Building Healthy Analytics to Guide Better Decisions

By Sidique Jarr-Koroma
and Michael Kelly

Let's assume you're aware of the power of data to enable better strategic decisions. And let's say you now have an opportunity in your organization to push more data-driven decision-making. Maybe you've been learning about data science, or maybe the COVID-19 pandemic opened your eyes to how data impacts critical decisions, or maybe you've taken a new position and this is your first opportunity for improving analytics.

Regardless, if you're making decisions with significant impacts on your business, you want to make sure the numbers are guiding you. As advocates for analytics, we fully support your turn to data for that guidance. There is a catch, though. We support only *good analytics*. Making strategic decisions based on low-quality analytics can lead to multimillion-dollar mistakes and erode your firm's competitive advantage!

We'll discuss three key factors that will help support your new data-driven decision-making:

☐1

ENSURE THAT YOUR DATA ACCURATELY REPRESENTS YOUR REALITY

Whether you've recognized the need for increased visibility on your own or are incentivized to report out key metrics, your organization is now taking steps toward becoming more data-driven. A few starting questions in this journey should be: Are the processes that create your data standardized? Are they well governed? Is the data itself well governed? And finally, Do you trust the outputs? Until the answer to all four is yes, the data collected will never accurately reflect what is happening in your organization.

When you're making decisions with significant impacts on your business, you want to make sure high-quality analytics are guiding you.

To exemplify what we mean, imagine you are managing the emergency department (ED) at a hospital. While there are many complex aspects of running an emergency department, the one we are going to focus on now is intake. There are several key metrics that indicate the efficiency and efficacy of your ED's intake process. Some of these metrics are performance-based, like measuring the time until a walk-in patient is seen by a physician or other provider, while others are more descriptive of the patient population, like capturing the number of patients suffering from similar ailments. The standard metrics used for external reporting by many hospitals often have long and technical names. The one we will use for this example is a performance-based metric, called "Initial ECG Acquisition Within 10 Minutes of Arrival at the emergency department in Persons With Chest Pain." Simply put, this metric measures your ED's success rate of performing an electrocardiogram,[1] or ECG, on a person experiencing chest pains within the first 10 minutes of their arrival.

The key considerations for this metric include: (1) When does the 10-minute clock start? Is it when the patient walks in the door or when they first report chest pains? (2) Who is responsible for starting the clock and tracking time during this critical moment? (3) How are they tracking completion? Does the clock stop when the ECG is first ordered, attempted, or successfully administered? When is the ECG officially recorded? (4) Are the answers to the preceding questions consistent across all staff shifts? Are hospital leaders, administrative staff, and practitioners all aligned on the answers to the preceding questions? These considerations reveal that even though a process to accomplish a task is standardized, that may not directly translate to consistent and accurate data capture. Each question represents process variables that must be defined to attain consistent data capture.

The uncertainties around how the data for this critical metric is captured diminish the trust that it is an accurate representation of reality. This example demonstrates how thoughtful data collection is just as important as good process control. Steps should be taken to ensure that the data being collected paints an accurate picture of your reality. In this example, clear parameters and governance around when the 10-minute window starts, who starts it, and what actions constitute a successfully administered ECG set by the process owner are critical to consistent and accurate data capture. Inappropriate data collection here, and for other metrics, will invalidate your analysis before you even start. Once you've cleared this hurdle, you can move to the next step.

02

UNDERSTAND WHAT YOUR DATA CAN TELL YOU

At this point, we'll assume that you have quality data on your hands. The processes to generate and collect your data are refined and well governed, maybe even automated. The next key element is understanding what your data is able to tell you. The best way to think of this step is, What *can* I learn from my data? instead of, What do I *want* to learn from my data?

Many people make the mistake of collecting data, manipulating it until it aligns with their hypothesis, and then calling their decisions "data-driven." Manipulating data until it fits your predefined narrative is not data-driven decision-making. Following your data wherever it leads is. To do this, you should first explore your data without defined targets. This type of analysis is often called "unsupervised learning" and includes techniques like clustering and trending. We won't get into the specific math, but the goal of these types of analysis is to understand what type of information, insights, and complete pictures *can* be gleaned from your data. Let's walk through an example of the tension between what you want from data versus what your data can actually tell you.
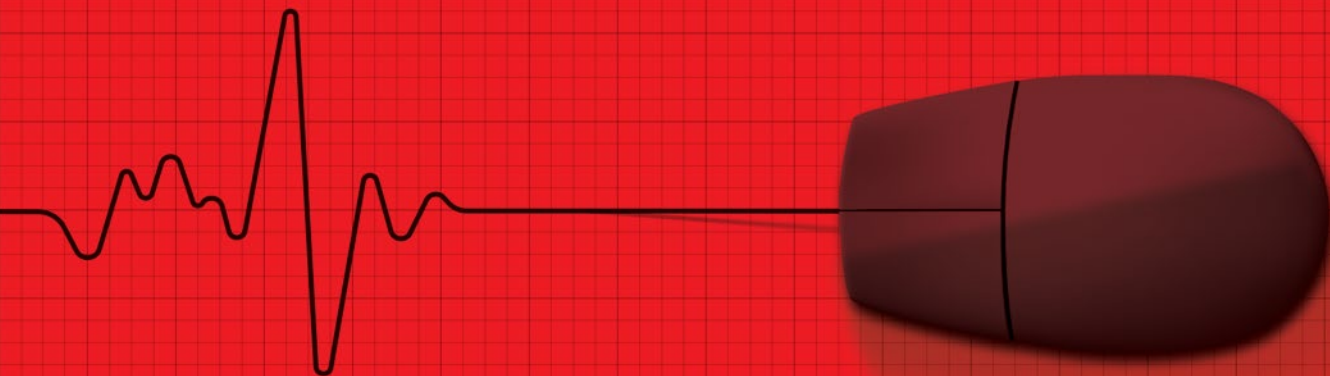
We'll continue with the emergency department example. Patients with a choice between emergency rooms tend to prioritize speed. To attract patients, it is becoming increasingly common[2] for emergency departments to advertise short average wait times. While there is no universal definition for this wait time, it is likely we want some version of "Time to Provider" (TTP). This is generally calculated as the time from patient arrival until they are first seen by a physician, physician's assistant, or nurse practitioner, and is a common metric that you are probably already capturing for every patient. But what version of TTP is the most appropriate?

If it is purely a marketing-driven effort, you may be tempted to either guarantee some maximum wait time or anchor patient expectations with an average. Unfortunately, neither of these is going to work out very well. Some patients, such as those arriving on an ambulance already in critical condition, have an extremely short TTP, and other less urgent cases, such as walk-ins with a sprained ankle, may wait several hours. Furthermore, many emergency departments see dramatically different volumes throughout the day and week. A weekday evening arrival will likely confront a much less crowded department than an early-morning arrival, and their wait time may vary greatly as a result. After looking at several weeks' or months' worth of recent data, you'll realize that to "guarantee" a maximum wait time you would have to set it at several hours, which isn't going to help your marketing. And with so much variability, an average, whether mean or median, isn't going to be representative of many patients' experience.

Given these challenges, we could reevaluate the approach. What we wanted to do was advertise how quickly our emergency staff could treat you. What we found was wide variance and seasonality. "See a doctor in 30 minutes, as long as it isn't too crowded, and your injury is serious enough" just doesn't feel like a great slogan. Some emergency departments choose to display "live" averages on billboards or websites[3] to try to outmaneuver the seasonality issue, but the triage process will always prioritize sicker patients (as it should), making these averages a loose guideline at best. Perhaps the marketing team could focus on quality instead.

Of course, we've gone to all the trouble to collect and validate our Time to Provider data, so it would be nice to find something useful to do with it. If we followed an unstructured learning approach from the outset, we might find some patterns that are useful. For one thing, looking at repeating patterns over time, we'll get a great sense of when the department is most crowded—or, more specifically, when the demand for health care services of incoming patients exceeds the supply.

While this data didn't meet our needs for marketing, it turns out you *can* still use it to make operational improvements. An operational leader could consider capacity levers, especially reviewing staffing schedules to adjust personnel levels to match the need. It is possible that the critical constraint is physical room space or something else harder to change. But, leveraging your findings, you could try Lean experiments or other process improvements and keep collecting data to see what might move the needle.

The takeaway here is that good data collection doesn't guarantee a particular data-driven story. You have to follow the data to see what appropriate conclusions you can draw, and not the other way around.

03

USE YOUR DATA APPROPRIATELY TO GUIDE FUTURE DECISIONS

If you have good data collection, and a good understanding of the information contained in the data, you can move on to our third step: determining whether the data can help you make strategic decisions.

To explore a strategic decision that is directly enabled by good analytics, we'll consider trying to make a facility investment decision. In our hypothetical, you are again running the emergency department of your hospital. There are 40 patient rooms in the department, but 10 of them are small, out of date, and less capable of supporting sick patients. You believe quality of care may be lower in these rooms, and they can't adequately treat as many patient conditions. Of course, major construction will require a significant capital campaign, so it needs a clear justification. Ultimately, you would like to advise the hospital leadership to either renovate these rooms, change the number of rooms through reconstruction, or find an alternative use for the space entirely. We know you are an experienced consumer of strategic data and know how to answer these questions.

To guide the decision, we need to understand how many rooms the department needs at a time. We could probably look at an average (multiply total patient arrivals by average length of stay) over the past year or two, but we are trying to predict future need over many years, with potentially changing conditions. Our past data will contain seasonal patterns, as well as long-term trends related to the overall community and our relative market share. Separating these effects from each other in order to build a forecast could require years of internal data to sort out. There could be too many variables to control for, and we would find ourselves in a similar position to the marketing-driven effort discussed above, and we'd fail the same data tests. The information we have internally isn't enough.

Instead, we'll project the need for emergency space from more causal variables. Nearly all our patients come from within this local area (let's say a county), and government data is available to show population and aging trends by geography. In the health care industry, there are groups collecting data on total utilization of various hospital resources by age bracket. Combined, these data sets can give us both past and future estimates of total emergency room visits within the county. Now we layer in our own data on usage over the last few years to understand our market share and any changes thereof. Instead of just projecting our patient count data forward, we can now use our market share trends against the larger trends (with their greater statistical heft) and estimate total patient volume with greater confidence.

Once we have estimated arrivals, we can combine with other operational metrics such as length of stay to build a Monte Carlo[4] model that will simulate future needs. We'll again skip the specific math for this article, but you can model how often the "Nth" room in the department is needed. If built appropriately, this alternative would even allow us to use the recent internal occupancy data as a validation set, increasing confidence in the model. With knowledge of how often each room will be used, our team can use financial metrics and determine that either (a) the department can get by just fine with only the 30 existing high-quality rooms, (b) it absolutely needs to upgrade for a total of 40 (or more) to serve the community, or (c) something in between. Some cost and revenue estimates, requiring additional data, can help to decide whether an investment is appropriate.

Ultimately, we believe in the value of data and analytics to drive strategic decision-making, but recognize that value comes only from high-quality data and good analytics. To give yourself the best chance of success, you first have to ensure that your data collection is consistent, accurate, reliable, and meaningful. Having achieved that, don't just ask a preformed question—make sure you understand what information the data can tell you. With that information in hand, when faced with critical strategic business problems, you can develop key insights to guide your decisions.

All of these steps are not single-use processes, nor are they limited to individual departments. The more they are adopted at all levels of your organization, the better. Quality data and analytics will support continuous improvement and better decisions. And with these practices in place, when conditions change (whether it's sudden and dramatic during a pandemic or more mundane), you'll be more prepared to respond appropriately and maximize your chance of success. ✛

Sidique Jarr-Koroma
sidique.jarr-koroma@jabian.com

Michael Kelly
michael.kelly@jabian.com

Sources:
1   *Electrocardiogram is a record or display of a person's heartbeat used to confirm that the heart is functioning normally*

2   *Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628484/*

3   *Source: https://health.howstuffworks.com/medicine/are-er-wait-time-billboards-accurate.htm*

4   *Monte Carlo simulation is a probabilistic method that uses randomized events to show likelihoods of various outcomes, often of complex systems*