“Lies, Damned Lies, and Statistics”

In This White Paper

Statistical Lies and Misdirection .................................................................................................................. 1
What Makes a Good Metric? ............................................................................................................................. 2
The Problems of a Small Sample Size .............................................................................................................. 2
Honest and Dishonest Measurements ........................................................................................................... 3
So What Can You Measure? ............................................................................................................................ 3

Mark Twain wrote the title line in his autobiography, citing Disraeli – and the sentiment probably predates that British Prime Minister as well.

Some professionals ask about Legal Project Management, or anything else that’s different: What are the statistics to back up these claims?

I’ve seen a number of metrics and statistics, and even contributed a few myself. However, most, when analyzed, lead back to the Twain/Disraeli quote.

Statistical Lies and Misdirection

One problem is that it’s easy to make any number do anything you want in statistics.

For example, do want to make your scores go up – client satisfaction scores, say, or organizational-health indexes? It easy. Do two surveys (called a longitudinal survey / study). On the first, arrange the list of answers with low scores on the left:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9

Now repeat the survey in a few weeks or a month, but this time arrange high scores from the left:

☐ 9 ☐ 8 ☐ 7 ☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

Scores will improve from the first survey to the second.

Here’s another example. Managers at a call center (that shall remain nameless, but I have direct knowledge of the event) decided that a good measure of customer-representative effectiveness would be the length of phone calls, with shorter being better. Their intentions were good. After all, someone who could answer a question in less time was more valuable, both because they presumably were better at finding an answer and because they could handle more calls in a day.

(Before reading on, you might want to guess the unintended consequences.)
One result: representatives became abrupt with callers, since it was imperative to get them off the phone.

Another result: Very difficult, time-consuming problems correlated with a higher incidence of “technical difficulties” such as disconnections… as in employees breaking off a call that would kill their numbers in hope that when the customer called back someone else would get stuck.

(The good news with this example is that the manager recognized the problem within a month and dropped the metric. The bad news is that such management behavior and acknowledgement of mistakes is unfortunately rare.)

What Makes a Good Metric?
There are two real questions a manager (or statistician) must answer in any attempt to measure something. Both are very hard to answer in a meaningful way.

1. Figure out what you really want to understand: what measure do you believe can be isolated as a factor in contributing to positive movement?

2. Can you actually measure it?

For example, client retention (technically, repurchase intent) is the #1 metric in pretty much any business. Yet you can’t measure it directly. While it’s fairly easy to isolate behaviors that can drive this metric down, it’s extremely difficult to identify those that drive it up.

So sometimes people measure client satisfaction instead. “CSat” has a positive correlation with retention, so it’s not totally bogus. However, so many outside factors that drive retention that the correlation is relatively low, especially over a small sample.

The Problems of a Small Sample Size
It’s the small sample size that makes measurement of professional programs really, really hard.

Small samples present two problems. One is endemic to any small sample – how do you be sure your sample is representative of the larger population? This problem manifests in medical research and engineering trials just as it does in the legal world.

There are statistical methods to define a “good” sample, usually striving for a 95% “level of confidence” – meaning nineteen times out of twenty, real-world results will match the sample results. (That’s a very simplified explanation of confidence, but good enough for purposes here.)

That sample size required for a valid survey is considerably larger than the number of pilot projects possible at any legal practice, with a handful of exceptions. (E-discovery, with millions of documents, in an exception. So is patent prosecution by a large corporation – e.g., Microsoft or IBM, with many thousands of very similar filings each year. Beyond that, you can find occasional one-off situations, but they’re uncommon in “BigLaw.”)

The second problem concerns any metric that measures the way people “feel” about something, whether political candidates, their team’s chances in the playoffs, or Legal Project Management.
Fights with a spouse, getting the kids to school late, or the throb of a twisted ankle all have highly significant effects on these types of scores. For example, during the DOJ/Microsoft trial in the mid-1990s, Microsoft survey teams could often tell how the press reported the courtroom goings-on the previous day by very significant fluctuations in daily customer-satisfaction scores! And that was with thousands of people polled on a given day. How much bigger – and unpredictable – are the effects when you measure a couple of dozen respondents?

These effects, too, can be smoothed out with repeated large-sample-size surveys, but few if any law practices have enough people to make these surveys valid or to perform a real multivariable regression analysis – to say nothing of inevitable “survey fatigue.”

Honest and Dishonest Measurements

Very little in the legal world meets the two-question criteria honestly – a distinct behavior-driving metric that you can measure directly. It’s always possible to gin up numbers to prove or disprove pretty much anything, and most of the metrics I’ve seen – not just in legal – seem to fall into the category of interesting but statistically suspect numbers. Most such metrics exists as superficial answers to a manager’s request rather than real attempts to measure progress.

(Just because a number has no statistical validity doesn’t mean it’s wrong. Rather, it’s not necessarily true, not as valid as someone might claim.)

One potential measure of Legal Project Management effectiveness I’ve heard is hours per matter, with fewer being better. However, such a number is meaningful only if you can identify parallel matters (among other problems). Such a measure could prove valid in large-scale patent prosecution or e-discovery, say, but it doesn’t align very well in other environments. And even if you could measure it, it might not tell you what you want. Some lawyers might use the time savings to probe deeper or do better work on a given matter within the same number of hours, for example. That would also be a positive outcome but wouldn’t show up in your measurements.

So What Can You Measure?

Not all metrics must be quantitative. Qualitative (e.g., “feel”) metrics may lack statistical validity, but as noted above, they can be quite useful. Indeed, they’re often more honest that good-sounding but bogus quantitative measures (e.g., “four out of five dentists recommend…”).

For example, ask these questions of pilot teams:

- Do you feel better about your job, about the work you’re doing?
- Do you feel more in control of the work, the clients, your time?
- Does the team function more smoothly?

In my experience, the effects of Legal Project Management are often significant enough, especially in self-aware pilot teams, to deliver reasonable valid answers to these questions. (“Reasonable” means useful, not statistically valid.)
If you’re trying to gauge whether a program is working, consider simply ask these types of open-ended questions and then talk about the results. For one thing, if people believe they’re making progress, not only is it likely to be true, but the feeling of forward movement will encourage them to keep at it, to keep improving.

If management insists, of course, you can create numbers from surveys and such.

And even statistically shaky numbers needn’t be damned lies. Even if the exact number isn’t meaningful in itself, it can point to the truth.