

# BIG DATA INSIGHTS & PREDICTIVE ANALYTICS

OCTOBER 2018





## SSRS – RESEARCH REFINED, BIG DATA INSIGHTS REDEFINED

Big data means big opportunity. In a new landscape focused on insights – derived faster and cheaper than ever before – one might ask how survey research and social science, more broadly, can fit or add value.

## BIG DATA OPPORTUNITIES

There are big opportunities with big data, however there are also big challenges.

Big data are often not explicitly designed or generated for the insights they are asked to support. Much like secondary data analysis, investigators, researchers and businesses alike have to find ways to use or repurpose the data for gathering desired insights.

These data offer many new avenues of enquiry and can produce insights at faster speeds than we have typically seen in social and survey sciences. Big data can and is being used to answer many questions relating to the when, where, who and to some extent how things are happening in various industries and domains of interest.

**But the one key aspect of big data that seemingly still falls short is the one domain that social and survey sciences have been honing for decades – the “why”.**

We know that big data can tell us which products are returned, where the largest volume of returns are being made, as well as when they are made. However, these data are not providing insights into why a product is being returned – at least not in any systematic fashion.

So if full scope insights that seek to understand the entire picture of how, what, where, when and why are of interest, then the survey process can add value and round out the insights.

**As a premier social science and survey research firm, SSRS sees the path forward in this new modern era as a two-way street.**

---

On the one hand, data science approaches and methods can be leveraged and used to enhance survey sample designs and data collection protocols to make them even more efficient than they already are.

Using well-tuned machine learning models to better predict respondents prior to fielding surveys is one way we are already working to incorporate the power of data science within our world class survey and social science data collection methods.

Leveraging the power of a rich array of administrative data including census-related data, Twitter, and other commercially available sources, we are also plotting a course for more efficient sampling designs that can leverage a wider array of auxiliary variables to improve efficiency in stratification, sampling, and weighting.

But beyond these applications, survey and social science research can also add to the advances made in data science by creating additional variables to power predictive models, by understanding how to better tune machine learning methods using population-based estimates and for evaluating non-survey data sources within our advanced rubric of total survey error expanded to big data.

The path forward certainly has lanes for both survey and big data, but we see that successfully traversing this complex pathway will involve ingenuity around data triangulation. In our view, the new calculus for survey and social science research firms like SSRS revolves around the ability to triangulate different data sources, each of which add pixels to the picture.

It's clear that most pictures, to be fully appreciated, should have a full frame that can showcase the pixels in all their splendor – complete and comprehensible. And so it is with insights.

**We believe that comprehensive, full scope insights can be derived by agile triangulation of multiple sources of information including big data and survey data.**

However, it isn't just machines or methods that create these insights automatically.

It is said that putting garbage in often results in getting garbage out. By leveraging our survey expertise with subject matter experts and new expertise in data science and machine learning methods, we at SSRS are well poised to travel along this new path.

**SSRS is research refined, and we are using an agile, thoughtful, efficient, and rigorous data triangulation approach to redefine how insights are derived.**



## SSRS BIG DATA, PREDICTIVE ANALYTICS & DATA SCIENCE

Survey data provides the yin to big data's yang. It has a number of features in marked contrast to much Big Data.

Big Data is often transactional or behavioral. It can tell you that a person bought a ticket to an event, visited the hospital for a malady, or clicked on a website. But it cannot clearly tell you the motivations for why such behavior occurred in most cases, nor then predict who else might enact the same behavior.

**Big Data is a target of opportunity: It is largely gathered in the wild for purposes unrelated to a given research question.**

---

Survey research can probe persons specifically regarding that research question. Why did that person buy the ticket? Why did that person have to go to the hospital? Survey data can answer these questions by developing hypotheses on the whys and asking questions to test those hypotheses.

So confirmed, such survey data can then be applied to the Big Data to predict who else will buy tickets or be susceptible to a certain malady.

The turn to data science is largely an explosion of predictive analytics. Machine learning opens up new possibilities for the researcher in two important ways.

First, by allowing the machine to do the learning, surprising discoveries can occur that might be missed in strict hypothesis testing.

Second, a large range of data, Big Data, can be explored, meaning, it creates huge leaps in the ability to work with broad data. Which 30 of 300 indicators best predict whether someone will purchase a product? And which variables interact with others in powerful ways? Machine learning offers great new opportunities to be able to analyze such data.

**Survey research and Big Data, married together, offer insights that neither can provide alone.**

---

Big Data will often be deep but not wide; survey data wide but not deep. Linking both data together allow for models using wide data that are then scored to deep data. Most importantly, as noted earlier, Big Data is typically data of opportunity.

Survey research is purposeful. As such, the marrying of both allows researchers to leverage Big Data with research questions that survey data is built specifically to answer, again then allowing one to model such findings to a deep population.

The possibilities of this combination are only limited by the ability to ask survey questions; and to ask them to a sufficient number of people who are linkable to the Big Data.

The linkage can be direct, that is by linking person "a" in the survey data to the same person "a" in the Big Data, or by modelling person types in the survey data, and then scoring the same person types in Big Data.

**SSRS is known for its methodological prowess, deep client service, and high quality research.**

---

Those same principles serve SSRS well in a Big Data and data science environment. We have provided cutting edge insights to large customer databases by applying surgical survey research to understand the "why", and then modelling those results to the full

databases, allowing our clients the ability to classify database members (customers, users, etc.) into highly useful classes of behavior.

We model purchase behavior that helps customers understand who is "in the pocket," who is unpersuadable and should be avoided as to not expend limited capital, and who is in the middle "goldilocks zone", ready for the right persuasion toward customer status.

We model message susceptibility that identifies which people the message will resonate with most. We model likelihood to disconnect, so that you can develop strategies to avoid such behavior in your customers. We uncover the whys in terms of who among your customers are avid fans of your product, so that those who are avid can become rabid fans, and those who are casual fans can become avid fans. And we model highly specified behavior, so that specific outreach can be made to maximize product sales within specific segments.

All this is possible and more, with the combination of Big Data and survey data, and the application of predictive analytics to those married datasets.

SSRS is here to help take your insights to the next level.

## ABOUT THE AUTHOR

**David Dutwin, Ph.D.**

**SSRS EVP & Chief Methodologist**



David Dutwin, Ph.D., is primarily responsible for sampling designs, project management, executive oversight, weighting and statistical estimation. He is an active member of the survey research community, having served in the American Association for Public Opinion Research as a member and a chair of special task forces, a member of the Standards, Communications, and Heritage Committees; teaching multiple short courses and webinars; as the Student Paper winner of 2002; and as the 2016 Conference Chair. He was elected to the AAPOR Executive Council in 2017 and serves as the 2017 Vice President/2018 President. David is a Senior Fellow with the Program for Opinion Research and Election Studies at the University of Pennsylvania.

He holds a Masters of Communication from the University of Washington and his doctorate in Communication and Public Opinion from the Annenberg School for Communication at the University of Pennsylvania. David attained his Bachelors in Political Science and Communication from the University of Pittsburgh

He has taught Research Methods, Rhetorical Theory, Media Effects and other courses as an Adjunct Professor at West Chester University and the University of Pennsylvania for over a decade. David is also a Research Scholar at the Institute for Jewish and Community Research. His publications are wide-ranging, including a 2008 book on media effects and parenting; methodology articles for Survey Practice, the MRA magazine Alert!, and other publications; and a range of client reports, most recently on Hispanic acceptance of LGBT, which he presented to a Congressional briefing in 2012.

## CONTACT SSRS FOR MORE INFORMATION

[info@ssrs.com](mailto:info@ssrs.com) | 484.840.4300 | [@ssrs\\_research](https://www.ssrs.com)