The Weakest Link in Diagnostics May Be the Sample Itself

By Mark D. Lim

Anyone who has had blood drawn at a doctor’s office is familiar with much of the information that results from its analysis, such as glucose, cholesterol, lipids, and cell counts. In the past two decades, two trends have given rise to a revolution in the next generation of diagnostic technologies.

First, molecular biology is increasingly finding molecules in bodily fluids (such as blood and saliva) and tissues that provide information about a patient’s health state to guide the decisions made by a physician (“biomarkers”). Second, engineering advances have enabled these tests to be performed with equipment that is smaller, less expensive, and requires less sample for analysis. These trends are creating market opportunities that, for DNA sequencing technologies alone, are predicted to exceed $20 billion with opportunities beyond niche diagnostic markets.¹ This is encouraging more entrepreneurs to join this rapidly developing market by applying advanced analytics and novel sensors to increase the speed, precision, and accuracy for interrogating human samples, while also reducing per-analysis costs.

The Problem of Variability in Molecular Analysis

A central task for most molecular diagnostics is the measurement of a biomarker or panel of biomarkers; that is, determining which molecules (typically proteins, DNA, RNA, or small molecules) are present and in what quantities, and how variations in the presence and/or quantity of those molecules correlate with to determine if a population is at-risk of developing or harboring disease. Such broad applications of genomic- and proteomic-based diagnostics to public health are as vulnerable as any research to flaws that prevent their reproducibility, as has been reported recently.² While there are many factors that may contribute to irreproducibility in the use of diagnostic technology, there are some sources of data variability and bias that can be mitigated when engineering the workflow from human sample to molecular answers on a population scale.
health or illness. For diagnostic technologies that are being developed to distinguish the well from the sick, track the progress of disease outbreaks, or predict susceptibility to future illness, the characteristics of the analytes detected by a platform must be in context of the entire population intended to be screened. This statement seems obvious, but several efforts to develop diagnostics fail to account for regional normality; that is, samples used to create and benchmark the test did not represent the actual population where it will be used. Epidemiologically-relevant demographics that developers must consider include age, gender, natural history of disease, lifestyle, co-infections, geographies, and environmental exposures. Failure to take these variables into account can invalidate a given diagnostic technology or biomarker for application to a specific population or geography.

The biomarker(s) targeted by a diagnostic also need to be chosen with care — and it’s important to consider their quantitative abundance in any given volume of fluid. For example, the gold standard for determining the concentration of glucose in blood is analysis of glucose itself. However, there is significant market demand in diabetes care for methods that can determine blood glucose without the need for a finger stick or venipuncture, and the R&D landscape is crowded with research aiming to measure glucose in saliva or tears. However, any solution that is acceptable to physicians and regulators (giving good metrics for patient care) and patients (non-invasive and convenient) still requires additional foundational research to benchmark glucose measurements in non-invasive specimens to “gold-standard” blood-based measurements.

**Pre-Analytical Variability Threatens the Reliability of Many New Technologies: Sample Collection**

In addition to the inherent variability among human subject populations, variability in the handling of sample materials themselves can render study results irreproducible. The high analytical sensitivity of next-generation diagnostics potentially amplifies these sources of sample-to-sample variation. When this variability noise is confused with signal, it confounds the very purpose of disease surveillance and screening; what appears to be real person-to-person differences can be the result of mere differences in sample collection and handling. Ignoring variability in the collection and handling of samples can result in the detection of a signal that should have been noise (a false positive result), which consumes resources to verify. Conversely, a signal that was buried in the noise results in the failure to detect a true risk (a false negative result).

Sampled fluids and tissues begin to change as soon as they are removed from the human body, being broken down by both internal factors (e.g., degradative enzymes) and external (e.g., the presence of microorganisms, in samples like saliva and stool). When developing a reliable diagnostic technology, researchers and end users must account for these sources of instability that might result in unacceptably large sample-to-sample variability. Additional variability can also be introduced by inconsistent pre-analytical methods for collecting, transporting, and preparing a sample. These processes are often manual and vary in waiting times (samples left on benchtops before further processing), temperatures, humidity, reagents, and handling conditions.

The impact of sample collection conditions on the reproducibility of diagnostic results is greater in resource-limited environments, particularly for unstable analytes such as RNA or proteins. As mentioned above, samples begin to degrade as soon as they are collected, and refrigeration or freezing are typical means of slowing sample degradation. Preservatives are often added to samples as well, and must be chosen carefully. For example, preservatives added to blood storage tubes (such as anti-coagulants, clotting agents, etc.) have been shown to interfere with later analysis by various platforms.

Consistency of sample preparation is another goal of diagnostic technology developers. All diagnostic analytical platforms require the target molecules to be purified and concentrated from a sample before analysis. This process, called sample preparation, is often as, if not more, complex than the analysis itself. On average two-thirds of the real estate and fluidic components on commercially-available test cartridges are dedicated to sample preparation, typically customized to the analytic platform, biospecimen type, and specific analyte(s).

Developers of new diagnostic technology and its end-to-end workflow therefore must, as an inherent part of their process, evaluate all potential sources of variability for tests that analyze molecules as diverse as nucleic acids (DNA and RNA), proteins, and small molecule metabolites. Unfortunately, there is no “cure-all” method to mitigate all sources of pre-analytical variability for all platforms and analytes. Technology developers must try their best to standardize the means of collecting and transporting samples to the instrument. A consistent, effective, and rigorously followed protocol for sample collection and handling is needed to mitigate variability in the sample, which in turn allows for more accurate diagnosis.
collection and preparation is essential for mitigating post-collection alterations caused by the illustrative factors mentioned above.

To illustrate the process of developing a sample collection technology, we discuss below one example, among the simplest and most useful methods available in resource-constrained parts of the world.

**Case Study: The Promise of Dried Blood Spot Cards**

Dried blood spot cards continue to capture the imagination of multiple communities since they were first demonstrated by Robert Guthrie to simplify newborn screening in the early 1960s; DBS is still central for routine screening of specific metabolic, endocrine, and genetic disorders within the developed world. For this use case, DBS offers the ability to take a small volume of blood (15 to 60 µL) from a newborn without a needle (beyond a lancet for the heel prick) and place it onto an absorbent card. DBS is also useful in the drug development setting, as researchers use them to collect small volumes of blood from small, fidgety animals while maximizing the number of sampling opportunities. DBS cards are also useful in public health and population-wide disease surveillance because they simplify the logistics of transporting a self- or simply-collected sample from a remote location to a central laboratory. DBS cards are inexpensive to manufacture by robust roll-to-roll processes, making them an effective technology for large population screening.

Because of their usefulness, the biopharmaceutical industry has devoted significant resources to evaluate the performance of DBS. Most studies have uncovered multiple ways in which DBS introduce variability into downstream analyses. Commonly cited issues include:

- Patient-to-patient differences in red blood cell concentrations can affect the measured levels of other blood components in an unpredictable manner;
- DBS cards themselves pre-separate blood components in a variable manner before drying, in a process called chromatographic separation; and
- Efficiencies of extracting blood components from cards varies with the target component, card type, solvent, and other experimental conditions.

When collecting samples from newborns, or in drug development, DBS is typically used in controlled settings like a hospital or centralized laboratory. This is not the case, however, when DBS is used as the sample collection medium for diagnosis or disease surveillance in remote regions. In such areas, variability in later analysis of DBS-collected samples may be caused by:

- Contamination by the handler pre- or post-collection
- Contamination from dust, dirt, or insects when dried on an open surface
- Contamination from contact with other non-dried cards
- Variation in drying times due to temperature and humidity
- Degradation of analytes from exposure to sunlight
- Fluctuations in environmental conditions after drying, and during storage and transportation

There have been far fewer studies evaluating DBS for use in remote diagnosis or surveillance. Those studies that exist similarly reveal unpredictable performance, highlighting the need to carefully characterize any potential variability in relation to specific analytes and platform technology. For example, DBS is often recommended when collecting samples from patients in remote areas when assessing HIV viral load to monitor efficacy of a specific therapy, or perform early infant diagnosis. However, the ability of DBS to permit uniform extraction and processing of RNA (the key analyte in this case) has only been reported recently. Some researchers report inconsistent measurements of HIV viral load that appear to be dependent on the chosen analytical platform. Under these circumstances, a quantified measurement that incorrectly represents an individual’s health status can result in an incorrect assessment of the patient’s response to anti-HIV therapy.

No approach has been shown to mitigate all sources of variability when using DBS as the front-end sample, collection, and transport medium. In fact, DBS cards are largely unchanged since initially described by Dr. Guthrie. Non-cellulose DBS cards have been developed for cases where cellulose may affect the extraction of target molecules. Other versions of DBS cards include preservatives that stabilize nucleic acids and proteins. DBS-like formats have also been developed to collect and store derivatives of whole blood that are prepared on- or off-card, such as dried plasma and dried serum cards. There are also DBS accessories that protect cards from environmental conditions or cross-contamination, enhance dehydration through desiccants, and devices that control the volume deposited onto the card.

As is true for any sample collection technology, the value of DBS in any strategy — research, diagnostics, or surveillance — needs to be counterbalanced with the resources required to evaluate and minimize sources of data variability. This can only be done by identifying each step in the integrated collection-to-result workflow and evaluating where variable processes and
conditions impact downstream analytical results. An important set of basic principles was published through a collaboration between multiple pharmaceutical companies and the FDA.\textsuperscript{7, 8} Even though these procedures are focused on a drug development use case, they serve as an important resource to estimate the level of rigor for qualifying DBS processes and technologies for other use-cases.

**Hope Comes From the Intersection of Technology and Rigorous Processes**

To fully exploit the remarkable advances in diagnostic science and engineering that are now emerging, it is essential to account, to the greatest extent possible, for the experimental variability that exists before the actual diagnostic analysis. Deep diligence in all these variables — disease, epidemiology, validation of biomarkers and the sample type from which they are isolated, as well as considerations for sample collection, preservation and transport, and the choice of a compatible analytical platform — is laborious, but essential to maximize the value of data in disease surveillance and public health assessments. This diligence is best achieved by analyzing replicate samples collected and handled under the range of concentrations and conditions that simulate the real-world user setting, with an eye towards assessing impact on analytical linearity, trueness, detection limits, consistency, and precision. When done rigorously, this process leads to an end-to-end, sample-to-answer diagnostic architecture that can provide value across the entire public health spectrum, from health monitoring to the surveillance and eradication of endemic disease, and the quenching of disease outbreaks.

Mark D. Lim is a Program Officer at the Bill and Melinda Gates Foundation’s diagnostics team, focused on investing in the development of diagnostics for neglected tropical diseases and cervical cancer. Prior to joining the Gates Foundation, Lim served as an Associate Director at FasterCures, a Center of the Milken Institute, focused on mechanisms for multi-stakeholder partnerships to biomedical research, in addition to providing counsel to several patient organizations with active funding programs. Previously, Lim served a supporting role as the Chief of Technical Staff to Col. Daniel Wattendorf at DARPA’s Defense Sciences Office, on a program that had efforts in new diagnostic and vaccine capabilities. Lim received his Ph.D. in Chemistry at the University of California, Santa Barbara and completed an NCI-funded postdoctoral fellowship in cancer nanotechnology at the University of California, San Francisco and the University of California, Berkeley.

**REFERENCES**