Mathematical Models to Enhance the Value of Information from Current Laboratory Platforms

Molly Webster[1] and Vikram Sheel Kumar[2*]

Buckminster Fuller once said, "It is one of our most exciting discoveries that local discovery leads to a complex of further discoveries." Every minute, a clinical laboratory generates local discoveries. Examples are a creatinine value of 3 mg/dL (265 μmol/L), a potassium concentration of 6 mEq/L, or a hematocrit of 24%. Each observation provides the clinician with the potential to discover the cause of an urgent problem. Some researchers believe we can discover disease earlier by using existing laboratory platforms. For example, while performing a complete blood count (CBC),[3] we can collect up to 80 000 measurements at a time. "Clinically, the only thing typically used is the average red blood cell volume and maybe the variance," said Dr. John Higgins, assistant pathologist at Massachusetts General Hospital in Boston. "It was shocking to me." Higgins wanted to know what could be done with some of what he considers as high-throughput data. We spoke with Higgins about his pioneering work on a mathematical model that he hopes will make use of the wealth of these data to identify various forms of disease and increase opportunities for the early detection of illness.

Why Is This Invention Important?

The body is known to create 200 × 10^9 red blood cells (RBCs) per day. The body also loses the same number of cells through senescence, thereby maintaining an overall stability of the RBC population; however, little is known about the aspect of the 120-day life cycle of RBCs that occurs after they leave the bone marrow. To try to elucidate more about the RBC aging process, researchers began studying the volume and hemoglobin mass of RBCs, only to find that the 2 parameters are linked: New reticulocytes have a high volume and contain a high hemoglobin mass, and both variables decrease as the cell ages (although cell volume decreases more than hemoglobin mass, meaning that the hemoglobin concentration actually increases). "Biology is messy and noisy[,] and any time one sees a level-of-order increase, for example, when correlations between naturally independent quantities increase over time, it is likely that the body has a good reason to want that to happen," said Higgins.

Higgins was one of the first to notice that the strength of this correlation changes over time, and he was a leader in trying to quantify the rate at which it increases over the RBC life span. By using a mathematical model and the excess of high-throughput data gathered during a routine CBC, Higgins is able to study a patient's blood in a different way.

For example, Higgins has reported that the body's production of fewer reticulocytes in patients with iron deficiency anemia (IDA) appears to be counterbalanced by delayed clearance of old cells. This delay allows the body to maintain an appearance of normalcy—a stable total volume of the RBC population—although there is an underlying issue. The traditional clinical anemia tests diagnose the disorder by looking at the mean RBC volume, but the tests can be fooled by the body's trickery: A body could already be fighting IDA despite appearing healthy. "If the body is able to delay the removal of aged red blood cells, it masks the deficiency," said Higgins. "Which is great for your body, but it actually confounds the diagnosis. We can identify a blood test from someone already working hard to compensate decreased production."

In the patient population used in their study, Higgins and Mahadevan were able to identify IDA 75% of the time and at least 1 month before traditional tests (1). Clinicians know that extra time is critical for IDA, because it is often symptomatic of other illness, such as colon cancer in postmenopausal women and in adult men. "In the case of colon cancer, a tumor can lead to low-level bleeding in the intestinal tract," said Higgins. "As the tumor grows, bleeding slowly leads to iron deficiency—and by the time you realize it is anemia, in some cases the tumor may have progressed to a
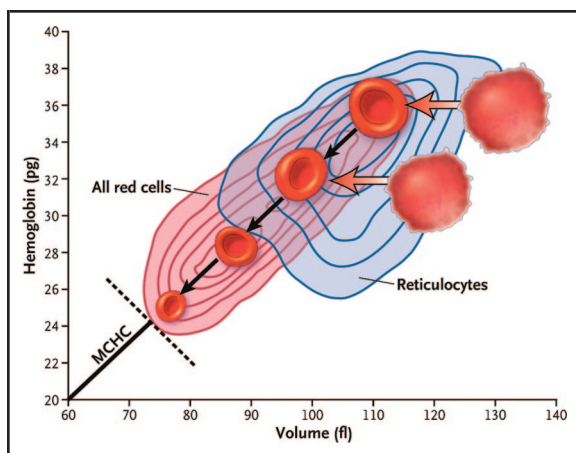
Fig. 1. Model for the typical volume and hemoglobin changes occurring in circulating red blood cells.

Reticulocytes enter the circulation with a range of volumes and hemoglobin masses (blue contours). The volume and hemoglobin mass of each cell decrease over time such that the cell's hemoglobin concentration tends toward the mean concentration—the mean corpuscular hemoglobin concentration (MCHC). Most cells are removed from the circulation and recycled before reaching a clearance threshold, which is represented schematically as a dashed line perpendicular to the MCHC line. Reproduced with permission from Weatherall *(2)*.

point beyond help." With the approach of Higgins, the clinician has a chance for earlier diagnosis. Furthermore, the model can also distinguish between IDA, anemia of chronic disease, and thalassemia trait.

How Does It Work?

Higgins uses steady-state volume and hemoglobin measurements to create a mathematical model describing "how volume and hemoglobin change over time during RBC maturation and clearance."

The mathematical model—developed by Higgins and first described in a 2010 report in the *Proceedings of the National Academy of Sciences*—calculates rates to understand how quickly volume decreases over time and how that rate changes with time. The simplicity of a software approach to better diagnostics comes about by its ability to be integrated with existing laboratory analyzers.

Models of RBC populations demonstrate the maturation of RBCs from reticulocytes (Fig. 1). The statistical distribution of these populations can be seen on a plot of cellular hemoglobin mass against cellular volume. A cell starts as a reticulocyte, which has a large volume and a normal hemoglobin mass, about 20% and 15% larger, respectively, than the typical RBC. The rate of changes during the fast phase is quantified by

the β parameters in the model. β_v quantifies the rate of volume change during the fast phase, and β_h quantifies the rate of hemoglobin change during this phase. The α parameter quantifies the rates of slow change in both variables. During the slow phase, the RBC moves generally along the mean corpuscular hemoglobin concentration line, and the rate of progression is quantified by α. Ultimately, when the RBC becomes too small ("critical volume"), the cell is terminated.

The model can use these parameters (α, β_v, β_h, critical volume) as well as others to create a statistical distribution of RBC volumes and hemoglobin masses. In particular, Higgins and Mahadevan initialize their RBC population simulator with reticulocyte volume, hemoglobin mass distribution, and a randomly chosen set of parameter values. These parameters are then iteratively modified to reach a point at which the virtual CBC measurements performed on the estimated population distribution are similar to actual CBC results. The estimated population distribution and parameter values provide a much deeper perspective about an RBC population than what is immediately apparent from a CBC analysis.

Where Can This Technology Fit in the Clinical Laboratory?

Beyond understanding anemia disorders, Higgins believes that this mathematical diagnostics approach can be applied to understanding various dynamic systems in the body, including white blood cells. "We have a gold mine of information that's telling us a lot more about what's happening in vivo," said Higgins. "We can use modeling as a tool to infer rates and other features."

Currently, both Siemens Healthcare Diagnostics and Abbott Laboratories, whose high-throughput analyzers Higgins used in creating the model, are looking into advancing the scientific research of Higgins and his colleagues. In an e-mail to *Clinical Chemistry*, Don Wright, of Abbott Hematology Scientific Affairs, said that in cell populations associated with disease, it is not outside the realm of possibility for other mathematical models to have predictive effects. "As the dynamics of cellular analysis continues to evolve with new methods and more accurate probes are discovered, the mathematical relationships will play a key role in simplifying the volume of new data and making it useful to clinicians," said Wright.

Some may ask why the mathematical diagnostics approach of Higgins is unique. After all, journals these days do not have a dearth of published algorithms. Wright explained, "Unlike some past algorithms, this approach represents a shift from looking at overall RBC population indices . . . to individual cellular information that identifies unique RBC subpopulations for anemia diagnosis." The most important aspect of this innovation could be the simplicity of enriching for the local discoveries that could be made from standard laboratory analyzers. Could this be the end of average medicine? At the very least, this work emphasizes the additional information potentially available in an automated CBC analysis for enhancing and extending its use beyond iron deficiency.

References

1. 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.
2. Weatherall DJ. Systems biology and red cells. N Engl J Med 2011;364:376–7.