Review of the Role of Biostatistics in the Life Cycle of Commercial Molecular Diagnostic Assays

By Robert Makowsky, PhD, Global Head of Biostatistics, and Sian Dibben, Almac Diagnostic Services

Introduction

New therapies may be most effective for individuals with specific genetic variants. Today, many early phase trials are enriched such that subjects with the variant are enrolled and those that do not harbor the variant may not be eligible for enrollment. For these trials, companion diagnostic (CDx) assays using molecular biomarkers to measure features of DNA and RNA may be tested in parallel (or after drug release) for the purpose of indicating which populations may most likely respond to disease treatment, or are less likely to suffer side-effects of biomarkers. Well known examples include *BRCA1/2* mutations for sensitivity to PARP inhibitors and *KRAS* mutations for resistance to EGFR inhibitors.

For these assays to be considered reliable diagnostic tools, they must demonstrate analytical and clinical validity during pivotal or follow-on studies. But fit for use assays are often not available and must be developed *de novo* or at least modified from existing assays. We propose that the best approach for a successful CDx application with its associated drug is to integrate CDx planning from the beginning of the drug development cycle and provide objective biostatistical support throughout the CDx cycle, including testing in trials.

In this article, we focus on how and when biostatisticians are involved, what specialized information they provide, and how they can help maximize the chances of success while minimizing risk and cost. Throughout this article, we also highlight the critical role of biostatisticians as integral to the design, development, validation and commercialization of CDx products.

Designing Assays

The aim of a well-designed diagnostic development process is to produce a robustly designed product with an efficient pathway to regulatory approval – again, bearing the end in mind. The start of the cycle begins with the discovery of genes or clinical markers, a vital and in-depth process that serves as the source for CDx design. Once a marker or genetic aberration(s) has been identified, suitable assay candidates capable of reliably identifying the patient sub-population must be reviewed to determine which assay options are sufficiently sensitive, accurate and precise to be fit for purpose.

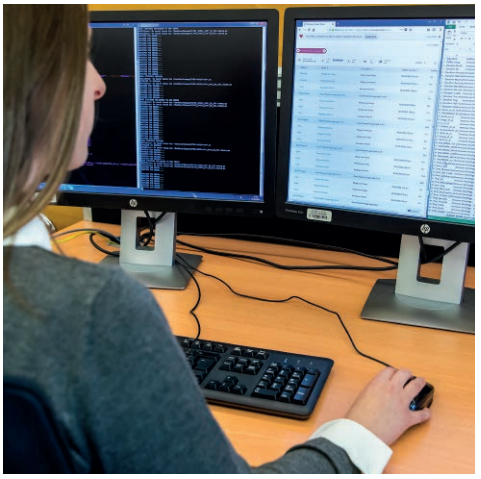
To understand the various factors that affect performance of the assay options and then develop an assay fit for use, biostatisticians need to have an integral role when designing developmental experiments. Such experiments will set the parameters that define platform/process requirements (e.g., number of samples, molecular inputs) or sample properties (e.g., sample composition) that will constitute test usage. They will also determine if performance criteria can be met (e.g., reasonable molecular input amount, sample preparation conditions) with the genetic material that is available. If a statistically robust process is followed in the selection of markers, materials, and assay development experiments, the product stands the best chance of meeting its validation requirements and being clinically relevant.

Example 1 – Designing a single TP53 assay for multiple sample types

In one example, Almac Diagnostic Services developed an NGS assay for the tumor suppressor gene *TP53*, which is known to have homozygous mutations in approximately 50% of cancers. Since some drug researchers prefer to therapeutically stabilize wild-type P53 protein (the product of the *TP53* gene) whereas others prefer to identify P53 mutant cancers for other therapies, our goal was to produce a single assay that could be applied to various cancer types. One important question was how deep the sequencing needed to be for the assay to perform as required. To biostatisticians, more data (that is, more reads) are almost always better, but there are cost and time implications. In this case, the biostatistician worked with the medical and analytical team to identify a sequencing depth that would produce a sufficiently analytically sensitive assay (the ability to detect a true mutation within the limits of assay detection), but also distinguish samples that truly did not have any mutations (analytical specificity).

Example 2 – Selecting reference genes for PCR assays

In another example, we have worked with several client teams to develop qPCR gene expression assays for their commercial needs. One of the most important decisions taken for assay development is the choice of reference genes to complement the marker genes during the qPCR reaction. Since clients typically are well-acquainted with their needs and system under test, our biostatisticians have adopted a customized approach by working with the client to choose the best reference gene(s). Once these reference genes have been chosen, other parameters (e.g., assay input, lower limit of quantitation (LLoQ) determination, and the number of replicate wells) need to be determined to optimize assay reproducibility and limit failure rate. (see **Figure 1** and Example 4 for more detail on reproducibility).



Analytical Validation Processes

Following assay development, test performance metrics mandated for regulatory approval must be established, finalized and validated. This decision process involves a complex interplay of disease considerations, regulatory requirements, the chosen assay, as well as statistical expertise and biological knowledge. As always, finite resources dictate the potential scale of studies, often requiring biostatisticians to weigh in on defining acceptance criteria and study designs (e.g., number of samples needed, including reference and control, reagent shelf-life conditions) that will satisfy medical directors and improve the likelihood of passing agency reviews, while keeping to a budget.

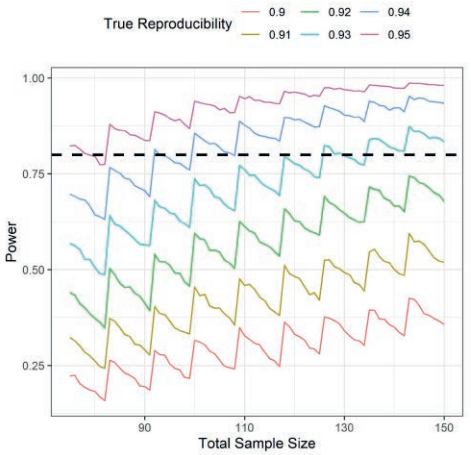


Figure 1: Example line plot that depicts the estimated statistical power while varying the sample size (across the x-axis) and true reproducibility (different colored lines).

Example 3 – Validating a ddPCR assay for circulating gene marker

As one example, a client needed a droplet digital polymerase chain reaction (ddPCR) assay for detection of a single, freely-circulating genetic aberration in blood (specifically, circulating, cell-free DNA, ccfDNA). Liquid biopsy assays have several clinical benefits (e.g., less invasive than solid tumor biopsies, easier serial evaluation over the course of a disease) but pose unique challenges during analytic validation. One of the primary challenges is that the rarity of the target mutation necessitates high DNA inputs to improve the chance of the target being present in the sample, which necessitates a large amount of blood plasma. In the case of a request for a 95% detection capability of an analyte at a level of 1 part in 20,000 (by volume), the biostatistics team examined the developmental data to determine the range of usable droplets per well and combined this with Poisson sampling theory. They determined that four wells per subject would be required, with each well containing 350 ng of cell free DNA input, to have a sufficiently high chance of detecting the analyte at the targeted level. Levels of ccfDNA vary widely based on disease and tumor burden, but this case was particularly challenging to accommodate clinically since the levels of cell-free DNA are typically at concentrations of < 10 ng per mL of blood plasma. To obtain the necessary amount of DNA, approximately 350 mL of blood would be needed. Obviously, this was not practical, so the choice of ddPCR as an assay using ccfDNA was abandoned.

Example 4 – Establishing assay reproducibility

One of the central performance characteristics of any assay is its reproducibility. In reproducibility studies, we typically aim for the entire 95% confidence interval (CI) of the reproducibility estimate to exceed a particular performance goal, which is determined by the biostatistician working with the medical director and the product development team. For example, suppose the performance goal is >= 85% reproducibility. This means that the same samples should give the same detection call at least 85% of the time. For the 95% CI of the reproducibility estimate to be expected to exceed 85% during analytic validation, the assay's true reproducibility must exceed 85%. Suppose for this example that the assay is expected to be ~ 95% reproducible. The biostatistician will determine how many samples are required to have at least an 80% probability of exceeding the performance goal (also referred to as 80% statistical power).

But notice that to estimate the necessary

sample size, an estimate of the reproducibility must already be available, which is challenging. The sample size estimate is only as good as the assumptions, so the biostatistician should appropriately estimate a "worst case" scenario that diminishes the assay performance. **Figure 1** depicts the estimated power for a reproducibility experiment where the performance goal is that the lower 95% CI must exceed 85%, the true reproducibility is between 90% (hypothesized worst case) and 95% (hypothesized reproducibility), and the analysis is based on a binomial proportion. Notice that the increase in power is not continuous (monotone increasing) but more a pattern of slanted steps.

Example 5 – Adaptive designs for CDx and precision medicine applications

Some studies may benefit from an adaptive sample size design – that is, a design for which the sample size is not fixed at the start of the experiment, but instead varies based on the results obtained. Such study designs are intended to minimize both sample size and risk when conditions are appropriate. We have found that when samples are rare, or the cost of running the assay is very high, an adaptive sample size design can be beneficial.

To cite one example, we worked on an assay that allowed for eight samples per run. We used an adaptive sample size approach, where runs of eight samples were performed *one at a time*, and the results were analyzed by the biostatistician to determine 1) if the study could stop before more samples were run and definitively declare the study met or failed the acceptance criterion, or 2) if the conclusion was ambiguous and another run was necessary. Although this approach requires more time overall to complete the study, and possibly more working hours from the laboratory technicians and biostatisticians, the approach may be preferable when assays are expensive or samples are limiting and precious.

Critical Steps For Clinical Validation

While analytical validation is important to move to the clinical stage, clinical validity and clinical utility must be demonstrated independently to receive regulatory approval (e.g., FDA) and qualify for reimbursement for clinical use (CMS). In an ideal scenario, the assay is fully locked after development, put through CDx-level analytic validation, and then used in the drug's registrational trial prospectively as part of the enrollment criteria. The trial assesses the clinical validity of the device by demonstrating its ability to identify the appropriate patient population and ensure clinical utility for the prospective use of the CDx and drug combination.

Unfortunately, in our experience, a fully locked CDx is rarely ready for registrational studies. Analytical validation studies can be lengthy, and pharmaceutical companies typically do not want to delay the drug trial to accommodate the assay. In such a situation, an initial clinical trial assay (CTA) is often used during the trial in place of the final CDx. The analytical validation requirements for a CTA are much less than a CDx and more likely to meet the timelines of the pharmaceutical company as well as reduce the initial investment. This situation means that the clinical trial samples must be later rerun with the CDx for the purpose of clinical validation, which can lead to multiple biases. A biostatistician can help to determine what proportion of samples should be rerun with the CDx, as well as if any stratification of sampling should occur.

In some cases, the CTA can be the same device as the CDx, only differing in the level of supporting validation and regulatory approval. In this case, the clinical bridge would be easier and the assay outcomes would be expected to agree 100%.

Example 6 – Addressing differences between CDx and CTA assays

Sometimes there are fundamental differences between the CDx and CTA, whether in the chemistry, equipment, or analysis pipeline. In such instances, the biostatistician must determine what biases are possible. One example we have seen is where the CDx is more analytically sensitive than the CTA, and the CDx can detect the analyte of interest at a lower amount than the CTA, creating a potential discrepancy between CDx and CTA results.

While all CTA + subjects (those identified by CTA as more likely to respond to the drug) would be expected to be CDx +, some proportion of the CDx + subjects (those identified by CDx assay as more likely to respond to the drug) would be expected to be CTA – (i.e., those identified by CTA as less likely to respond to the drug). See **Figure 2** for a visual representation of this concern. This potential discrepancy leads to spectrum bias when the CDx effectiveness is measured, since patients at the lower end of the CDx "spectrum" (red and labeled as CTA – and CDx + subjects) will not be enrolled and therefore response data is limited to the blue patients that are both CTA + and CDx +. In such situations, the biostatistician must evaluate the impact of the missing data (subjects that are CDx + and CTA -) since the clinical validation goal is to estimate

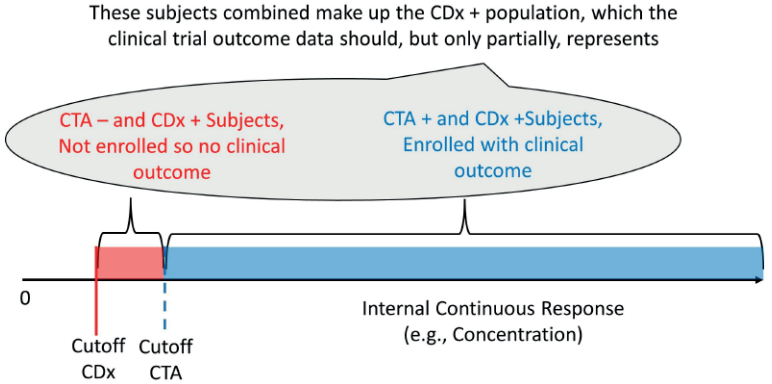
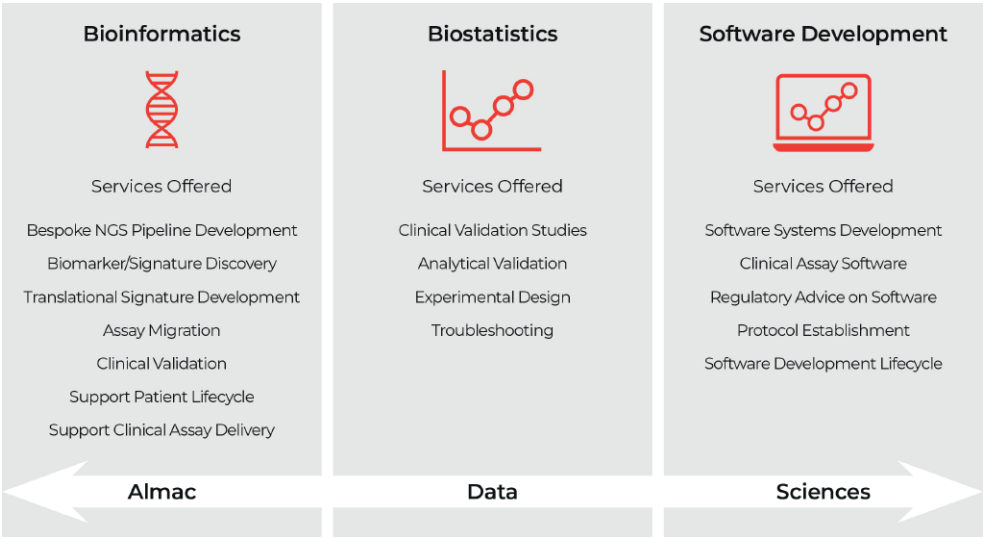


Figure 2: Plot depicting differential analytical sensitivity of two assays, a CTA and a CDx, the latter of which has a lower cutoff and therefore better analytical sensitivity.



performance using the entire CDx + population. This may be represented to regulatory authorities as a sensitivity analysis or theoretical clinical effectiveness estimates.

Protocols For Commercializing An Assay

Once a diagnostic assay has received regulatory approval and begins commercialized use, the manufacturer has new requirements to ensure consistent and reliable production of highly sensitive molecular diagnostics. We describe in this section two important checks that help ensure quality for reagent and control sample release in kit manufacture and surveillance of in use kits.

Acceptance testing protocols

Before reagents and controls can be released for clinical use, the reagent production process, the quality of the products themselves, and the acceptance testing protocols must be assessed (excellent compendia are available on the topic of acceptance testing; for more detail, see Reference 1). A biostatistician collaborates with the production team to develop an acceptance testing protocol, which typically integrates a probabilistic method to assess the suitability of components of an assay.

A knowledgeable biostatistician can help ensure that the risks to the manufacturer and consumer are appropriately mitigated using the most efficient acceptance testing design. The biostatistician should work with the manufacturing team to discuss options for the outcome that is assessed such as qualitative or quantitative variables, as well as simple, multiple, or sequential sampling plans. This is combined with probabilistic risks at both the consumer and manufacturer level to derive a sampling plan. One common plan is the (n=59, a=0) single attributes-based plan, where 59 random samples are evaluated and if the number of defectives does not exceed 0, then the probability that the lot contains >5% defective units is less than 5%. This sampling plan also has a 95% chance of passing a lot with a true defective rate of 0.087%.

Reagents and controls

Process controls are commonly run with clinical samples to ensure that no contamination is present and that the assay is working appropriately. Making such controls that perform reliably can be difficult, and manufacturers typically check that controls perform properly before being released. By establishing acceptance testing to assess the reliability of a process control lot, the consumer is protected from receiving an unacceptable lot (and subsequent false results) while also protecting the manufacturer from accidentally failing a good lot.

Even though reagents and controls have been assessed for quality prior to release, surveillance of assay performance in the testing lab is also necessary. Surveillance of the controls across clinical sample runs can help determine if the overall process, which includes all components including materials, operators and equipment is in control contemporaneously. Surveilling controls allows for the quicker identification and correction of any problems that could lead to erroneous results. Statistical process control (SPC) is an exceptionally powerful approach to ensure that the testing process is "on target with minimal variability."[2]

Concluding Remarks

The goal of this article was to provide insight into how biostatisticians can be used to develop, validate, and support commercialized molecular companion diagnostics across the life cycle of the production and release process. Biostatisticians play a central role in the identification of variables to monitor, how to collect and compile data, setting limits, and producing visual aids for the life cycle, including surveillance chart reviews and interpretations, ensuring that data is interpreted correctly and overseeing the SPC system. This holds true for all phases, from development to commercialization, and allows the biostatistician to appropriately design studies, propose sample sizes, and understand and communicate the risks involved at each step. PMQ

Robert Makowsky, PhD

Robert leads the biostatistics team at Almac Diagnostic Services. He holds a PhD in quantitative biology from the University of Texas at Arlington and did his post-doctoral fellowship in the Section on Statistical Genetics at the University of Alabama at Birmingham. He has spent his graduate education and career working on statistical problems, generally involving genetic data. He has experience working at the US FDA, small startups, large device manufacturers, and as a consultant. With the support of Almac Diagnostic Services, he is an active contributor at the Clinical and Laboratory Standards Institute (CLSI).

Sian Dibben

Sian is a member of the Biostatistics team at Almac Diagnostic Services. She has a Masters is Bioinformatics from Manchester University and previously worked for CRUK before joining Almac 11 years ago. She led the bioinformatics team at Almac Diagnostic Services until recently joining the biostatistics team. She is experienced in biomarker discovery and subsequent development of validated clinical assays.

About Almac Diagnostic Services

At Almac Diagnostic Services, we partner with a range of global biopharma clients to design, develop, and validate molecular diagnostic assays. Our support reaches far beyond the physical components of an assay, though, because a successful CDx requires appropriate design control, software, clinical testing, and commercialization. While many diverse personnel provide support throughout the process, one vital contributor is the biostatistician. For more information on Almac Diagnostic Services' range of Data Science solutions including bioinformatics, biostatistics and software development, visit: **https://www.almacgroup.com/diagnostics/supporting-services/data-sciences/**

References

1. Taylor, W.A. (1992). Guide to Acceptance Sampling. Taylor Enterprises Inc, Lake Villa, Illinois.
2. Wheeler, D.J., Chambers, D.S. (1992). Understanding Statistical Process Control. SPC Press, Knoxville, Tennessee.