So You Don't Trust Your Data...

Data integrity is really important if you want to reduce the risk of your decision making.

STACEYBARR

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

Contents

Overview

The management team sit around the monthly performance report, debating. But they're not debating the interpretation of the trends or signals in the data. They're not debating the causes that sit beneath those trends or signals. And they're not debating the pros and cons of various performance improvement ideas. No, they're debating whether or not the data has enough integrity to even consider using it. Not again!

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

What is data integrity and why does it matter?

You depend on the quality of data and information to provide a stable foundation for your decision making. Decision making often involves responding to something, so you need your data to validly describe what you are responding to so that you choose the right responses.

Whether your data is quantitative (based on numbers) or qualitative (based on perceptions), it's integrity depends on 5 widely recognised qualities.

Data must be relevant.

Make sure the data you have selected is directly appropriate to the purpose of the performance measure you selected it for. This is the same as making sure that your data is capable of answering the right question.

Be careful of data that seems interesting: it doesn't mean it is relevant. Trying to gather more data than you really need, especially in surveys, can negatively impact on the other dimensions of data integrity (below).

Take for example, a survey. Surveys are widely used to gather information about groups of people such as customers, employees or attendees at a training course or conference. The most common mistakes made in surveys are usually associated with one or more of the 5 Rs of data integrity.

The way in which the questions on the survey form are chosen and phrased may not be the best for gathering information that is relevant to the survey's purpose. It is important to distinguish between which questions will gather interesting data versus which questions will gather useful data.

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

Data must be representative.

It is important that the data you collect are observable events or characteristics that describe the *full* scope of what your performance measure is supposed to be measuring. This means that it is unbiased.

The last thing you need is for your data to tell you only what the "squeaky wheels" have to say, drowning out the valid and important and balancing views of the "well oiled wheels".

Often surveys are based on samples for reasons of cost and time efficiency, which is fine as long as the sample is chosen at random. Often the sample is a volunteer sample which is not representative of the whole population. In volunteer samples, people have the choice about whether or not they could be bothered to participate in the survey. Their attitudes to what the survey is about strongly influence this choice which introduces bias in the data that is ultimately collected.

For example, a volunteer survey about how satisfied employees are with working in an organisation will likely result in mostly those people with strong opinions responding, which will not adequately represent those people who don't really care either way. You can easily see how decisions based on such data can result in over- or under-reacting to what is really going on.

Data must be reliable.

Collect enough data and collect it carefully to ensure that it is accurate and continues to be accurate as you collect it over time.

Would you rely on one day's rainfall to draw conclusions about annual rainfall? What about five days' rainfall? How many days rainfall would you need to get a good estimate of annual rainfall? And what would this number depend on?

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

Even when your sample is chosen at random, there is still potential for it to be inadequate for your purpose. The reliability of sample data depends greatly on the size of the sample. A sample of 100 people will give you more reliable results than a sample of 10 people.

There are statistical formulae for calculating the best sample size for your purpose and these formulae can take into account the type of analysis you want to do, the type of data you are collecting, your budget and the degree of reliability you want. Interestingly, *the size of your population has very little, if anything at all, to do with the size of the sample you will need*. To get more information about survey sampling, talk to a survey statistician from your local statistics bureau, university or market research firm.

Reliability is also affected quite significantly by the way that questions are asked. If survey questions are ambiguous, leading, loaded or sensitive, then you can't be sure whether the question that was answered was really the same as the question that was asked.

Data must be readable.

Unless the data you collect is clearly defined, legibly presented, makes sense to its users and can be easily interpreted and understood by them, it won't matter how relevant, representative or reliable it is. It just won't be used.

What your local doctor writes on your prescriptions and what a research statistician presents in a 3-dimensional contour chart are potential examples of poor readability in data.

Even after you have collected your survey data, you're still not completely clear of jeopardising its integrity. How you analyse and present your data will determine how readable it is. It is very tempting to ask a lot of questions on survey forms to "take full advantage of the effort we are putting in to collect this data". The drawback is that you are likely to end up with too much information and your audience will drown in it rather than use it.

Making survey data readable means only answering the burning questions that drove the research in the first place and keeping the analysis simple (i.e. don't fill the pages with

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

technical statistical jargon and complex three-dimensional, dual axes graphs). You can always go back and do deeper or more sophisticated analysis later, after more burning questions have been discovered.

Data must be realistic.

Trade off the degree to which your data is relevant, representative, reliable and readable with the level of resources you will need to invest to make it so. Make sure the value you get from using your data is greater than the effort you invested in getting it.

Beware of the temptation to invest in sophisticated automatic data capture systems (such as bar-coding and voice recognition software). If you haven't got a simple manual system working well first, then these systems are likely to cost you much, much more than the savings they appear to promise.

Many surveys turn out to be a waste of time simply because they were not organised and conducted realistically. A survey is realistic if the resources you consumed to collect, analyse, present and use the data are significantly outweighed by the value you got in applying the data.

For example, there is no point choosing a sample size of 500 employees to know with 99% confidence that the percentage of satisfied employees is between 78.1% and 78.3% when you can know with 95% confidence that the percentage is between 77% and 79% with a sample of 100. The second example gives you very acceptable reliability for less than half the cost.

While this discussion has focused on surveys, the 5 Rs of data integrity apply in a similar way to any other form of data. The most important thing to do if you want to maintain the integrity of your performance measures is to clearly define those measures, and the level of integrity that they need. And it won't ever be 100%.

www.staceybarr.com | info@staceybarr.com | +61 1800 883 458

STACEYBARR

So You Don't Trust Your Data...

DATA INTEGRITY IS REALLY IMPORTANT IF YOU WANT TO REDUCE THE RISK OF YOUR DECISION MAKING.

About the author

Stacey Barr

Stacey Barr is a globally recognised specialist in organisational performance measurement. She discovered that the struggles with measuring business performance are, surprisingly, universal. The biggest include hard-to-measure goals, trivial or meaningless measures, and no buy-in from people to measure and improve what matters. The root cause is a set of bad habits that have become common practice.

Stacey created PuMP®, a deliberate performance measurement methodology to replace the bad habits with techniques that make measuring performance faster, easier, engaging, and meaningful.

Stacey is author of Practical Performance Measurement and Prove It!, publisher of the Measure Up blog, and her content appears on Harvard Business Review's website and in their acclaimed ManageMentor Program.

Discover more about Stacey and practical performance measurement at www.staceybarr.com.