics for the acutely ill patients would reflect a transition to a new pathologic quasisteady state, integrating the effects of a rapidly acting disease process. To compare the magnitude of the baseline healthy fluctuations with the acute disease changes, we scale the drift and diffusion parameters by an arbitrary time constant. To focus on the effects of WBC dynamics instead of WBC count, we restricted our analysis to CBCs with normal WBC counts. Each of the 13 parameters obtained for patients belonging to the two cohorts, healthy ("control") and acutely ill ("study"), were compared using the Kruskal–Wallis nonparametric test. Fig. 2 reveals the statistical significance of the difference between each individual model parameter for the two groups. The patients in the two groups were thus indistinguishable based on the WBC count, but Fig. 2 shows that 7 of 13 model parameters differ between the two groups with statistical significance.

The differences in median parameters for the two groups show that acute pathologic processes significantly perturb WBC population dynamics and that these perturbations are detectable, even with this coarse model of WBC population dynamics, and

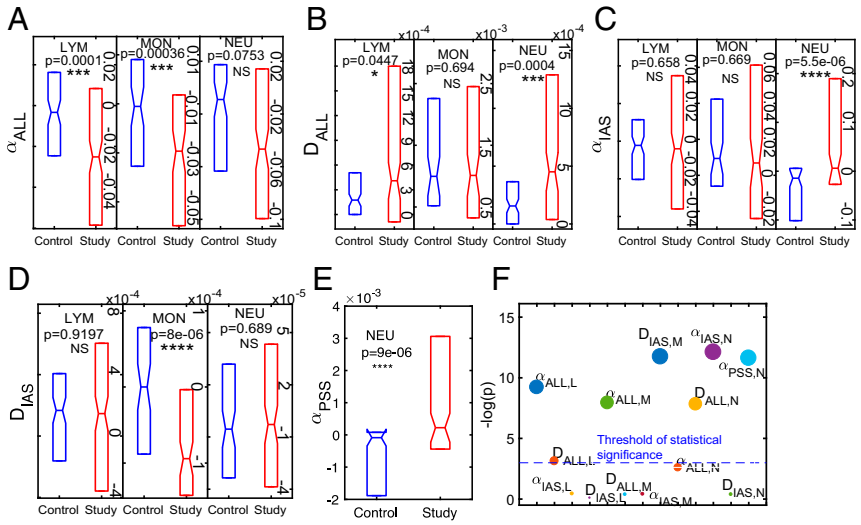


**Fig. 2.** Healthy and acutely ill patients with normal WBC counts can be distinguished on the basis of WBC population dynamics. *A–E* compare individual model parameters for healthy individuals (control) and patients who are acutely ill (study). All WBC counts for all subjects were normal, and the control and study groups were, therefore, indistinguishable based on only WBC count. Each of the 13 model parameters is compared separately between the two groups with a Kruskal–Wallis test. *F* displays log-transformed *P* values [−log(p)] for each model parameter (subscript L is for lymphocytes, M is for monocytes, and N is for neutrophils). Points above the dashed line [−log(0.05)] are significant, and 7 of 13 model parameters are different with statistical significance. Each cohort for each parameter and lineage contains results for all CBCs with good parameter fits ($n >$ 70 in all cases). *Methods* and *SI Results* have more details. LYM, lymphocytes; MON, monocytes; NEU, neutrophils; NS, not significant. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

are also sufficiently independent of changes in absolute WBC counts. Parameters characterizing the ALL dynamics (or size) are significantly different for all three WBC subtypes, with an increasing distribution width for neutrophils in acute illness and a decreasing median for both lymphocytes and monocytes. This pattern would be consistent with a neutrophil-predominant response to acute processes leading to heterogeneity in the neutrophil size distribution compared with a relative reduction in the size distributions of both lymphocytes and monocytes. There is also an increase in the granularity of the neutrophil populations for the acutely ill cohort. Motivated by cellular mechanisms, our models include nonuniform drift fields. For healthy individuals, we expect the dynamics to fluctuate around an average state with no net bias, consistent with the small median drift parameter values for the healthy population as shown in Fig. 2.

We next focus on WBC population dynamics in ACS. In some cases of ACS, the myocardial ischemia progresses to the point of myocardial cell death, necrosis, and tissue infarction (heart attack). We hypothesize that the earliest stages of myocardial ischemia trigger an inflammatory response altering WBC population dynamics and that these altered dynamics are detectable well before the ischemia has progressed to the point of significant myocardial cell death. We, therefore, expect that WBC dynamics will be altered before an elevated level of cTn has leaked into the bloodstream (20, 21), and it might be possible to detect these altered dynamics with our model. ACS patients often have initially normal (low) cTn levels when they first present to the emergency room with symptoms and must wait hours before cTn becomes elevated, delaying the timing of necessary treatment (22, 23). We modeled WBC population dynamics for patients who presented with a normal cTn and compared parameters for those whose cTn remained normal (low) with those whose cTn became elevated (high). Absolute WBC count correlates with risk of MI (24), and to focus on differences in WBC dynamics that are independent of the number of WBCs, we matched patients in each group by absolute WBC count (counts match to within $0.5 \times 10^3$ cells per 1 μL). We compared distributions of each model parameter in the two groups. Fig. 3 shows that 6 of 13 parameters are significantly different between these two groups according to a Kruskal–Wallis test.

The differences in median parameters for the two groups show that changes in WBC population dynamics preceded elevation in cTn for some patients in this cohort and that these changes were detectable even with this coarse model of WBC population dynamics and also even when controlling for absolute WBC count. These groups were also indistinguishable based on the WBC differential (Fig. S1). The parameters distinguishing the two groups with statistical significance using Kruskal–Wallis test are lymphocyte $D_{ALL}$ and $\alpha_{ALL}$, neutrophil $\alpha_{ALL}$ and $\alpha_{PSS}$, and monocyte $\alpha_{IAS}$ and $D_{IAS}$. The median $D_{ALL}$ is elevated when cTn is elevated, as was seen for the acutely ill group in Fig. 2, but the difference in neutrophil $D_{ALL}$ in this comparison is not statistically significant, perhaps because the group with normal cTn is also likely suffering from some sort of acute illness warranting the cTn testing. These other presumably non-ACS conditions may have nonspecifically altered neutrophil dynamics. Earlier identification of patients whose cTn is likely to become elevated in the near term would enable earlier intervention in ACS and possibly improve outcomes for at least some of these patients suffering from one of the most common causes of sudden death in the world. To begin to assess the diagnostic utility of this approach for assisting in the risk stratification of patients being evaluated for ACS, we developed a cross-validated ensemble boosted tree model (AdaBoost) to classify patients at the time of a normal cTn using a subset of the model parameters chosen by feature selection ($D_{ALL,L}$, $\alpha_{ALL,L}$, $\alpha_{PSS,N}$, $\alpha_{IAS,M}$) and the data from the patients shown in Fig. 3 (25). We then identified an independent set of patients with initially normal cTn levels and assessed the accuracy of the same classifier when predicting which of those patients would have an elevated cTn level in the next 48 h. Fig. 4 shows the classifier performance for both the training and validation sets using a receiver operating characteristic (ROC) curve (26) and a confusion matrix. ROC curves are commonly used in medical diagnosis to assess classifier accuracy. A classifier based on WBC population dynamics can thus identify more than 70% of patients who will have an elevated cTn in the next 48 h. Sensitivity is more important than specificity in a screening test, and as shown in Fig. 4, it is possible to choose a decision threshold for this classifier such that, for every four actual MI patients identified who can receive more rapid and appropriate intervention, only about one patient not having an actual MI would begin unnecessary early treatment, corresponding to a sensitivity of 40% and a specificity of 90% when negative and positive cases have equal prevalence.

## Discussion

We have developed a mathematical model of the population dynamics of neutrophils, lymphocytes, and monocytes that coarsely quantifies the lumped effects of cellular production,

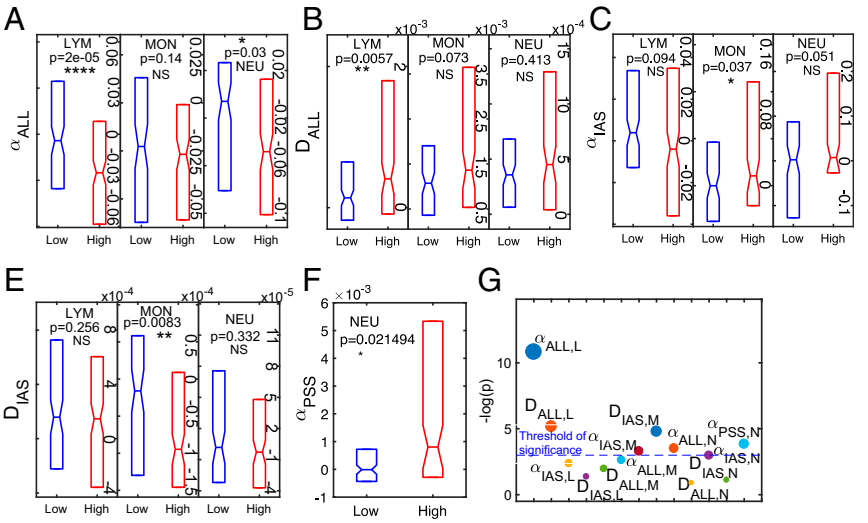


**Fig. 3.** WBC population dynamics can distinguish patients with stably low troponin from those whose low troponin is likely to become elevated. *A–E* compare individual model parameters for patients with stably low cTn (low) and patients whose initially normal cTn becomes elevated (high). Low and high subjects were matched to have WBC counts within $0.5 \times 10^3$ cells per 1 μL. Low and high groups were, therefore, indistinguishable based on only WBC count. Each of the 13 model parameters is compared separately between the two groups with a Kruskal–Wallis test. *F* displays log-transformed *P* values [–log(p)] for each model parameter (subscript L is for lymphocytes, M is for monocytes, and N is for neutrophils). Points above the dashed line [–log(0.05)] are significant, and 6 of 13 model parameters are different with statistical significance. Each cohort for each parameter and lineage contains results for all CBCs with good parameter fits ($n > 110$ in all cases). *Methods* and *SI Results* have more details. LYM, lymphocytes; MON, monocytes; NEU, neutrophils; NS, not significant. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

maturation, activation, and senescence. We found that patients with acute disease processes have significantly different WBC dynamics from those who are healthy, even after controlling for WBC count. We further showed that the model may be helpful in identifying patients being evaluated for MI who are most likely to be confirmed as having an MI in the near future and thus would likely benefit from immediate aggressive intervention.

This model generates hypotheses that require follow-up investigation about the underlying pathophysiological perturbations to WBC population dynamics caused by acute disease in general and MI in particular. For instance, the median and variance of the lymphocyte $D_{ALL}$ and $\alpha_{ALL}$ parameters in Fig. 3 are higher for patients who progress to MI, suggesting that the median and variance of the distribution of the volumes of circulating lymphocytes both increase for patients developing MI. Lymphocytes are hyperproliferative cells that continue to proliferate on activation, and this widening of the size distribution could be attributed to the presence of greater numbers of naïve cells, smaller activated cells entering the circulation, or larger activated cells undergoing additional proliferation (27–29). The ALL measurement generally reflects cell size, and the significant difference in $\alpha_{ALL}$ shown in Fig. 3*A* suggests enrichment of smaller neutrophils in advance of a clinical diagnosis of MI compared with patients who remain stable. Smaller neutrophils have been shown to be older (30), and this enrichment may reflect a consumption of younger neutrophils or a state of activation, which is associated with a reduction in cell size (31, 32). The significance of $\alpha_{PSS}$ for neutrophils is consistent with an increased number of immature or band neutrophils in the circulation, as has been shown before (20, 33). The increased monocyte $\alpha_{IAS}$ for patients progressing to MI (and the statistically significant $\alpha_{ALL}$ for study vs. control groups) suggests an increase in the heterogeneity of the population, which could be attributed to accelerated proliferation or maturation and increased production of new cells (34). These interesting mechanistic hypotheses require dedicated follow-up study, and our model enables the identification and analysis of informative sets of patients.

Patients in our study were evaluated using traditional cTn measurements, and it is possible that some patients with normal cTn who are accurately identified as high risk for cTn elevation would have tested positive using a high-sensitivity cTn assay. Follow-up analysis using one of these assays is warranted. Because we focus on patients who are either stable or being treated for a significant acute illness, we do not explore the timescale for WBC population dynamics. Future study of the timescale for changes in WBC dynamics, particularly in chronic illness, may also be informative and useful. We currently use a normal CBC from each patient as the initial condition, and future work will explore the possibility of using a normal template or ensemble of normal starting points in cases where a normal CBC for the patient is not available.

We have focused on ACS, but we expect that many other diseases are associated with altered WBC population dynamics and that study with this sort of dynamic model will generate insight into human physiology overall, providing potentially useful biomarkers. More generally, our study shows how semi-mechanistic modeling of the high-throughput datasets already routinely collected in hospital clinical laboratories can help realize the vision of personalized medicine.

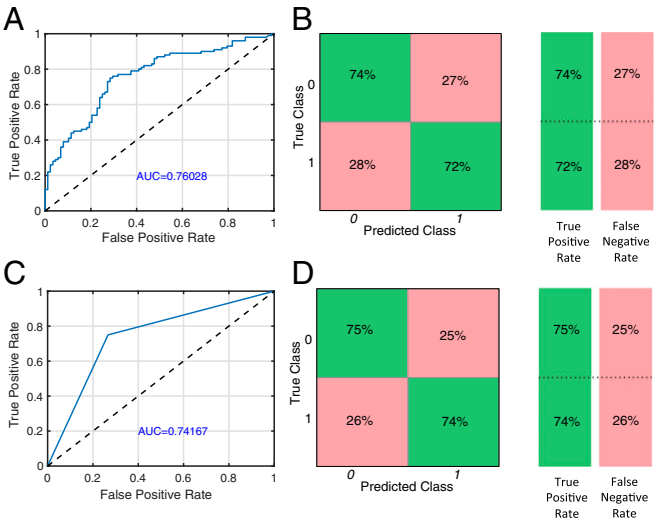


**Fig. 4.** Diagnostic accuracy of model parameters for predicting which patients with normal cTn will have an elevated cTn in the subsequent 48 h. ROC curves were generated for a decision tree classifier using parameters from the model of WBC population dynamics. *A* shows the ROC curve for the training dataset used to develop the classification tree with fivefold cross-validation, and a confusion matrix (*B*) shows the performance of one decision threshold on the training set. This same classifier was applied to an independent set of patients ($n_{low} = 60$, $n_{high} = 60$), with ROC curve (*C*) and confusion matrix (*D*) showing similar diagnostic accuracy and area under the curve (AUC) as expected for a cross-validated classifier.

## Methods

**Human Materials.** All patient data were accessed under a research protocol with waiver of consent approved by the Partners Healthcare Institutional Review Board.

**Patient Blood Sample Collection and Measurement.** All CBC measurements were made on an Abbott Cell-DYN Sapphire automated hematology analyzer (Abbott Diagnostics). The analyzer makes several optical measurements on each cell, including ALL representing cell size, IAS representing cellular complexity, PSS representing nuclear lobularity, depolarized side scatter (DSS) representing cytoplasmic granularity, and cellular DNA fluorescence primarily used to detect nucleated RBCs (10). We expected that cellular size, cytoplasmic complexity, and nuclear lobularity would provide the strongest and most useful correlates of cellular maturation and activation and therefore focused our analysis on the ALL, IAS, and PSS measurements. Since lymphocytes and monocytes are mononuclear, the relevance of nuclear lobularity is limited for these cells, and hence, the PSS measurement is studied only for neutrophils.

**Control and Study Cohorts.** The healthy cohort (control) is obtained from serial measurements of patients who had not had an abnormal CBC index in the prior 2 y and who had not visited the hospital within the prior 250 d to select for patients whose only medical visits were likely for annual physical examination. We excluded parameters on a per WBC-type basis when minimum goodness-of-fit criteria were not satisfied (*SI Results*). Final analysis for the control cohort included 102 CBC pairs for the neutrophil model, 113 for the lymphocyte model, and 74 for the monocyte model. The study group included patients who had a cardiac troponin T (cTnT) test (Roche Diagnostics) ordered at Massachusetts General Hospital between June of 2012 and July of 2015, a CBC at the same time as the cTnT, and also, a normal CBC measurement at least 60 d before the cTnT measurement. CBC pairs were included for modeling of each WBC subtype as above when goodness-of-fit criteria were achieved, yielding 139 CBC pairs for the neutrophil model, 173 for the lymphocyte model, and 97 for the monocyte model.

**"Low" and "High" Cohorts.** Within the study group described above, we identified a low group containing 153 patients who had a normal cTnT measurement ($\leq$0.03 ng/mL) followed by another normal cTnT. These patients were treated at Massachusetts General Hospital and had cTnT measurements between June of 2012, and July of 2015. We also identified a high group containing 201 patients with an initially normal cTnT that became elevated within the subsequent 48 h. We treated an elevated cTnT (>0.09 ng/mL) as diagnostic for an MI (8). Patients with MI often have elevations in their WBC count (24, 35), and we therefore controlled for WBC count, ensuring that each low–high pair differed by less than $0.5 \times 10^3/\mu L$ and that each WBC count was $<17 \times 10^3/\mu L$. These two groups were thus indistinguishable based on WBC count, and Fig. S1 shows that they are also indistinguishable based on WBC differential counts.

**"Validation" Cohort.** A validation set of 120 patients was considered to test the accuracy of a classifier built from the low and high model parameters. Patients who had cTnT measurements at Massachusetts General Hospital after August of 2015 and normal CBCs at least 60 d prior were considered in this cohort.

**ACKNOWLEDGMENTS.** We thank Kent Lewandrowski, Aliyah Sohani, and Ricardo Paxson for helpful discussions. We also thank Shawn Murphy, Henry Chueh, and the Partners Health Care Research Patient Data Registry group for facilitating use of their database. We appreciate financial support from the NIH, the Life Sciences Research Foundation (LSRF), and Abbott Diagnostics. A.C. is a Good Ventures Fellow of the LSRF. J.M.H. is supported by NIH Director's New Innovator Award DP2DK098087. Portions of this research were conducted on the Orchestra High Performance Compute Cluster at Harvard Medical School, an NIH-supported shared facility that consists of thousands of processing cores and terabytes of associated storage partially provided through a National Center for Research Resources Grant 1S10RR028832-01. None of the funders played any role in the decision to submit for publication.

1. Statland BE, Winkel P, Harris SC, Burdsall MJ, Saunders AM (1978) Evaluation of biologic sources of variation of leukocyte counts and other hematologic quantities using very precise automated analyzers. *Am J Clin Pathol* 69:48–54.
2. Horne BD, et al.; Intermountain Heart Collaborative Study Group (2005) Which white blood cell subtypes predict increased cardiovascular risk? *J Am Coll Cardiol* 45: 1638–1643.
3. Tamhane UU, et al. (2008) Association between admission neutrophil to lymphocyte ratio and outcomes in patients with acute coronary syndrome. *Am J Cardiol* 102:653–657.
4. Gijsberts CM, et al. (2015) Hematological parameters improve prediction of mortality and secondary adverse events in coronary angiography patients: A longitudinal cohort study. *Medicine (Baltimore)* 94:e1992.
5. Menezes AA, Vilardi RF, Arkin AP, Cohen MJ (2017) Targeted clinical control of trauma patient coagulation through a thrombin dynamics model. *Sci Transl Med* 9: eaaf5045.
6. Kochanek KD, Murphy SL, Xu JQ, Arias E (2014) Mortality in the United States, 2013 (National Center for Health Statistics, Hyattsville, MD), NCHS Data Brief no. 178.
7. Wilkinson JM, Grand RJA (1978) Comparison of amino acid sequence of troponin I from different striated muscles. *Nature* 271:31–35.
8. Adams JE, 3rd, et al. (1993) Cardiac troponin I. A marker with high specificity for cardiac injury. *Circulation* 88:101–106.
9. Apple FS, Collinson PO; IFCC Task Force on Clinical Applications of Cardiac Biomarkers (2012) Analytical characteristics of high-sensitivity cardiac troponin assays. *Clin Chem* 58:54–61.
10. De Smet D, et al. (2010) Use of the cell-dyn sapphire hematology analyzer for automated counting of blood cells in body fluids. *Am J Clin Pathol* 133:291–299.
11. Lam SW, Leenen LP, van Solinge WW, Hietbrink F, Huisman A (2011) Evaluation of hematological parameters on admission for the prediction of 7-day in-hospital mortality in a large trauma cohort. *Clin Chem Lab Med* 49:493–499.
12. Van Kampen NG (2007) *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam), 3rd Ed.
13. Higgins JM, Mahadevan L (2010) Physiological and pathological population dynamics of circulating human red blood cells. *Proc Natl Acad Sci USA* 107:20587–20592.
14. Tzur A, Kafri R, LeBleu VS, Lahav G, Kirschner MW (2009) Cell growth and size homeostasis in proliferating animal cells. *Science* 325:167–171.
15. Neumann FR, Nurse P (2007) Nuclear size control in fission yeast. *J Cell Biol* 179: 593–600.
16. Webster M, Witkin KL, Cohen-Fix O (2009) Sizing up the nucleus: Nuclear shape, size and nuclear-envelope assembly. *J Cell Sci* 122:1477–1486.
17. Beck WS (1998) *Hematology* (MIT Press, Cambridge, MA), 5th Ed.
18. Cornbleet PJ (2002) Clinical utility of the band count. *Clin Lab Med* 22:101–136.
19. Daubert MA, Jeremias A (2010) The utility of troponin measurement to detect myocardial infarction: Review of the current findings. *Vasc Health Risk Manag* 6:691–699.
20. Madjid M, Awan I, Willerson JT, Casscells SW (2004) Leukocyte count and coronary heart disease: Implications for risk assessment. *J Am Coll Cardiol* 44:1945–1956.
21. Marjot J, et al. (2017) Quantifying the release of biomarkers of myocardial necrosis from cardiac myocytes and intact myocardium. *Clin Chem* 63:990–996.
22. Mair J, et al. (1991) Cardiac troponin T in diagnosis of acute myocardial infarction. *Clin Chem* 37:845–852.
23. Hamm CW, et al. (1997) Emergency room triage of patients with acute chest pain by means of rapid testing for cardiac troponin T or troponin I. *N Engl J Med* 337: 1648–1653.
24. Menon V, et al. (2003) Leukocytosis and adverse hospital outcomes after acute myocardial infarction. *Am J Cardiol* 92:368–372.
25. Freund Y, Schapire RE (1995) A desicion-theoretic generalization of on-line learning and an application to boosting. *Proceedings of the Second European Conference on Computational Learning Theory, EuroCOLT '95*, ed Vitányi P (Springer, Berlin), pp 23–37.
26. Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin Chem* 39:561–577.
27. Thompson CB, et al. (1984) Size-dependent B lymphocyte subpopulations: Relationship of cell volume to surface phenotype, cell cycle, proliferative response, and requirements for antibody production to TNP-Ficoll and TNP-BA. *J Immunol* 133: 2333–2342.
28. Carstairs K (1962) The human small lymphocyte: Its possible pluripotential quality. *Lancet* 1:829–832.
29. Mackay CR (1993) Homing of naive, memory and effector lymphocytes. *Curr Opin Immunol* 5:423–427.
30. Casanova-Acebes M, et al. (2013) Rhythmic modulation of the hematopoietic niche through neutrophil clearance. *Cell* 153:1025–1035.
31. Hoffstein ST, Friedman RS, Weissmann G (1982) Degranulation, membrane addition, and shape change during chemotactic factor-induced aggregation of human neutrophils. *J Cell Biol* 95:234–241.
32. Athens JW, et al. (1961) Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *J Clin Invest* 40:989–995.
33. Kawaguchi H, et al. (1996) Band neutrophil count and the presence and severity of coronary atherosclerosis. *Am Heart J* 132:9–12.
34. Wang SY, Mak KL, Chen LY, Chou MP, Ho CK (1992) Heterogeneity of human blood monocyte: Two subpopulations with different sizes, phenotypes and functions. *Immunology* 77:298–303.
35. Koren-Morag N, Tanne D, Goldbourt U (2005) White blood cell count and the incidence of ischemic stroke in coronary heart disease patients. *Am J Med* 118: 1004–1009.