

Deloitte Review

Issue 21 | July 2017

Complimentary article reprint



Predictably inaccurate

The prevalence and perils of
bad big data

By John Lucker, Susan K. Hogan, and
Trevor Bischoff

Illustration by Jon Krause

Deloitte.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.

Deloitte provides audit, tax, consulting, and financial advisory services to public and private clients spanning multiple industries. With a globally connected network of member firms in more than 150 countries and territories, Deloitte brings world-class capabilities and high-quality service to clients, delivering the insights they need to address their most complex business challenges. Deloitte's more than 200,000 professionals are committed to becoming the standard of excellence.

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, the "Deloitte Network") is, by means of this communication, rendering professional advice or services. No entity in the Deloitte network shall be responsible for any loss whatsoever sustained by any person who relies on this communication.

Copyright © 2017. Deloitte Development LLC. All rights reserved.



Predictably *in*accurate

The prevalence and perils of bad big data

By John Lucker, Susan K. Hogan, and Trevor Bischoff

Illustration by Jon Krause

“We’re not that much smarter than we used to be, even though we have much more information—and that means the real skill now is learning how to pick out the useful information from all this noise.”

—Nate Silver¹

IS OUR LOVE AFFAIR WITH BIG DATA LEADING US ASTRAY?

SOCIETY and businesses have fallen in love with big data. We can’t get enough: The more we collect, the more we want. Some companies hoard data, unsure of its value or unclear if or when it will be useful to them but, all the while, reticent to delete or not capture it for fear of missing out on potential future value. Stoking this appetite is the sheer growth in the volume, velocity, and variety of the data.

Most of all, many business leaders see high potential in a fourth V: *value*. Given our ability to access and (potentially) understand every move our current and potential customers make, coupled with

access to their demographic, biographic, and psychographic data, it seems logical that we should be able to form a more intimate, meaningful relationship with them. Every data point should move the business at least one step closer to the customer.

Yet despite all the digital breadcrumbs, it turns out that marketers might know less about individual consumers than they think. The numbers don't lie—or do they? What if much of this data is less accurate than we expect it to be?

Perils ranging from minor embarrassments to complete customer alienation may await businesses that increasingly depend on big data to guide business decisions and pursue micro-segmentation and micro-targeting marketing strategies. Specifically, overconfidence in the accuracy of both original and purchased data can lead to a false sense of security that can compromise these efforts to such an extent that it undermines the overall strategy.

This article explores the potential adverse consequences of our current love affair with big data. Evidence from our prior² and current primary research, supported by secondary research, highlights the potential prevalence and types of inaccurate data from US-based data brokers, as well as the factors that might be causing these errors. The good news is that strategies and guardrails exist to help businesses improve the accuracy of their data sets as well as decrease the risks associated with overreliance on big data in general.

PERSONAL DATA THAT'S BOTH INCOMPLETE AND INACCURATE

"It's pretty scary how wrong data collected about you can be—especially if people make important decisions based on this incorrect information. This becomes more frightening as more and more decisions become information-based."

—Survey respondent

TO better gauge the degree and types of big data inaccuracies and consumer willingness to help correct any inaccuracies, we conducted a survey to test how accurate commercial data-broker data is likely to be—data upon which many firms rely for marketing, research and development, product management, and numerous other activities. (See the sidebar “Survey methodology” for details.) Some of the key findings:³

- More than two-thirds of survey respondents stated that the third-party data about them was only 0 to 50 percent correct as a whole. One-third of respondents perceived the information to be 0 to 25 percent correct.
- Whether individuals were born in the United States tended to determine whether they were able to locate their data within the data broker's portal. Of those not born in the United States, 33 percent could not locate their data; conversely, of those born in the United States, only 5 percent had missing information. Further, no respondents born outside the United States and resid-

ing in the country for less than three years could locate their data.

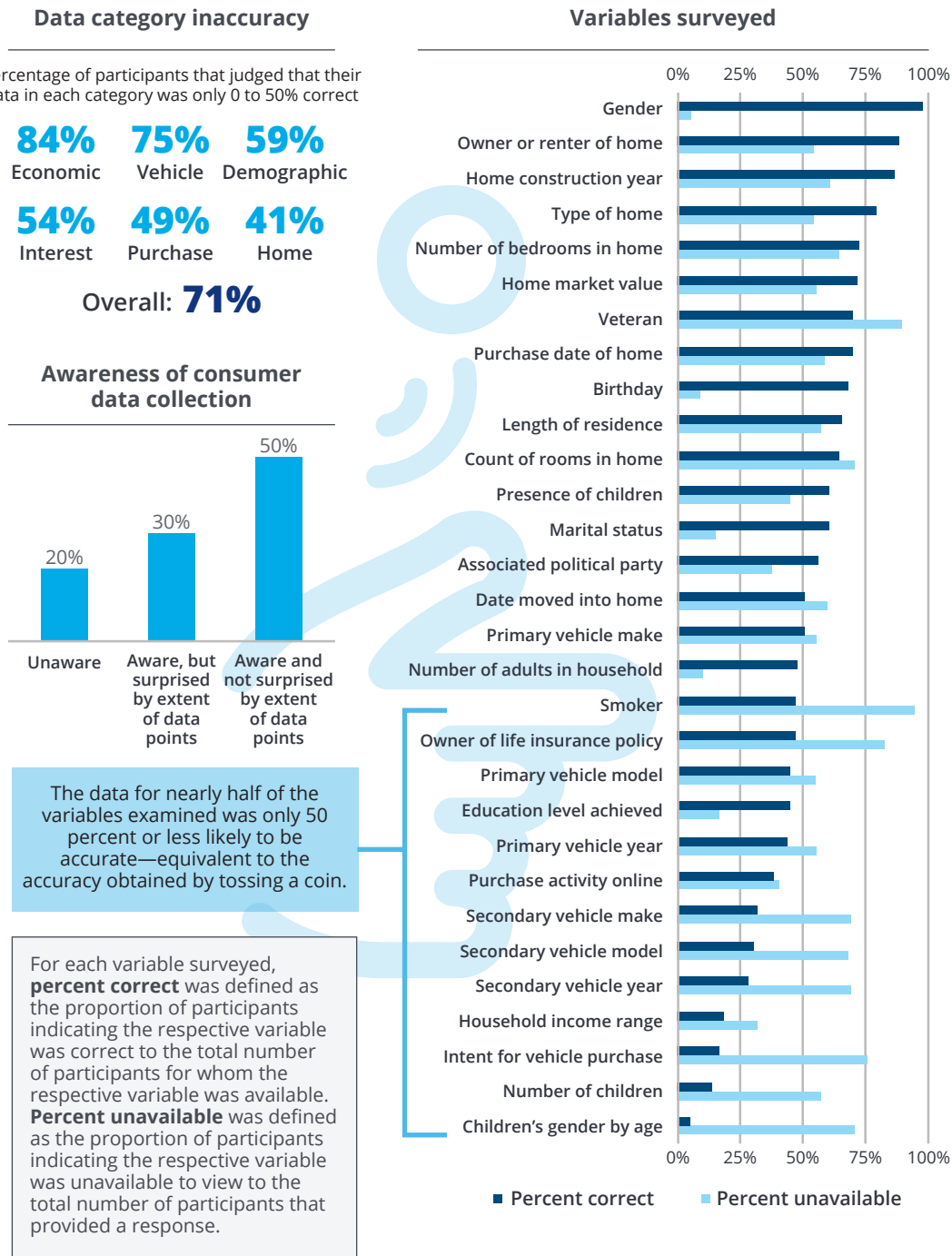
- The type of data on individuals that was most available was demographic information; the least available was home data. However, even if demographic information was available, it was not all that accurate and was often incomplete, with 59 percent of respondents judging their demographic data to be only 0 to 50 percent correct. Even seemingly easily available data types (such as date of birth, marital status, and number of adults in the household) had wide variances in accuracy.
- Nearly 44 percent of respondents said the information about their vehicles was 0 percent correct, while 75 percent said the vehicle data was 0 to 50 percent correct. In contrast to auto data, home data was considered more accurate, with only 41 percent of respondents judging their data to be 0 to 50 percent accurate.
- Only 42 percent of participants said that their listed online purchase activity was correct. Similarly, less than one-fourth of participants felt that the information on their online and offline spending and the data on their purchase categories were more than 50 percent correct.
- While half of the respondents were aware that this type of information about them existed among data providers, the remain-

SURVEY METHODOLOGY

Our survey asked 107 Deloitte US professionals to privately and anonymously review data made available by a leading consumer data broker, a broker with a publicly available, web-based portal that presents users with a variety of personal and household data. Respondents, all between 22 and 67 years of age, completed the rapid-response, 87-question survey between January 12–March 31, 2017.

Respondents viewed their third-party data profiles along a number of specific variables (such as gender, marital status, and political affiliation), grouped into six categories (economic, vehicle, demographic, interest, purchase, and home). To calculate the “percent correct” for each individual variable, we took the number of participants who indicated that the third-party data point for that variable was correct, and divided it by the total number of participants for whom third-party data were available for that variable. To determine respondents’ views of the accuracy of the data for each *category*, we asked them to indicate whether they felt the category data was 0 percent, 25 percent, 50 percent, 75 percent, or 100 percent accurate.

Figure 1. Reported accuracy of third-party consumer data from our respondents



ing half were surprised or completely unaware of the scale and breadth of the data being gathered.

Figure 1 outlines other inaccuracies or omissions related to date of birth, education level, number of children, political affiliation, and household income. Clearly, all of these types of data are potentially important to marketers as they target different consumer segments.

Can we count on individuals to correct their own data?

“While I wasn’t surprised by the extent of the data collected, it was interesting to see it. I was actually surprised at how little data there was about me (I am an avid online shopper), and how incomplete the ‘cyber me’ picture is. I’m not complaining about it, though.”

—Survey respondent

Survey respondents were provided with the opportunity to elaborate on why they thought their data might be wrong or incomplete. Most commonly, the available information was outdated—especially vehicle data. Many others saw the data as characterizing their parents or other household members (spouses or children) rather than themselves. The most-mentioned feeling among respondents was surprise—not at the amount of correct data available, but rather that the information was so limited, of poor quality, and inconsistent. In essence, for many respondents, the data seemed, as aptly put by one respondent, “stale.”

“There was lots of information that didn’t exist about me. And of the data that did exist, much seemed inconsistent with other data.”

—Survey respondent

Interestingly, even after being offered the opportunity to edit their data via the data broker’s online portal, few respondents chose to do so. While approximately two-thirds of respondents reported that at least half of their information was inaccurate, only 37 percent opted to edit their data.

The most common best reason for the decision to edit (given by 31 percent of respondents who chose to edit) was to improve the information’s accuracy. The second most common response was a decision to edit only what seemed relevant (provided by 17 percent of respondents opting to edit). Another 11 percent of respondents who opted to edit cited privacy and nervousness about their data being “out there.” Other respondents noted the desire to reduce or avoid targeted messaging and political mailings, as well as the hope of improving their credit rating (even though, presumably unknown to them, this type of marketing data has no direct connection to how credit scores are derived). The most commonly edited categories were demographic data and political party data.

Why did so many respondents elect not to edit their data? Most often, people cited privacy concerns. Other reasons included no perceived

value in editing and ambiguity regarding how third parties might use the data. Table 1 gives an overview of the most common reasons for the decision to edit or not.

“I’m skeptical and cautious about what could be done with this data. Even assuming the best of intentions and integrity by people who might consume this data, I cannot imagine a scenario that would also be in my or my family’s best interest. I would actually prefer less personal information about me to exist publicly. So, obscure, inaccurate, or unreliable data is what I consider to be the next best thing.”

—Survey respondent

THE PERILS OF RELYING ON BAD DATA

OUR survey findings suggest that the data that brokers sell not only has serious accuracy problems, but may be less current or complete than data buyers expect or need. Given that a major US marketing data broker hosts the publicly available portal used for our survey, these findings can be considered a credible representation of the entire US marketing data available from numerous data brokers. The impacts of inaccurate or incomplete data are many, ranging from *missed opportunities* to just plain *misses*.

Table 1. Common reasons driving decisions to edit or not to edit data

Why did you edit your data?	Why didn't you edit your data?
<ul style="list-style-type: none"> • To make data more accurate/better • Corrected only where I perceived that it was valuable/worth the effort • Privacy/nervous that this data is even out there • To reduce/avoid targeted ads/offers 	<ul style="list-style-type: none"> • Privacy • No perceived value/not worth the time and energy • Not interested/don't care what data they have on me • Cautious/unclear how the information will be used • Lack of time to edit • Not my job to do their work/fix errors • Against targeted marketing

Source: Deloitte analysis.

Deloitte University Press | dupress.deloitte.com

Missed opportunity 1: Underestimating customer worth and not capitalizing on the power of habit

"I wish I spent only that much. My purchasing data seems significantly understated from what I know I spend in the categories indicated."

—Survey respondent

Understanding the spending behavior and power of current and potential customers is very important to firms. Many marketers extrapolate this information based on three key categories: current income, modeled net worth, and prior purchasing behavior. Consumers are creatures of habit—our past spending behavior is one of the best indicators for marketers to determine not only how much we will spend in the future, but what types of items we are likely to purchase. This can guide predictions on how much revenue a company can expect to see in the coming year, as well as any cross-selling or up-selling efforts.⁴ Given this information's importance to marketers, and the incredible number of digital breadcrumbs that consumers leave behind, we were surprised to find such a high level of inaccuracy. More often than not, respondents indicated that the household income data provided by the broker was incorrect, with purchasing data often underestimated, suggesting that marketers relying on this information to guide their targeting efforts may be leaving potential revenue on the table.

Missed opportunity 2: Decreased customer loyalty and revenue

"[The data] stated that I own a property that is actually owned by my parents, and at the same time, it failed to list the property that I currently do own."

—Survey respondent

Another area of significant inaccuracy was home residence and vehicle ownership, which was quite surprising given the readily available public records for each. As stated previously, home data was more accurate than auto data, but still considerably inaccurate overall. Respondents suggested that the data in these two categories was often outdated—potentially by five to ten years.

One of the highest-expenditure periods in an individual's life is when she makes a household move. Not only are these moves expensive—households incur significant ancillary spending as well, even with local moves. When moving from one geography to another with a different climate, the consumer often starts from scratch in numerous product categories (new wardrobe, home furnishings, outdoor equipment, and so on). A marketer wouldn't want to miss this transitional moment, in which consumers spend more money than they typically would as well as form new behaviors—including purchasing routines and loyalties. Without a timely and relatively accurate picture of a consumer's residence changes, the marketer could miss out on influencing momentary pur-

Figure 2. What do people think about their own big data profiles? A sampling of comments from our respondents



Deloitte University Press | dupress.deloitte.com

chases, subsequent add-on purchases, and, potentially, building long-run customer loyalty.

Corroborating our findings, a third-party data quality study found that 92 percent of financial institutions rely on faulty information to better understand their members, a rate likely attributable to human errors and flaws in the way multiple data sources were combined. Fully 80 percent of credit unions believe the inaccuracies have affected their bottom line, causing an

average 13 percent hit on revenue. Additionally, 70 percent of financial institutions blame poor data quality for ongoing problems with their loyalty efforts.⁵

Miss 1: Moving the customer relationship along too fast

"I'm annoyed that nothing is private anymore. I rarely use advertisements for purchasing decisions anyway, and I wish I could stop receiving them altogether."

—Survey respondent

It should go without saying that micro-targeted messaging is full of pitfalls—regardless of the accuracy of the data on which it is based. Take, for example, the father who learned about his daughter’s pregnancy through retailer offerings that came in the mail after the retailer detected purchasing behavior correlated with pregnancy.⁶ While evidence suggests that consumers are becoming more receptive to personalized marketing, marketers still need to be thoughtful and tread lightly in this area.⁷ This word of warning is consistent with recent research identifying similarities between interpersonal relationship development and business and customer relationships,⁸ as well as existing theories regarding healthy relationship development. Particularly, self-disclosure of personal information is meant to follow a reciprocal and progressive course, with initial mutual sharing of surface-level personal information over time evolving to a more intimate level of exchange.⁹ Too much, too soon from either party can come across as invasive and creepy—and disrupt the relationship that has developed so far. This means that demonstrating a ballpark knowledge of your customer early on may be more beneficial than demonstrating an intimate or precise knowledge. Recent research has corroborated this idea, suggesting that semi-tailored or customized advertising can lead to a 5 percent increase in intent to purchase. However, advertising that gets too specific, by seeming to zero in on one individual as opposed to a general demograph-

ic group profile, may be viewed as invasive and a little too close for comfort. This latter situation can lead to a 5 percent *decrease* in intent to purchase.¹⁰

Miss 2: Delivering the wrong or inappropriate micro-targeted message

*“Some of the misses were really bad, like my political party and my interest in tobacco!”
—Survey respondent*

Probably worse than getting too close is getting it wrong. When a marketer tries to make a personal connection through messaging using wrong or inappropriate information, the effects can range from humorous—such as a twentysomething receiving AARP membership invitations¹¹—to sad. The latter was the case with a recently mailed discount offer that, while sent to a live person, included an (accurate) reference to not only a recently deceased family member but the way this person died—embedded into the recipient’s mailing address. The firm that had given the offer, which didn’t believe it could have sent out this mailing until receiving the physical proof, claimed this blunder was the result of a rented mailing list from a third-party provider.¹² While reported cases such as this last example are rare, basing a personalized message around wrong or inappropriate information, and subsequently delivering the wrong micro-targeted message to customers, can not only diminish the effect of marketing efforts, but do more damage than good. This adverse reaction is often referred

to as a boomerang effect: causing a customer to move from a neutral, nonexistent, or positive attitude toward the company to a negative one.¹³

Miss 3: Assessing risk inaccurately

Both private and public health care institutions often create and rely on big data models to understand their patients' future needs and potential life spans. Such risk models, however, go beyond managing an insurer's bottom line by helping identify high-risk clients.¹⁴ Inaccurate data can prompt inaccurate assessments such as determining financial risks,¹⁵ life expectancies,¹⁶ and medical care needs, which can lead to inappropriate insurance payments at best.¹⁷ At worst, if public health groups that use these risk models to guide strategic decisions around global public health initiatives miss the mark, it can contribute to deaths. These deaths could be due to misidentification of vulnerable or at-risk populations, which could be avoided if the right treatments were made available to them.¹⁸

Miss 4: Predicting inaccurate outcomes

While most of us have learned to cut weather forecasters some slack, we are fixated on the many "scientific" and "statistically significant" crystal balls: models used to predict the outcomes of our elections,¹⁹ football games, and horse races. Yet models meant to determine precautions to be taken have often been off the mark. For example, in 2013, a search engine-

based flu-tracking model forecast an increase in influenza-related doctor visits that was more than double what the Centers for Disease Control and Prevention (CDC) predicted.²⁰ While the CDC based its predictions on various laboratory surveillance reports collected from across the United States, the culprit behind the social media tracking tool's wildly different result was what some researchers have called "big data hubris": the mistake of assuming that big data can substitute for, rather than supplement, traditional methods of data collection and analysis.²¹

HOW DID THE DATA GET SO BAD?

UNFORTUNATELY, our primary research findings are not unique but, rather, a glimpse into the general state of affairs: Big data is often inaccurate,²² and companies relying on inaccurate big data can suffer significant consequences. Since we reviewed only the fields available to us, it's important to note that inaccuracies almost certainly extend beyond the fields and attributes highlighted in this article, especially the less common or more esoteric fields, such as whether an individual is a veteran.

So how does this information wind up so far off the mark? There are many possible causes, such as human error, collection or modeling errors, and even malicious behavior. To make matters worse, a data set is often victim to more than one type of error. Some examples of how errors can arise:

Big data is often inaccurate, and companies relying on inaccurate big data can suffer significant consequences.

- Outdated or incomplete information may persist due to the cost and/or effort of obtaining up-to-date information
- An organization that uses multiple data sources may incorrectly interweave data sets and/or be unaware of causal relationships between data points and lack proper data governance mechanisms to identify these inconsistencies
- An organization may fall prey to data collection errors:
 - Using biased sample populations (subject to sampling biases based on convenience, self-selection, and/or opt-out options, for instance)²³
 - Asking leading or evaluative questions that increase the likelihood of demand effects (for example, respondents providing what they believe to be the “desired” or socially acceptable answer versus their true opinion, feeling, belief, or behavior)
 - Collecting data in suboptimal settings that can also lead to demand effects (for example, exit polls, public surveys, or any mechanism or environment in which respondents do not feel their responses will be truly anonymous)
 - Relying on self-reported data versus observed (actual) behaviors²⁴
- Data analysis errors may lead to inaccuracies due to:
 - Incorrect inferences about consumers’ interests (for example, inferring that the purchase of a hang-gliding magazine suggests a risky lifestyle when the purchaser’s true motive is an interest in photography)²⁵
 - Incorrect models (for instance, incorrect assumptions, proxies, or presuming a causal relationship where none exists)
- Malicious parties may corrupt data (for example, cybercrime activity that alters data and documents)²⁶

Understanding the causes of these errors is a first step to avoiding and rectifying them. The next section explores the next steps companies can take along the path to utilizing big data in the right way.

A BIG DATA PLAYBOOK: PRESCRIPTIONS FOR SUCCESS

THERE is growing recognition that much big data is built on inaccurate information, driving incorrect, suboptimal, or disadvantageous actions. Some initial efforts are under way to put in place regulations around big data governance and management.²⁷ Regulatory agencies, such as the Federal Trade Commission and the National Association of Insurance Commissioners, are beginning to consider more oversight on data brokers as well as how models utilizing their data are used. However, savvy firms already engaged in big data should not wait for agencies to act, especially given the uncertainty around how effective or restrictive any eventual regulations will be. Based on our market experience and observations, here are some guidelines, advice, and remedies to consider to help you avoid shooting yourself in the foot when utilizing big data.

Increase the likelihood that more of your big data will be accurate

*"If they were more clever, they could cross-reference the home data with household income data to find major discrepancies."
—Survey respondent*

Ask and expect more from big data brokers. Perhaps our expectations for big data are too high—but it's possible that we are asking too little of data brokers, especially given the study results we describe here. The role of data brokers has evolved over time. Traditionally, firms looked to data brokers to provide mailing lists and labels for prospective customers and, perhaps, to manage mailing lists and track current customers' purchasing behavior. However, the information that brokers provide now plays a much more integral role in our strategies, digital interactions, and analytic models. Consequently, we should be asking for more accountability, transparency, and continuous dialogue with these organizations. (See the sidebar, "What to ask your data brokers.")

Know the data sources. While you certainly want to understand where your own data come from, knowing the source and lineage is particularly important for information you source through data brokers. However, our research suggests data brokers fall on a spectrum when it comes to revealing their sources. Not all brokers organically generate the data they sell; rather, many license information to each other, as different brokers cater to various data use cases and business niches.

Put steps in place to verify that the brokers from which you source have adequate control over their data's accuracy, including control over and transparency regarding their data sources. Understand the surveillance proce-

WHAT TO ASK YOUR DATA BROKERS

Demand transparency regarding:

- Data source(s): the lineage of the data fields and values, timing of maintenance, update processes
- Data collection, validation, and correction methods
- Any relationships and interdependencies—for instance, interrelatedness between data sources and model inputs
- Model inputs and assumptions

Ensure ongoing communications with data sources in order to be kept abreast of any:

- Inaccuracies found in existing data sets
- Changes to models and/or assumptions and the rationale for such changes, as well as transparency to model logic and metadata
- Changes to categories and the rationale for such changes

Verify the appropriateness of the manner in which you are using their data:

- Explain to the broker how you are using data, and verify that their information is appropriate and sufficiently accurate for your context

Consider specifying accuracy and performance standards in your data broker contracts.

dures they have in place with these sources to track changes, measure accuracy, and ensure consistency. Develop and maintain processes to be notified of inaccuracies in the data, and understand how often information is validated or updated. Consider the significance of a five-year age difference: 20-year-olds are buying different products than those aged 25, just as those who are 25 are at a different stage in life than 30-year-olds.

Explore the data yourself. Before you use any big data (especially externally sourced) to guide your decisions and marketing strategies, do an exploratory data analysis yourself. If possible, test a sample for inaccuracies or inconsistencies against data fields you already have or can validate. On your own, consider digging into the data and doing validity checks, exploratory analysis, and data mining against individual and industry information. Does what

you are seeing make sense? For example, one of the authors of this very article was labeled as having an old-fashioned dial-up Internet connection rather than the actual broadband connection.

Alternatively, hire an expert to look at this data. Also, realize that internally gathered information often relies on a combination of sources—which could be external or outdated—and is also prone to human error, so the same verification tests should be performed here as well. A proper data governance framework can go a long way in helping to ensure your information is accurate, timely, and valuable.

Consider big data to be one more tool in the toolkit, not a replacement toolkit

Keep expectations for big data in check.

It is often the case that big data might be directionally correct but still inaccurate at an individual level. The good news for firms and marketers is that big data analytics built on such “semi-accurate” information can provide predictive power overall. However, it is a mistake to expect individual micro-predictions to carry the same level of accuracy.²⁸

Use and draw conclusions from big data judiciously. Big data is a great tool for marketers, but it should be thought of as a tool in the decision-making and marketing toolkit, not a replacement for the already existing toolkit. Consequently, don't rely too heavily on a limited number of data points, especially if accuracy is a potential peril. If you decide to do

any micro-messaging, consider limiting its geographies and scope to avoid some of the perils we discussed earlier. Additionally, soliciting customer feedback on the data not only improves the prospect of more accurate data—it increases transparency within the relationship. However, as our findings suggest, you can't count on your customers to fill in the gaps adequately and accurately.

Complement big data with other decision-making tools.

While big data is and will remain a powerful tool for firms and marketers when used appropriately, we've already explored the dangers of overreliance on it—which could also result in marketers losing faith in their own experience and intuition to help guide decisions.²⁹ Therefore, executives should complement the decisions derived from big data with their own insights based on experience and other research methods and sources (such as small-sample qualitative research). Regardless of the data quality, a good rule of thumb is to not over-rely on the data and outsource too many decisions.³⁰

Continually connect with customers

Be nimble and responsive. Continually assess data sources and appropriateness of methodologies, models, and assumptions; frequently revisit and assess questions and category fit with changing target demographics and categories. Also, measure how successful target marketing efforts have been since incorporating insights from big data. Beyond quantitative

or objective measures, create feedback opportunities within your micro-targeting. After collecting feedback, spend time reviewing, incorporating, and adjusting your strategies based on this feedback. When appropriate, respond directly to those providing feedback—recent research suggests this may not only increase the likelihood of additional feedback, but also make the customer feel more valued and encourage an ongoing dialogue.³¹

Reward customers for correcting their data. While our study suggests that consumers are unlikely to correct information provided by a big data source, it's worth exploring their willingness to take corrective action for their own data if the request comes from a firm with which they have a relationship—and for which they see more direct value from such

an action. Additionally, in an effort to thank customers for not only their patronage but for updating personal information, firms can offer incentives for their corrective efforts. The benefits could be many: accurate customer data; an active, direct line of communication; and, ultimately, a deeper connection with customers.

Regardless of our current infatuation with big data, we must remember that data should never take center stage at the expense of the customer. Firms that understand big data's limitations (and advantages) can add it to their marketing and analytical arsenal, aiming to foster and preserve customer relationships and the trust that they work so hard to develop and maintain. ●

***John Lucker** is an advisory principal with Deloitte & Touche LLP, Global Advanced Analytics & Modeling market leader, and a US leader of Deloitte Analytics.*

***Susan K. Hogan** is a market insights manager with Deloitte Services LP.*

***Trevor Bischoff** is a senior consultant within the financial services sector at Deloitte & Touche LLP.*

*The authors would like to thank **Ashley Daily**, **Adam Hirsch**, and **Michael Greene**, who served as inspiration and fueled our enthusiasm for this current research. We also thank **Negina Rood**, **Junko Kaji**, **Aditi Rao**, and **Kevin Weier** for their contributions.*

Endnotes

1. NPR, "'Signal' and 'noise': prediction as art and science," October 10, 2012, <https://n.pr/UPXRS4>.
2. John Lucker, Ashley Daily, Adam Hirsch, and Michael Greene, "Predictably inaccurate: Big data brokers," *LinkedIn Pulse*, November 18, 2014, www.linkedin.com/pulse/20141118145642-24928192-predictably-inaccurate-big-data-brokers.
3. All percentages relating to respondents were calculated on a base of the number of respondents for whom third-party data was actually available in the categories of interest; the calculations excluded respondents for whom the third-party data was unavailable.
4. Lucker et al., "Predictably inaccurate."
5. Thomas Schutz, "Want better analysis? Consider the data," *Credit Union Journal*, May 16, 2014, www.cujournal.com/opinion/want-better-analysis-consider-the-data.
6. Morgan Hochheiser, "The truth behind data collection and analysis," *John Marshall Journal of Information Technology and Privacy Law* 33, no. 1 (2015): pp. 32–55, <http://repository.jmls.edu/jitpl/vol32/iss1/3>.
7. Victoria Petrock, "Are consumers warming to personalized marketing services?," eMarketer brief, July 26, 2016, www.emarketer.com/Brief/Consumers-Warming-Personalized-Marketing-Services/5500941.
8. Susan K. Hogan, Rod Sides, and Stacy Kemp, "Today's relationship dance: What can digital dating teach us about long-term customer loyalty?," *Deloitte Review* 20, January 23, 2017, <https://dupress.deloitte.com/dup-us-en/deloitte-review/issue-20/behavioral-insights-building-long-term-customer-loyalty.html>.
9. Irwin Altman and Dalmas A. Taylor, *Social Penetration: The Development of Interpersonal Relationships* (New York: Holt, Rinehart and Winston, 1973).
10. Ithaca College, "Online creep: Targeted ads may have opposite effect of marketers' intent," *Science Daily*, April 8, 2015, www.sciencedaily.com/releases/2015/04/150408171201.htm.
11. Joshua Lederman, Twitter post, September 29, 2016, 1:48 p.m., <https://twitter.com/joshledermanap/status/781596504351907840>.
12. Natasha Singer, "Oops! Health insurer exposes member data," *New York Times*, November 10, 2014, <https://nyti.ms/2qycdOU>.
13. Sharon S. Brehm and Jack Williams Brehm, *Psychological Reactance: A Theory of Freedom and Control* (New York: Academic Press, 1981).
14. Leslie Scism, "Life insurers draw on data, not blood," January 12, 2017, *Wall Street Journal*, www.wsj.com/articles/the-latest-gamble-in-life-insurance-sell-it-online-1484217026.
15. Ankur Aggarwal et al., "Model risk—daring to open up the black box," *British Actuarial Journal* 21(2), December 22, 2015, http://journals.cambridge.org/abstract_S1357321715000276.
16. Scism, "Life insurers draw on data, not blood."
17. Rachel S. Karas, "Stakeholders urge CMS to factor Rx drugs in risk assessment pay, question other CMS ideas," *InsideHealthPolicy's Daily Brief*, April 28, 2016.
18. Jane Freemantle et al., "Indigenous mortality (revealed): The invisible illuminated," *American Journal of Public Health* 105(4), April 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4358192/.
19. Jim Rutenberg, "A 'Dewey defeats Truman' lesson for the digital age," *New York Times*, November 9, 2016, <https://nyti.ms/2JL43lb>.
20. Vasileios Lampos, Andrew C. Miller, Steve Crossan, and Christian Stefansen, "Advances in nowcasting influenza-like illness rates using search query logs," *Scientific Reports* 5 (2015), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4522652/>.
21. Larry Greenemeier, "Why big data isn't necessarily better data," *Scientific American*, March 13, 2014.
22. Daniel A. McFarland and H. Richard McFarland, "Big data and the danger of being precisely inaccurate," *Big Data & Society*, July–December 2015, pp. 1–4; <http://journals.sagepub.com/doi/full/10.1177/2053951715602495>.
23. StopDataMining.me, "Opt out list," www.stopdatamining.me/opt-out-list/, accessed May 2, 2017.
24. McFarland and McFarland, "Big data and the danger of being precisely inaccurate."
25. Scism, "Life insurers draw on data, not blood."
26. Mark Ward, "How fake data could lead to failed crops and other woes," *BBC*, March 21, 2017, www.bbc.com/news/business-38254362.

27. Carten Cordell, "Transparency advocates pitch the next big thing in big data," *Federal Times*, March 23, 2017, www.federaltimes.com/articles/with-data-act-reports-looming-transparency-advocates-pitch-the-next-big-thing.
28. Lucker et al., "Predictably inaccurate."
29. Ulrich Hoffrage, "Overconfidence," in Rüdiger F. Pohl, editor, *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (New York: Psychology Press, 2004); Tore Håkonsson and Tim Carroll, "Is there a dark side of big data—point, counterpoint," *Journal of Organization Design* 5(5), July 12, 2016, <http://link.springer.com/article/10.1186/s41469-016-0007-5>.
30. Nigel Harvey, "Confidence in judgment," *Trends in Cognitive Sciences* 1(2), May 1997, pp. 78–82, [www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613\(97\)01014-0](http://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(97)01014-0).
31. Susan K. Hogan and Timothy Murphy, *Loving the one you're with*, Deloitte University Press, June 17, 2016, <https://dupress.deloitte.com/dup-us-en/focus/behavioral-economics/how-behavioral-factors-influence-customer-rewards-incentives.html>.

Deloitte.
University Press

 Follow @DU_Press #DeloitteReview

Subscribe to receive email updates at
dupress.deloitte.com

