BIORANKINGS
TECHNICAL REPORT SERIES

Virtual Biostatistics Core

A New Model for Life Sciences R&D

On May 25, 2021, the Wall Street Journal published the obituary of Paul Songer, founder and CEO of Furnco, a large bricklaying and construction company. The last sentences read:

> *Many steel companies at the time were using their own bricklayers and other craft workers. Mr. Songer offered a faster, more efficient alternative by contracting with his outside company.*

Mr. Songer raised a classic build or buy business decision. The build option allows steel companies to own the entire pipeline from making the steel to putting it together into buildings. This option requires hiring and managing workers from a myriad of crafts and levels of expertise. The attraction is control and capturing all the profit of the building process. However, by owning the pipeline their core strength – steel manufacturing - is diluted and likely will lead to lower quality and less innovation across the entire process which hurts them competitively.

The buy option makes three things possible. One, steel executives can focus on their core strength and innovate the steel making process to be more efficient, be done at lower costs, and make steel with higher value. Two, Mr. Songer can focus on his core strength and innovate the construction process making it more efficient, be done at lower cost, and make buildings with higher value. Three, working as partners can lead to more buildings requiring more steel.

Biotech companies face a similar build or buy decisions with respect to biostatistics. The build option allows them to retain control of the entire R&D process. By hiring a biostatistician or accepting a substitute, they dilute their core strength in biology by having to manage a technical area that at other institutions – universities and large pharma – is almost always obtained through a core service. By accepting a substitute, a second option for biotech, they weaken their R&D pipeline by getting statistics from non-statisticians.

This Technical Report discusses the Build vs. Buy dilemma and shows the advantage of buying a Virtual Biostatistics Consulting Core service.

CORE STRENGTHS OF BIOTECH

Biotech companies commercialize biology for medical and agriculture markets. To investors they are initially unprofitable, have unusually long development times to product approval and sales, and have a high failure rate.[1] Yet investors pour money into these companies because of the ROI if they succeed.

The biotech business is often centered around academic scientists/founders with support from entrepreneurs, often the scientist, and investors.[2] The scientific discoveries from the academic labs to be commercialized are licensed to the company at early stages of development requiring years of additional basic R&D.

With this model, the biotech start-up can be considered an extension of an academic lab requiring the same skill sets including wet bench science, biological specialties (e.g., immunology, neuroscience, genomics or other -omics, molecular biology, chemistry), strategic vision, innovation, and scientific creativity. Computational support for data management and bioinformatics is often required.

Biotech start-ups use core services to supplement their technical expertise in a cost effective way. An agriculture animal feed company whose core strength is nutrition, may want to measure the impact of an additive on the animal gut microbiome. They can build those services to control the process by buying a sequencing machine and hire staff to run it and create a bioinformatics pipeline for annotation.

Or they can buy these services and send samples to a microbiome sequencing company which have the staff and pipeline in place. The same company may want to run animal studies of their compounds before testing it in farm animals. They can build and manage their own mouse facility or outsource this to an existing animal model company who have the facilities, staff, and expertise to do this work quickly, accurately, and at lower cost.

Since microbiome and animal studies are not part of their core strength, and likely not be needed on a regular basis, buying these services from vendors who can share costs across different clients and have optimized the processes already can be cost effective, efficient, and faster than doing the work themselves.

As mentioned, many biotech start-ups come from universities who provide researchers core fee-for-service for these same reasons. It distributes the operating costs of the cores across all the users and allows cost-efficiencies and innovation to be added more easily. Biostatistics are core services available at almost every university and few academic labs hire their own biostatisticians. (Unfortunately, labs often will use inferior substitutes, which we will discuss below, not understanding how wrong the results can be and how they hurt their research.)

CORE STRENGTHS OF BIOSTATISTICS

Biostatistics is about summarizing and making inferences from experimental data. This work requires an advanced understanding of mathematics and statistics to be done right. Some of the core strengths and services of biostatistics include:

- Hypothesis generation
- Experimental design
- Sample size, and power calculations
- Statistical analyses and machine learning
- Interpretation of results
- Developing new analysis methods
- Writing Statistical Analysis Plans and reports
- Writing software for pipelines
- Statistical review of proposed research
- Education

Biostatisticians are trained in basic principles and can apply their skills across biotech specialties. For example, analyzing plant growth with different fertilizers in different fields and measuring cholesterol reduction from a new drug over time in patients are both analyzed using analysis-of-variance (ANOVA). However, biostatisticians have deep understanding of ANOVA to know the plant study use a random block design and the cholesterol study a mixed model with contrast to measure the rate of cholesterol decrease over time. In other words, deep understanding of statistics allows the optimal method to be used for a specific experiment, and therefore increase the probability to obtain a significant result.

Interpreting statistical results also requires statistical training. For example, claiming a result is real when it is not (i.e., false-positive) wastes money and time in follow-up experiments. This is particularly alarming, and produces a significant risk to the company, because high throughput technologies generate lots of variables where false positives are guaranteed to be found.[3] Having results interpreted by experts who understand how Large P Small N data leads to an explosion in false positives, or how regression diagnostics are needed to ensure your conclusions are not the result of a single outlier (which can easily happen), will save you money and wasted time pursuing the wrong leads.

COST OF BIOSTATISTICS SUBSTITUTES

Biotech executives should ask if they are using a biostatistics substitute and evaluate if this is a good decision for the business. Substitution generally happens in one of two ways. First, many companies have someone who applies statistics to their data but is not a card carrying statistician (a substitute). Often these substitutes apply statistics to data with no understanding as to why. In these cases, an overabundance of graphs, charts, and numbers are presented to look knowledge-able but as the results are interrogated an inability to explain analysis rationale is quickly uncovered.

Bioinformaticians are often called on to be a substitute since they have expertise in data management and computing. However, bioinformaticians are not trained in biostatistics as is made clear by comparing the core curriculum at Boston Univ. for PhD programs in biostatistics and bioinformatics side-by-side (See Table on Page 3). Bioinformatics PhD students are required to take 1 introductory class in statistics, 1 in computational biology, and the rest in research. In contrast, biostatistics PhD students take a dozen classes in advanced statistics. These include mathematical statistics and probability, estimation (how do you summarize data), hypothesis testing (how do you calculate P values), linear models (how do you do regression and ANOVA), and other advanced biostatistics.

A second common substitute is the “do-it-yourself” lab scientist with some experience analyzing data, perhaps from grad school or their post-doc. Their knowledge is generally limited to a few statistical methods used in earlier research, is not based on strong mathematical and statistical foundations (their training is biology and lab science), and lacks the breadth of experience to bring in

Boston Univ. PhD Core Courses	
Biostatistics	**Bioinformatics**
Linear Models	Computational Biology: Genomes, Networks, Evolution
Probability	Accelerated Intro. to Statistical Methods for Quantitative Research
Mathematical Statistics	Bioinformatics Challenge Project
Estimation Theory	Legal & Ethical Issues of Science and Technology
Hypothesis Testing	Laboratory Rotation System
Intermediate Statistical Computing and Applied Regression Analysis or Multivariate Analysis for Biostatisticians	Research Opportunities in Bioinformatics
Statistical Methods for Epidemiology	Bioinformatics Graduate Seminar
Analysis of Correlated Data	1 Elective in computer math
Generalized Linear Models with Applications	1 Elective in biology
Concepts and Methods in Epidemiology	1 Elective in non-research
3 Electives advanced biostatistics	
3 Electives from advanced math/biology/CS	

other statistical approaches when needed (e.g., knowing about categorical data analysis is unlikely when your background has been with continuous data). Spending time analyzing data means they will have less time for the lab science they were trained and hired to do. This takes their focus off science and has that person operating outside of their core area of expertise and comfort level.

COST OF INACTION

There is a tendency to overvalue what one already possesses, leading to inaction and a status quo bias. The cost of inaction must be considered when evaluating build or buy options. In many cases the cost of inaction can be high.

In addition to doing the wrong analyses and getting the wrong results and conclusions by not having a biostatistician, biotech leaders, researchers, and investors express wonder associated with FOMO. FOMO is the fear of missing out. Are there insights lying dormant in this data? Have sample size calculations been optimized and therefore money not wasted? Will these analysis results stand up to outside rigor?

Substitutes, DIY, and FOMO lead to inefficiencies in the data analysis which can have a direct impact on companies by increasing costs and time for product development. It is baffling why biotech will invest in hiring the right PhDs – companies developing treatments for neurological disease hire PhD neuroscientists – and contract for core services for biology, but accept biostatistics substitutes blindly and without the expertise to know if the right analyses were done and the right interpretations made.

VIRTUAL BIOSTATISTICS CORE

Our service provides biotech with the Build or Buy Biostatistics decision similar to steel manufacturing and building solved by Mr. Songer. He was able to show the steel manufacturing industry that some services are better bought than built in-house – after all, what do steel workers in Pittsburgh know about carpet installation? In the same way our Virtual Statistical Consulting Center gives your company the biostatistical analyses you need, not what a substitute can provide.

If you want to explore your Build or Buy decision for biostatistics let us know at contact@biorankings.com and we will schedule a time to talk.

REFERENCES

[1] https://www.investopedia.com/articles/fundamental-analysis/11/primer-on-biotech-sector.asp

[2] https://techcrunch.com/2019/08/09/biotech-researchers-venture-into-the-wild-to-start-their-own-business/

[3] https://blog.dominodatalab.com/the-curse-of-dimensionality/