The 5 most common pitfalls when developing biomarkers

Liesbeth Ceelen, General manager at BioLizard

Discovering a strong biomarker is an expensive and time-consuming process. One of the key takeaways is to take your time and look at the data before you start testing for biomarkers. Why? Because junk in equals junk out. **You can run the best models and the craziest statistical analysis. When your data is not up for the test, you'll lose a bunch of your precious time.** We can't imagine this creates fun status meetings with your CEO.

Throughout the years, we have seen a lot of mistakes being made within collecting, storing and using data. To prevent you from making the same mistakes, we'll share the five most common pitfalls when developing biomarkers.

How was your available data generated?

The first thing you need to know is if you have enough data available? When it comes to the amount of data you need, the answer is simple: the more, the better. **Sometimes, you think you have enough data. But in reality, you do not.** This mistake happens when you don't know how much data is actually needed to generate a reliable biomarker signature. It also occurs when you are unsure of how the available data was actually generated.

Imagine this: you have a decent amount of data. Some of it was generated one year ago, some two years, and some five years ago. The problem is that you don't know if the data from five years ago is generated in the same way as the more recent data. This makes statistical analysis a lot harder, and you'll need more data samples per set.

Different stages, different data. Track it.

Timing your data-mining is essential: you'll need enough data during different patient stages. **The most significant breakthroughs come from looking at differences within data over time.** Let's say a patient develops colorectal cancer. In that case, you'll need data from the moment of diagnosis, after the first treatment, after surgery, after one year, two years, three years, and so on. This is where companies discover their biggest breakthroughs. And when companies don't? It's often because they do not have enough data throughout those different stages.

We realize you can't track someone for the rest of their life. That's usually a waste of time anyway. But the key takeaway here is to keep your eyes open to when it's the most relevant time to track data. In our experience, **mastering the art of identifying these opportunities brings valuable information at a minimal cost.** Sounds like music to our ears!

The most overlooked clinical variable.

For your study to be successful, you need hundreds of people to enlist. When you build your case and control groups, you must add some variety: different age groups, sex, and ethnicity. All these characteristics have an impact on the effectiveness of a biomarker. Luckily, **there is a large enough variety in age and sex within most research. But ethnicity is often overlooked.**

For example, it is known that people with an Asian ethnicity have enzymes that take more time to break down alcohol. This makes them have a two-day hangover, while the rest of us are usually fine after one day of feeling like crap. This shows that genes differ in different ethnic groups, and so shall

your biomarker's effect. In other words: try to enroll enough people with different ethnicities in your test- and control groups.

Limited metabolomic and proteomic coverage.

When generating data – depending on the type of experiment – you'll generate five, ten, a thousand, 20.000, or even more proteins. These proteins develop a variety of data accumulated per entity (human, mouse, monkey,…). This is essential: **when your coverage is too small, you might not be able to find discriminating signals that distinguish cases and controls.** This means: no data to identify strong biomarkers. This then means: a big '*sayonara*' to all your invested time & money.

The wrong structure in your dataset.

The structure of your data is essential.

The lack of information on protocols means the generated data is almost impossible to interpret. Companies tend to overlook this aspect.

Most of the time, you can find out how data was generated. But in some cases, it was people who are no longer at the company that generated the data. This makes it almost impossible to track down how the data was collected and processed. Can you hear the bin opening up in the background? Yup, that's your valuable data gone.

Moral of the story: take care of your data.

These are just 5 of the most common pitfalls in finding the right biomarkers. We'd love to provide you with a more comprehensive checklist on how to build the perfect dataset. But discovering biomarkers is extremely research-specific. This makes it impossible to give you a one-size-fits-all

solution. Keeping these 5 pitfalls in mind is a good first step, as it will save you a lot of trouble. However, if you still need an extra pair of experienced eyes? We're also glad to help.

Remember, your data is the start of everything. Within it, you can find hidden treasures, or mess up parts of your research.

Want to make sure your data isn't junk?

Contact us

We'll tell you how we can help!