encore
HEALTH RESOURCES

Smart, Skinny Data vs. BIG DATA

By Dana Sellers, CEO, Encore Health Resources

One of the most popular phrases in all of information technology today is 'big data.' Maybe I should say BIG DATA. As the volume and type of data grows – the human genome, social media, video surveillance, and all kinds of imaging – the topic has become incredibly popular. Recently, at Encore, we started talking about a concept we call "smart, skinny data." We even sent out a press release about the subject, and Mort Meyerson, one of our Board members, sent me back a note. Now Mort's a guy who knows more about healthcare and data than I'll ever know, so I listen when he talks. He told me to be careful not to be dismissive of Big Data. Mort always makes me think, so I started thinking more on the subject…when do you need BIG DATA, and when do you just want a little smart, skinny data?

Big Data offers incredible opportunities for discovery with either pure data mining or loosely formed hypotheses. Researchers in London, mining data without a hypothesis, found a correlation between temperature increases and suicide rates. Here in the US, research found that social media data was nearly as good a predictor of the movement and prevalence of the flu as the CDC's clinical data. And I have a friend who has been working with claims and pharma data to predict heart attacks. Wow. Now that's pretty cool! He believes he can increase the accuracy of his model by adding social data, but it may be hard for a while. The people old enough to be having heart attacks aren't generally big users of social media—at least not yet!

But, while big data offers big opportunities, it also requires big effort. One definition of big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The FBI applied big data technology to the hundreds of terabytes of videos and pictures taken by individuals and surveillance cameras at the Boston Marathon with quick and successful results. That makes it clear that we should use big data when we have the resources to chase discovery or when we have a big problem to solve. But, it doesn't make it right for every situation. Big data might be a good way to identify major correlations or trends, but wouldn't be the best way to care for a specific patient at the bedside. And big data requires big resources.

Another popular phrase, particularly in healthcare, is "a single source of truth." For most hospitals and health systems, this typically means an enterprise data warehouse or EDW. While less complex than compiling and manipulating big data, this effort still involves combining data from across a heterogeneous application environment to represent the clinical, financial, market, and operational information about a complex business. An EDW is useful when you have some idea what you're going after and have the time to collect data from many sources, aggregate it, normalize it, and mine it for the answers. But it usually takes a while to collect a critical mass of data and start getting answers. Further, driving change from the insights you gain is hard unless you've been very disciplined in how you capture data. Before stakeholders want you to use information to drive change, they want to know how you captured your source data. Too often, they'll say, "that data's not accurate! "or "I don't believe that's

true!" All your efforts are wasted if you can't drive performance improvement or improvement in outcomes because your data's not trusted.

Given all this, analysts often turn to smaller data sets about specific disease states to focus their analyses. Many healthcare organizations refer to these as registries. Disease specific repositories allow for focused analysis and research and don't require the analyst to sift through information that isn't relevant to their analyses. They are convenient tools to track the natural history of a disease, evaluate the effectiveness of treatment protocols and monitor long-term outcomes for the patient population. Problems with registries can occur at the margins of disease definitions. Did I get all the patients I care about? Am I including patients that aren't relevant? As long-term tracking tools, registries require conscientious cohort design and don't lend themselves well to revisions.

And that gets us to smart, skinny data. Smart, skinny data describes an approach that focuses on capturing just the data you need to solve a specific problem, with a minimal set of surrounding data that might be relevant. Because you're not trying to boil the ocean, it's possible to exercise very tight data governance, to minimize the integration effort and capture data in a rapid timeframe, and to create a "chain of trust" so that stakeholders know exactly where the data comes from and what it means. Then, when you want to use that data to drive change, stakeholders are more willing to accept change because they trust the data that's driving the conclusions. The exciting part is that you're able to drive performance improvement and better outcomes from your data insights, and as your organization sees results, you can go back and bite off a little more smart, skinny data and tackle another strategic initiative.

I spoke with a hospital administrator recently who was suffering from metrics overload. He told me that he gets a crowded dashboard that has all kinds of pretty colors on it but it doesn't tell him what his biggest problem is or what levers he should pull right now to right the ship. At Encore, we focus on the data we know will be needed to support key strategic initiatives. Rather than trying to boil the data ocean, we focus on the data required to support these key initiatives. With so much going on in healthcare today, it's more important than ever to be focused on value.

As I thought about Mort's challenge about smart, skinny vs. BIG DATA, I reached the conclusion that Mort's right (as usual). Smart, skinny data isn't the only answer. Interestingly, we are seeing lots of opportunities where the data we rely on actually comes from an EDW, but we pull it into a smart, skinny data warehouse where we can perform trusted analytics on it and drive change. These folks are looking for smart, skinny solutions, either as an accelerator or as an application to help drive value out of their EDW.

There are many problems in healthcare that are so big we'll only solve them through big data. Big Data has huge promise in healthcare, but it's not the only tool in our kitbag. Let's start somewhere…somewhere we can show some results right now.

# # #