10

DEBUNKING MYTHS AND URBAN LEGENDS ABOUT HOW TO IDENTIFY INFLUENTIAL OUTLIERS

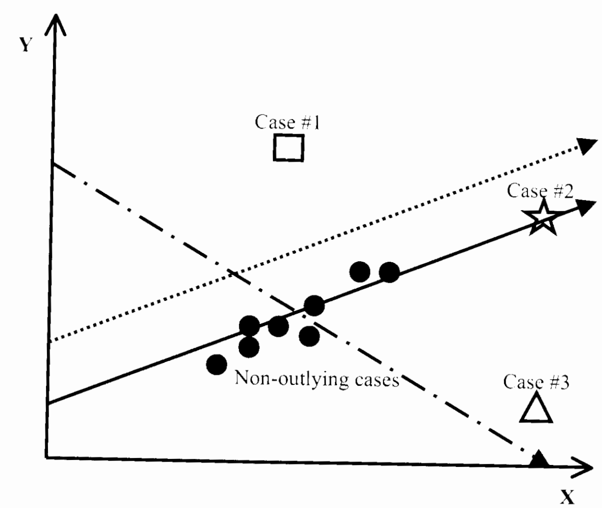
Herman Aguinis and Harry Joo

An outlier is an individual, team, firm, or any other unit that deviates markedly from others. *Influential outliers* are units that deviate markedly from the rest and, in addition, their presence has a disproportionate impact on substantive conclusions regarding relationships among variables. Due to their disproportionate impact on substantive conclusions, influential outliers constitute one of the most enduring and pervasive methodological challenges in both micro- (Orr, Sackett, & DuBois, 1991) and macro-level (Hitt, Harrison, Ireland, & Best, 1998) organizational science research.

There are many examples of substantive conclusions that have been changed based on how just a handful of influential outliers were identified in the same data set (Aguinis, Gottfredson, & Joo, 2013). For example, Hollenbeck, DeRue, and Mannor (2006) reanalyzed data collected by Peterson, Smith, Martorana, and Owens (2003), who investigated the relationships among CEO personality, team dynamics, and firm performance. Hollenbeck and colleagues (2006) showed that, of the 17 statistically significant correlations reported by Peterson and colleagues (2003), only one was actually significant for all 17 sensitivity analyses, in which each of the 17 individual data points (i.e., 17 CEOs) was removed one at a time. In other words, Hollenbeck and colleagues (2006) demonstrated that substantive conclusions regarding relationships among CEO personality, team dynamics, and firm performance changed almost completely depending on which cases were identified as outliers.

In spite of their pervasiveness and importance, there is confusion, misunderstanding, and a lack of clear guidelines on how to identify influential outliers. In particular, researchers frequently rely on three myths and urban legends (MULs) to identify such data points. These three MULs are rooted in the commonly invoked yet incorrect assumption that a data point, by virtue of being located far from others, *necessarily* has a large influence on substantive results (e.g., regression

coefficients, correlations). Although many researchers tend to assume that distance also means influence, this assumption often does not hold true because distance is a necessary but not sufficient condition for influence. To illustrate this point, consider Figure 10.1 (from Aguinis et al., 2013), which includes a scatter plot of a data set involving one predictor and one criterion. Regression analysis based on these data yields an R^2 of .73 when Cases #1, #2, and #3, which seem to be far from the rest, are excluded from the analysis. When Case #1, #2, or #3 is included one at a time, R^2 changes to .11, .95, or .17, respectively. So, the inclusion of each of these individual cases does change results regarding model fit in a substantive manner. Further, the inclusion of Case #1 or #3 reduces R^2 and also affects the model's parameter estimates (i.e., the intercept and/or slope). On the other hand, now consider Case #2, which is also far from the others—in terms of both the X and Y variable distributions. The inclusion of Case #2 in the analysis improves R^2 because of its location along the regression line. However, its inclusion or exclusion does not affect the intercept or slope parameter estimates. In short, although Case #2 is clearly far from other data points, it does not have influence on the



	Regression line	R^2	Slope	Intercept
With ● only	——→	.73	.83	5.34
With ● and □	⋯⋯→	.11	.83	9.13
With ● and ☆	——→	.95	.83	5.34
With ● and △	—·—·→	.17	-.22	30.34

FIGURE 10.1 Scatter plot illustrating that distance is not necessarily the same as influence in the context of regression

parameter estimates of this model, which illustrates that distance is not necessarily the same as influence.

The goal of our manuscript is to debunk three MULs about how to identify influential outliers. For each of these three MULs, we explain their nature, the kernels of truth behind them, and how the kernels of truth have been misapplied over time to form the MULs. Also, we illustrate these MULs using published articles with one important caveat: The practices we refer to are so pervasive that we could have illustrated them with dozens of examples. So we chose some illustrations with the purpose of making our points, but we do not wish to single these articles out as being particularly special in any way. In addition, after the discussion of each of the three MULs, we offer best-practice recommendations regarding how to identify influential outliers in the analytic contexts of multiple regression, structural equation modeling (SEM), multilevel modeling, meta-analysis, and time series analysis. We chose to address these particular data-analytic approaches because they are among the most popular and frequently used in organizational science research (Aguinis, Pierce, Bosco, & Muslin, 2009).

Three Pervasive Myths and Urban Legends about How to Identify Influential Outliers

We identified the three MULs through a content analysis of how authors of articles published in substantive organizational science journals identify influential outliers, as well as a review of recommendations on how to identify influential outliers offered in methodological sources. First, we content analyzed journal articles that mentioned the topic of outliers identified by Aguinis and colleagues (2013). The literature review focused on the following journals covering the years 1991 through 2010: *Academy of Management Journal*, *Journal of Applied Psychology*, *Personnel Psychology*, *Strategic Management Journal*, *Journal of Management*, and *Administrative Science Quarterly*. This process resulted in a total of 232 articles. Second, we reviewed the References list of each of these journal articles to locate and examine methodological sources upon which they may have relied regarding the issue of outliers. For example, we reviewed several textbooks that are typically used in training doctoral students in the organizational sciences (e.g., Cohen, Cohen, West, & Aiken, 2003; Tabachnick & Fidell, 2007).

As a result, we uncovered that the following are the three most common MULs about how to identify influential outliers: (1) Univariate cutoffs (e.g., top and bottom 1% of cases, cases that are more than 2 or 3 standard deviation [SD] units away from the mean) are sufficient for identifying influential outliers; (2) inspection of visual plots (e.g., histograms, scatter plots, residual plots) is sufficient for identifying influential outliers; and (3) absolute cutoffs based on multivariate test statistics (e.g., standardized residuals beyond ± 3 SD units, Cook's D values greater than 1) are appropriate for identifying influential outliers.

Table 10.1 includes a summary of the results of our content analysis, which revealed that 94 (i.e., 40.5%) of the 232 published substantive journal articles in

TABLE 10.1 Summary of Myths and Urban Legends about How to Identify Influential Outliers: Their Nature, Pervasiveness, Kernels of Truth, and How the Kernels of Truth Have Been Misapplied

Myth and Urban Legend (MUL)	*Pervasiveness of MUL (out of 138 articles that were transparent enough in reporting how influential outliers were identified)*	*Kernels of Truth Behind the MUL*	*How the Kernels of Truth Have Been Misapplied Over Time to Form the MUL*
Univariate cutoffs are sufficient for identifying influential outliers.	64.49% (89 articles)	• Univariate cutoffs help identify distant cases, which are often influential outliers. • Authoritative methodological sources describe univariate cutoffs in detail.	• Many authors have come to misunderstand distance as not only a necessary but also a sufficient condition for influence. • Many authors seem to have jumped to the incorrect conclusion that because an authoritative methodological source describes univariate cutoffs, the inference is that such cutoffs are recommended as sufficient means for identifying influential outliers.
Inspection of visual plots is sufficient for identifying influential outliers.	12.32% (17 articles)	• Superior alternative procedures for identifying influential outliers did not exist or were not available for practical use before the early 1980s. • Authoritative methodological sources give detailed descriptions of a number of visual techniques.	• Many authors have continued to use visual plots alone to identify influential outliers, even though improvements in statistical methods and computing technology no longer justify the exclusive use of visual plots on practical grounds. • Many authors seem to have jumped to the incorrect conclusion that an authoritative methodological source describing visual plots means that visual plots are sufficient means for identifying influential outliers.

(*Continued*)

TABLE 10.1 (Continued)

Myth and Urban Legend (MUL)	*Pervasiveness of MUL (out of 138 articles that were transparent enough in reporting how influential outliers were identified)*	*Kernels of Truth Behind the MUL*	*How the Kernels of Truth Have Been Misapplied Over Time to Form the MUL*
Absolute cutoffs based on multivariate test statistics are appropriate for identifying influential outliers.	8.70% (12 articles)	• Authoritative methodological sources seem to approve of the use of absolute cutoffs based on multivariate test statistics to identify influential outliers.	• Many authors have continued to use such absolute cutoffs to identify influential outliers, even though evidence has accumulated pointing to the superiority of research design–based cutoffs. Research design–based cutoffs are superior because they can vary from study to study depending on the characteristics of the particular research context such as sample size and number of predictors. Specifically, the "bar" for considering a case as an influential outlier is higher as sample size decreases and the number of predictors increases in a model.

Note: Of the 232 journal articles included in our review, 94 (i.e., 40.5%) were not sufficiently transparent in their reporting for us to determine precisely how influential outliers were identified. So these percentages are based on a total of 138 articles that provided sufficient information. Out of those 138 articles, 111 (i.e., 80.43%) articles relied on at least one of the three MULs.

our literature review were not sufficiently transparent in their reporting for us to determine precisely how influential outliers were identified. For example, in an article published in *Strategic Management Journal* in 2009, the authors "ran regression diagnostics to look for outliers and removed seven observations that substantially skewed regression results, consistent with normal practice." In another article published in *Journal of Applied Psychology* in 2008, the authors noted that "on the basis of an outlier analysis, three cases were dropped from the U.S. sample, as they contributed most to departures of multivariate kurtosis." This lack of transparency is obviously an issue that needs to be addressed, and later in our manuscript we

suggest that journal policies should motivate authors to include at least a few sentences on how they identified influential outliers. Of the remaining 138 articles that were sufficiently transparent in reporting how influential outliers were identified, 111 (i.e., 80.43%) relied on at least one of the three MULs. As a preview of the next sections, Table 10.1 offers a summary of each MUL, the kernels of truth behind each MUL, and how the kernels of truth have been misapplied over time.

Univariate Cutoffs Are Sufficient for Identifying Influential Outliers

According to this MUL, influential outliers are identified as cases that are far from others given a distribution of data points for a single variable. Among the 138 journal articles in our review that were sufficiently transparent in reporting how influential outliers were identified, 89 (i.e., 64.49%) relied on this MUL. As an example among micro-level studies, Stajkovic and Luthans (1997) conducted a meta-analysis, and "to estimate the relative stability of unbiased effect-size magnitudes . . . effect sizes positioned 1.5 to 3 lengths from the upper or lower edge of the 50 percent interquartile range . . . were considered outliers" (p. 1127). As an illustration in the macro domain, Henkel (2009) identified firms lying within the extreme 1% of the return on equity distribution and treated them as special cases "to restrict the influence of outliers" (p. 293). Note that our discussion here neither pertains to nor criticizes how researchers in these examples subsequently handled influential outliers. For example, Stajkovic and Luthans (1997) reported results with and without the influential outliers they identified—our focus is in on how outliers were identified.

There are two kernels of truth underlying this MUL—a conceptual-logical one and an authoritative one. First, the conceptual-logical kernel of truth is that univariate cutoffs help identify distant cases, which are often influential outliers. Specifically, univariate cutoffs "have some utility for identifying extreme cases" (Meade & Craig, 2012, p. 440). In turn, such "unusual cases that are far from the rest of the data . . . even one, can seriously jeopardize the results and conclusions of the regression analysis" (Cohen et al., 2003, p. 102). Accordingly, it seems logical to use "distance" as a proxy for "influence."

Second, the authoritative rationale used by some researchers is that the sources they cite describe univariate cutoffs in detail. For example, Cohen and colleagues (2003, chapter 4) discussed how to use boxplots and, in doing so, stated that "values of any outlying scores are displayed separately when they fall below $Q_1 - 3SIQR$ or above $Q_3 + 3SIQR$" (p. 108). Note that SIQR is the semi-interquartile range, or (Q3 – Q1)/2. Similarly, Tabachnick and Fidell (2007, chapter 4) stated that "cases with standardized scores in excess of 3.29 ($p < .001$, two-tailed test) are potential outliers" (p. 73). Other similarly influential and widely used textbooks that discuss univariate cutoffs to identify cases lying at a distance from others in a distribution include Tukey (1977) and Hildebrand (1986).

Unfortunately, the kernels of truth seem to have been misapplied over time in two ways. First, many authors have come to misunderstand distance as not only a necessary but also a sufficient condition for influence. Researchers using univariate cutoffs often produce false positives (i.e., deciding that a distant case is an influential outlier when it is not influential) and, sometimes, also false negatives (i.e., deciding that a case seemingly not far from others is not an influential outlier when it is influential; Aguinis et al., 2013). Thus, careful examination of data can reveal cases that are far from others but do not have influence on the results. For example, McCann and Vroom (2010) noted that a hotel had an unusually large number of rooms, yet further examination of that case revealed that its exclusion from the data set actually did not change any of the results in a substantive manner.

Second, many authors seem to have jumped to the incorrect conclusion that because an authoritative methodological source describes univariate cutoffs, the inference is that such cutoffs are recommended as sufficient means for identifying influential outliers. In fact, just because authoritative and widely used methodological sources describe univariate cutoffs, this does not mean that they have recommended that such cutoffs alone be used for identifying influential outliers. For example, although Cohen and colleagues (2003, chapter 4) described univariate cutoffs, they noted in the same chapter that outliers are given more detailed consideration later in chapter 10, in which they "encourage . . . the use of specialized statistics known as regression diagnostics which can greatly aid in the detection of outliers" (p. 394). Thus, Cohen and colleagues actually discouraged the sole use of univariate cutoffs to identify influential outliers.

Summary

The nature of the MUL: Univariate cutoffs are sufficient for identifying influential outliers. The kernels of truth: Univariate cutoffs help identify distant cases, which are often influential outliers. Also, authoritative methodological sources describe univariate cutoffs in detail. How the kernels of truth have been misapplied over time to form the MUL: Many authors have come to misunderstand distance as not only a necessary but also a sufficient condition for influence. Further, many authors seem to have jumped to the incorrect conclusion that an authoritative methodological source describing univariate cutoffs means that such cutoffs are recommended as sufficient means for identifying influential outliers.

Inspection of Visual Plots Is Sufficient for Identifying Influential Outliers

The second MUL involves using visual plots such as histograms, scatter plots, residual plots, and index plots as sufficient means for identifying influential outliers. Of the 138 articles in our review that were sufficiently transparent in reporting how influential outliers were identified, 12.32% (17 articles) relied on this MUL. Among these 17 studies, some used univariate visual plots (e.g., histograms), while others used multivariate visual plots. Multivariate visual plots include multiple variables (e.g., scatter plots) as well as plots of multivariate test statistics (e.g., residual plots, index plots). As an illustration in the micro domain, Blanton and colleagues (2009) examined a number of scatter plots "to determine visually if there were apparent outliers whose presence might have influenced the trend of the data within conditions" (p. 578). As an example in the macro domain, Bogert (1996) examined the data distributions of the dependent variables as the means to try to identify outliers that "unduly influenced the reported regression results" (p. 248). Note that our discussion here neither pertains to nor criticizes how researchers in these examples subsequently handled influential outliers. Specifically, Blanton and colleagues (2009) and Bogert (1996) reported results with and without influential outliers—which is a sound practice recommended by Aguinis and colleagues (2013).

There are four main reasons why the use of visual plots as a necessary and sufficient means for identifying influential outliers is inappropriate (Cohen et al., 2003; Iglewicz & Hoaglin, 1993; Ziegert & Hanges, 2009). First, similar to the first MUL, this practice relies on the incorrect logic that a case with a large distance from others necessarily means that the case also has a large influence on the study's results. Second, the determination of exactly which cases are identified as influential outliers in the same visual plot may vary from one researcher to another depending on a person's subjective judgment. Not surprisingly, some have described the practice as "a notoriously flawed approach for detecting outliers" (Ziegert & Hanges, 2009, p. 593) and "not a reliable way to identify potential outliers" (Iglewicz & Hoaglin, 1993, p. 9). Further, a cynical view is that researchers using visual plots are more likely to "find" influential outliers for the purpose of finding better support for one's hypothesis—an inappropriate practice that capitalizes on chance (Cortina, 2002) and borders on unethical research conduct (Bedeian, Taylor, & Miller, 2010). Third, visual plots used for identifying influential outliers "suffer in small samples because of the small number of comparators available" (Martin & Roberts, 2010, p. 258). In other words, the same cases may or may not be identified as outliers depending on the size of the sample. Fourth, the practice of using visual plots to identify influential outliers is usually accompanied by low transparency. Stated differently, replicating the decision of labeling a case as an outlier is difficult if a plot is not accompanied by a verbal description of exactly which cases were identified as influential outliers and why.

There are two kernels of truth underlying this MUL—a practical one and an authoritative one. First, in terms of practicality, superior methods (i.e., cutoffs based on multivariate test statistics that take into account the research design features of a study) simply did not exist or were not available for practical use before the early 1980s (Martin & Roberts, 2010). Accordingly, visual plots were a good practical alternative, although they only have limited ability to identify influential outliers.

The second kernel of truth is that authoritative methodological sources (e.g., textbooks) give detailed descriptions of a number of visual techniques, thereby possibly giving the impression that using them alone to identify influential outliers is acceptable. For example, Cohen and colleagues (2003, chapter 4) explained how to use a variety of visual techniques—although they do not state that using plots alone is the recommended procedure for identifying influential outliers.

Unfortunately, the kernels of truth have been misapplied over time in largely two ways. First, many authors have continued to use visual plots alone to identify influential outliers, even though there have been a number of developments that now make it practical to use better alternative procedures for identifying influential outliers. Specifically, seminal works by Cook (1977, 1979), Belsley, Kuh, and Welsch (1980), and Cook and Weisberg (1982) have provided more appropriate procedures (i.e., cutoffs based on multivariate test statistics that take into account research design features of a study). Another development is that high-speed computers have become more readily available, which have facilitated the implementation of these procedures. As a result, it is no longer justified on practical grounds to use visual plots as a sufficient means to identify influential outliers.

Second, many authors seem to have jumped to the incorrect conclusion that an authoritative methodological source describing visual plots means that visual plots are sufficient means for identifying influential outliers. In fact, just because authoritative methodological sources give detailed descriptions of a number of visual plots, this does not mean these sources recommend the use of visual plots as sufficient means for identifying influential outliers. Further, the same authoritative methodological sources (e.g., Cohen et al., 2003; Iglewicz & Hoaglin, 1993) describing various visual plots in detail also discourage the use of visual plots as sufficient means for identifying influential outliers, as noted in the discussion regarding the nature of the MUL.

Summary

The nature of the MUL: Inspection of visual plots is sufficient for identifying influential outliers. The kernels of truth: Superior alternative procedures for identifying influential outliers did not exist or were not available for practical use before the early 1980s. Also, commonly used methodological sources give detailed descriptions of a number of visual techniques. How the kernels of truth have been misapplied over time to form the MUL: Many authors have continued to use visual plots alone to identify influential outliers, even though improvements in statistical methods and computing technology no longer justify the exclusive use of visual plots on practical grounds. Further, many authors seem to have jumped to the incorrect conclusion that an authoritative methodological source describing visual plots means that visual plots are sufficient means for identifying influential outliers.

Absolute Cutoffs Based on Multivariate Test Statistics Are Appropriate for Identifying Influential Outliers

The third MUL involves using absolute cutoffs based on multivariate test statistics to identify influential outliers. According to this MUL, influential outliers are cases whose multivariate test statistic values exceed a numeric threshold, and this threshold is exactly the same regardless of a study's research design features such as sample size and number of variables investigated. As we will describe and illustrate later in our manuscript, considering research design features improves accuracy in the process of identifying influential outliers.

Of the 138 articles included in our review that were sufficiently transparent in reporting how influential outliers were identified, 8.70% (12 articles) relied on absolute cutoffs based on multivariate test statistics. As an example in the micro domain, Montes and Zweig (2009) looked for observations with standardized residuals beyond ± 3 SD units to identify data points that "might adversely affect the validity of the results" (p. 1249). As an illustration in the macro domain, Wright, Kroll, Krug, and Pettus (2007) looked for observations with residual values larger than 4 SD units to identify "firms that unduly influenced the regression results" (p. 86).

The kernel of truth behind this MUL seems to be based on the reliance on authoritative sources. For example, Belsley and colleagues (1980) actually did give credit to the utility of absolute cutoffs based on multivariate test statistics when they stated that "it is natural to say, at least to a first approximation, that any of the diagnostic measures is large if its value exceeds two in magnitude. Such a procedure defines what we call an absolute cutoff" (p. 28). Granted, Belsley and colleagues (1980, p. 28) also described and endorsed research design–based cutoffs based on multivariate test statistics—that is, cutoffs based on multivariate test statistics that take into account the research design features of a study. For example, observations with DFFITS values (i.e., similar to Cook's D but using a different scale) above or below $\pm 2\sqrt{\frac{(k+1)}{n}}$ are considered influential outliers, where k = number of predictors and n = sample size. But Belsley and colleagues (1980) do not seem to have stated explicitly that research design–based cutoffs are superior to absolute cutoffs based on multivariate test statistics. As another example of such ambiguity, Cohen and colleagues (2003, p. 404) stated that "a value of 1.0 or the critical value of the F distribution at $\alpha = .50$ with $df = (k + 1, n - k - 1)$ is used" regarding Cook's D, thereby making it seem that it does not matter whether a researcher uses absolute or research design–based cutoffs based on multivariate test statistics. Thus, the apparent approval of absolute cutoffs based on multivariate test statistics—which can be used across studies regardless of their design features—subsequently seems to have led to the widespread use of such absolute cutoffs to identify influential outliers.

The kernel of truth has been misapplied because many authors have continued to use such absolute cutoffs to identify influential outliers, even though evidence has accumulated demonstrating that the process of identifying influential outliers must include research design considerations (Andrews & Pregibon, 1978; Chatterjee & Hadi, 1986; Martin & Roberts, 2010). In other words, research design–based cutoffs, compared to absolute cutoffs, assess influence more accurately, as we describe in the following two illustrations.

First, consider DFFITS, which assesses the influence that a data point has on all regression coefficients in a regression model as a whole. The cutoff value for DFFITS in a study with a sample size of 100 (n = 100) and 10 predictors (k = 10) is $\pm 2\sqrt{(10+1)/100} = \pm 0.663$, the absolute value of which is about twice as large as that of ±0.346 in a study with the same sample size (n = 100) but 2 predictors (k = 2). Through this adjustment in the cutoff values for DFFITS is based on number of predictors, one can assess influence more accurately because as the number of predictors increases, so does the number of regression coefficients as a whole that a data point must affect to be an influential outlier. This "increased bar" for a data point to be influential is therefore reflected in the higher cutoff value for DFFITS.

As a second illustration, once again referring to DFFITS, note that the cutoff value in a study where k = 2 and n = 100 is $\pm 2\sqrt{\frac{2+1}{100}} = \pm 0.346$, the absolute value of which is substantially larger than that of ±0.173 for a study where k = 2 but n = 400. Through this adjustment in the cutoff values for DFFITS based on sample size, one can assess influence more accurately, because even if two cases cause the same overall amount of change in the same regression coefficients, the case in the model with the smaller sample size (i.e., fewer "competitors") is less influential than the other case in the model with the larger sample size (i.e., more "competitors"). To account for these differences in terms of "competition," the cutoff value for DFFITS decreases as sample size increases.

Summary

The nature of the MUL: Absolute cutoffs based on multivariate test statistics are appropriate for identifying influential outliers. The kernel of truth: Authoritative methodological sources seem to approve of the use of absolute cutoffs based on multivariate test statistics to identify influential outliers. How the kernel of truth has been misapplied over time to form the MUL: Many authors have continued to use such absolute cutoffs to identify influential outliers, even though evidence has accumulated pointing to the superiority of research design-based cutoffs. Research design-based cutoffs are superior because they can vary from study to study depending on the characteristics of the particular research context such as sample size and number of predictors. Specifically, the "bar" for considering a case as an influential outlier is higher as sample size decreases and number of predictors increases in a model.

Best-Practice Recommendations on How to Identify Influential Outliers

In this section, we offer best-practice recommendations on how researchers should proceed in terms of identifying influential outliers. These recommendations are necessary in light of the pervasiveness of practices based on the three MULs that

TABLE 10.2 Summary of Best-Practice Recommendations on How to Identify Influential Outliers

Recommendation	*Description*
Follow Aguinis et al.'s (2013) three-step approach, regardless of the particular data-analytic context.	• Step 1: Identify error outliers. • Step 2: Identify interesting outliers. • Step 3: Identify influential outliers.
Identify two types of influential outliers—model fit outliers and prediction outliers—in the context of multiple regression, SEM, or multilevel modeling.	• When identifying model fit outliers, use a two-step process: (1) identify cases that exceed cutoffs based on suitable techniques, and (2) check whether the removal of each previously identified case changes model fit. • When identifying prediction outliers, identify cases that exceed cutoffs based on suitable techniques. • These various techniques and cutoffs, as well as practical implementation guidelines, used in multiple regression, SEM, or multilevel modeling, are discussed in Aguinis et al. (2013).
Use research design–based cutoffs to identify influential outliers.	• For example, for DFFITS (i.e., used to assess the influence that a data point has on all regression coefficients in a regression model as a whole), the recommended cutoff is $\pm 2\sqrt{\frac{(k+1)}{n}}$, where k = number of predictors and n = sample size. • As another example, for Cook's D values of cases, the recommended research design-based cutoff is: F distribution at α = .50 with $df = (k + 1, n - k - 1)$.
Use visual techniques alone when there are no research design–based cutoffs available.	• For example, in SEM, it is acceptable to use index plots when using generalized Cook's D and single parameter influence. • An index plot includes case numbers on the X axis and test statistic values on the Y axis.
Identify influential outliers when using meta-analysis.	• Calculate the SAMD value for each primary-level study. • Regarding the recommended research design–based cutoff, use a scree plot of SAMD values, where SAMD values of primary studies are plotted from the highest to the lowest value on the Y-axis while the corresponding rank-ordered position of each primary-level study is denoted on the X-axis. Studies with SAMD values that lie above the "elbow" (i.e., the point that separates the steep slope from the gradual slope in the scree plot) are identified as influential outliers.
Identify influential outliers when using time series analysis.	• Use independent component analysis (ICA). • The recommended research design–based cutoff is: $\mu_i \pm 4.47\sigma_i$, where μ_i and σ_i are the mean and the standard error of the ith extracted component, respectively.

we described earlier. Table 10.2 offers a summary of the recommendations we discuss next, which also include illustrations of how these recommended procedures have been implemented in published articles.

Our first recommendation is to follow a sequential process consisting of three broad steps as identified by Aguinis and colleagues (2013). These three steps should be applied regardless of the particular data-analytic approach (e.g., regression, SEM, meta-analysis, multilevel modeling) used for assessing substantive questions and hypotheses. In the recommended sequential process, a researcher first needs to identify error outliers (i.e., outlying cases caused by undesirable reasons such as mistakes made in the research process), then interesting outliers (i.e., outlying cases caused not by mistakes but instead by potentially interesting substantive reasons) and, finally, influential outliers (i.e., outlying cases that are neither error nor necessarily interesting cases and that affect substantive conclusions of the study). Thus, identifying influential outliers in the third step of the process ensures that the cases identified as influential outliers are such, as opposed to other types of outliers.

In the particular context of multiple regression, SEM, and multilevel modeling, the third step in the aforementioned sequential process involves identifying two types of influential outliers: (1) model fit outliers (i.e., cases whose presence alters the fit of a model) and (2) prediction outliers (i.e., cases whose presence alters parameter estimates). When identifying model fit outliers, the researcher should first use suitable techniques and cutoffs as well as subsequently check whether model fit is changed by the removal of each identified case. This two-step process is necessary because the techniques suited for identifying model fit outliers assess the distance of a case from other cases instead of the influence of the case on model fit. When identifying prediction outliers, the researcher only needs to use suitable techniques and cutoffs and does not need to subsequently check whether model fit is changed by the removal of each identified case, because the techniques suited for identifying prediction outliers directly assess the influence of the case on parameter estimates. These various techniques and cutoffs, as well as practical implementation guidelines, used in multiple regression, SEM, or multilevel modeling are also discussed by Aguinis and colleagues (2013). For example, in line with Aguinis and colleagues' (2013) recommended two-step process for identifying model fit outliers, Baldridge, Floyd, and Markóczy (2004) first identified potential model fit outliers and subsequently checked whether each potential model fit outlier actually had influence on the fit of the model. As a result, Baldridge and colleagues (2004) found that three of the five potential model fit outliers were indeed model fit outliers. Had the researchers neglected the second step of checking whether each potential model fit outlier was in fact a model fit outlier, they would have erroneously identified two additional observations as influential outliers.

As an additional recommendation, we emphasize the importance of using research design–based cutoffs when using specific techniques to identify influential outliers. For example, Grant (2008) used *DFBETAS* (i.e., indicating whether the inclusion of a case leads to an increase or decrease in a single regression coefficient) and Cook's *D* to identify prediction outliers in his hierarchical regression model. For both techniques, Grant (2008) used research design–based cutoffs that take into account the number of cases and predictors in the model. As an illustration of model fit outlier identification in the analytic context of SEM, Goerzen (2007) first derived the Mahalanobis distance values (i.e., the length of the line segment between a data point and the centroid of the remaining cases), identified those cases that exceeded the research design–based cutoff used, and then checked whether the removal of the identified cases changed the fit of the tested models (though we recommend that such removal be done with one identified case at a time).

There are unique circumstances when it is appropriate to not use research design–based cutoffs and instead use visual techniques alone for identifying influential outliers (i.e., the second MUL we discussed). This recommendation applies to situations for which there are no research design–based cutoffs available for practical use. For example, in SEM, it is recommended that researchers use two techniques (i.e., multivariate test statistics)—generalized Cook's *D* and single parameter influence—to identify prediction outliers (Pek & MacCallum, 2011). Because there are no research design–based cutoffs available, it is acceptable to use index plots that include case numbers on the X axis and test statistic values on the Y axis.

Next, we offer recommendations for two additional data-analytic contexts: meta-analysis and time series analysis. Use of these two analytic techniques is fairly typical in organizational science research, yet recommendations on how to identify influential outliers in these contexts were not discussed by Aguinis and colleagues (2013). First of all, as mentioned earlier, the researcher must identify error and then interesting outliers before identifying influential outliers. To identify influential outliers in the context of meta-analysis, we recommend that researchers calculate the sample-adjusted meta-analytic deviancy (SAMD) statistic value for each observation, or the effect size estimate from each primary-level study included in the meta-analysis (Huffcutt & Arthur, 1995). This technique is recommended because an influential outlier is a function of both effect size and sample size, and SAMD takes into account both. The recommended research design–based cutoff involves using a scree plot of SAMD values, which are plotted from the highest to the lowest value on the Y-axis, while the corresponding rank-ordered position of each primary-level study is denoted on the X-axis (Arthur, Bennett, & Huffcutt, 2001). Primary-level studies with SAMD values that lie above the "elbow" (i.e., the point that separates the steep slope from the gradual slope in the scree plot) are identified as influential outliers. This cutoff takes research design features into consideration because the primary-level studies with SAMD values above the elbow are those that, compared to others, contribute substantially more to the variance across the primary-level studies in the particular meta-analytic data base at hand. So the exact location of the elbow varies from one meta-analysis to another. As an illustration of the recent use of this approach, Taylor, Russ-Eft, and Taylor (2009) conducted a meta-analysis of the transfer of training literature and used a scree plot of SAMD values to identify influential outliers.

Finally, to identify influential outliers in the context of time series analysis, we recommend the use of independent component analysis (ICA), accompanied by the research design–based cutoff of "$\mu i \pm 4.47\sigma i$, where μi and σi are, respectively, the mean and the standard error of the *i*th extracted component" (Baragona & Battaglia, 2007, p. 1973). MATLAB code for implementing ICA to identify influential outliers is publicly available online and has been developed by Bell and Sejnowski (1995; www.sccn.ucsd.edu/eeglab/) and Hyvärinen and Oja (2000; www.cis.hut.fi/projects/ica/fastica/).

Concluding Comments

The presence of outliers seems to be an unavoidable fact of life when conducting organizational science research (Aguinis & O'Boyle, 2014; O'Boyle & Aguinis, 2012). Thus, it is important that researchers address influential outliers appropriately, as well as report how they dealt with such cases openly and transparently. Our content analysis of 232 substantive journal articles that mentioned the term "outlier" revealed that about 40% did not provide sufficient information for us to understand the procedures that were implemented to identify these particular cases. Among studies that reported sufficient information on how authors identified influential outliers, about 80% of them have fallen prey to at least one of the three myths and urban legends that we described in our manuscript. Each of these MULs is inappropriate because they are based on the commonly invoked but incorrect assumption that a case with a large distance from others also *necessarily* has a large influence on the study's results.

As noted by Aguinis and colleagues (2013),

> without a description of the identification techniques used, a skeptical scientific audience might raise doubts about a study's substantive conclusions . . . [because] . . . a cynical view is that outliers are treated in such a way that their inclusion or exclusion from a data set is not based on sound and standardized practices, but on whether results favor one's preferred hypotheses.
> *(pp. 292, 297)*

We hope our manuscript will allow researchers to critically revisit common practices about how to identify influential outliers, as well as encourage researchers to adopt more appropriate practices. Also, we would like to offer the proposal that journal editors and reviewers make a proactive effort to ensure transparent reporting practices regarding outliers. This can be done by requiring authors of manuscripts describing empirical research to include at least a few sentences on how they identified influential outliers—this material may be included in a separate section titled "Outlier Identification and Management." Overall, we hope our manuscript will lead to the use of more appropriate and transparent practices for identifying influential outliers in future research.

Acknowledgments

Some of the material included in this manuscript was presented as professional development workshops at the meetings of the Academy of Management in 2011 (San Antonio, TX), 2012 (Boston, MA), and 2013 (Orlando, FL). We thank Bob Vandenberg for his highly constructive and detailed feedback that allowed us to improve our manuscript substantially.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*, 270–301.

Aguinis, H., & O'Boyle, E. (2014). Star performers in twenty-first-century organizations. *Personnel Psychology, 67*, 313–350.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of *Organizational Research Methods*: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods, 12*, 69–112.

Andrews, D. F., & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society, Series B*, 85–93.

Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah, NJ: Erlbaum.

Baldridge, D. C., Floyd, S. W., & Markóczy, L. (2004). Are managers from Mars and academicians from Venus? Toward an understanding of the relationship between academic quality and practical relevance. *Strategic Management Journal, 25*, 1063–1074.

Baragona, R., & Battaglia, F. (2007). Outliers detection in multivariate time series by independent component analysis. *Neural Computation, 19*, 1962–1984.

Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning and Education, 9*, 715–725.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, NY: John Wiley & Sons, Inc.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*, 567–582.

Bogert, J. D. (1996). Explaining variance in the performance of long-term corporate blockholders. *Strategic Management Journal, 17*, 243–249.

Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science, 1*, 379–393.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*, 15–18.
Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association, 74*, 169–174.
Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman and Hall.
Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management, 28*, 339–362.
Goerzen, A. (2007). Alliance networks and firm performance: The impact of repeated partnerships. *Strategic Management Journal, 28*, 487–509.
Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology, 93*, 48–58.
Henkel, J. (2009). The risk-return paradox for strategic management: Disentangling true and spurious effects. *Strategic Management Journal, 30*, 287–303.
Hildebrand, D. (1986). *Statistical thinking for behavioral scientists*. Boston, MA: Duxbury Press.
Hitt, M. A., Harrison, J. S., Ireland, R. D., & Best, A. (1998). Attributes of successful and unsuccessful acquisitions of US firms. *British Journal of Management, 9*, 91–114.
Hollenbeck, J. R., DeRue, D. S., & Mannor, M. (2006). Statistical power and parameter stability when subjects are few and tests are many: Comment on Peterson, Smith, Martorana, and Owens (2003). *Journal of Applied Psychology, 91*, 1–5.
Huffcutt, A. I., & Arthur, W., Jr. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80*, 327–334.
Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430.
Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers*. Milwaukee, WI: American Society for Quality Control.
Martin, M. A., & Roberts, S. (2010). Jackknife-after-bootstrap regression influence diagnostics. *Journal of Nonparametric Statistics, 22*, 257–269.
McCann, B. T., & Vroom, G. (2010). Pricing response to entry and agglomeration effects. *Strategic Management Journal, 31*, 284–305.
Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437–455.
Montes, S. D., & Zweig, D. (2009). Do promises matter? An exploration of the role of promises in psychological contract breach. *Journal of Applied Psychology, 94*, 1243–1260.
O'Boyle, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65*, 79–119.
Orr, J. M., Sackett, P. R., & DuBois, C.L.Z. (1991). Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology, 44*, 473–486.
Pek, J., & MacCallum, R. C. (2011). Sensitivity analysis in structural equation models: Cases and their influence. *Multivariate Behavioral Research, 46*, 202–228.
Peterson, R. S., Smith, D. B., Martorana, P. V., & Owens, P. D. (2003). The impact of chief executive officer personality on top management team dynamics: One mechanism by which leadership affects organizational performance. *Journal of Applied Psychology, 88*, 795–808.

Stajkovic, A. D., & Luthans, F. (1997). A meta-analysis of the effects of organizational behavior modification on task performance, 1975–95. *Academy of Management Journal, 40*, 1122–1149.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.
Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology, 94*, 104–121.
Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Pearson.
Wright, P., Kroll, M., Krug, J. A., & Pettus, M. (2007). Influences of top management team incentives on firm risk taking. *Strategic Management Journal, 28*, 81–89.
Ziegert, J. C., & Hanges, P. L. (2009) Strong rebuttal to weak criticisms: Reply to Blanton et al. (2009). *Journal of Applied Psychology, 94*, 590–597.