Correlation Coefficients: Appropriate Use and Interpretation

3 authors, including:

Patrick Schober
Amsterdam University Medical Center
272 PUBLICATIONS   **15,113** CITATIONS

Christa Boer
Amsterdam University Medical Center
268 PUBLICATIONS   **17,785** CITATIONS

Correlation Coefficients: Appropriate Use and Interpretation

Patrick Schober, MD, PhD, MMedStat, Christa Boer, PhD, MSc, and Lothar A. Schwarte, MD, PhD, MBA

Correlation in the broadest sense is a measure of an association between variables. In correlated data, the change in the magnitude of 1 variable is associated with a change in the magnitude of another variable, either in the same (positive correlation) or in the opposite (negative correlation) direction. Most often, the term correlation is used in the context of a linear relationship between 2 continuous variables and expressed as Pearson product-moment correlation. The Pearson correlation coefficient is typically used for jointly normally distributed data (data that follow a bivariate normal distribution). For nonnormally distributed continuous data, for ordinal data, or for data with relevant outliers, a Spearman rank correlation can be used as a measure of a monotonic association. Both correlation coefficients are scaled such that they range from −1 to +1, where 0 indicates that there is no linear or monotonic association, and the relationship gets stronger and ultimately approaches a straight line (Pearson correlation) or a constantly increasing or decreasing curve (Spearman correlation) as the coefficient approaches an absolute value of 1. Hypothesis tests and confidence intervals can be used to address the statistical significance of the results and to estimate the strength of the relationship in the population from which the data were sampled. The aim of this tutorial is to guide researchers and clinicians in the appropriate use and interpretation of correlation coefficients.   (Anesth Analg 2018;126:1763–8)

Researchers often aim to study whether there is some association between 2 observed variables and to estimate the strength of this relationship. For example, Nishimura et al[1] assessed whether the volume of infused crystalloid fluid is related to the amount of interstitial fluid leakage during surgery, and Kim et al[2] studied whether opioid growth factor receptor (OGFR) expression is associated with cell proliferation in cancer cells. These and similar research objectives can be quantitatively addressed by correlation analysis, which provides information about not only the strength but also the direction of a relationship (eg, an increase in OGFR expression is associated with an increase or a decrease in cell proliferation).

As part of the ongoing series in *Anesthesia & Analgesia*, this basic statistical tutorial discusses the 2 most commonly used correlation coefficients in medical research, the Pearson coefficient and the Spearman coefficient.[3] It is important to note that these correlation coefficients are frequently misunderstood and misused.[4,5] We thus focus on how they should and should not be used and correctly interpreted.

PEARSON PRODUCT-MOMENT CORRELATION

Correlation is a measure of a monotonic association between 2 variables. A monotonic relationship between 2 variables is a one in which either (1) as the value of 1 variable increases, so does the value of the other variable; or (2) as the value of 1 variable increases, the other variable value decreases.

In correlated data, therefore, the change in the magnitude of 1 variable is associated with a change in the magnitude of another variable, either in the same or in the opposite direction. In other words, higher values of 1 variable tend to be associated with either higher (positive correlation) or lower (negative correlation) values of the other variable, and vice versa.

A linear relationship between 2 variables is a special case of a monotonic relationship. Most often, the term "correlation" is used in the context of such a linear relationship between 2 continuous, random variables, known as a Pearson product-moment correlation, which is commonly abbreviated as "*r*."[6]

The degree to which the change in 1 continuous variable is associated with a change in another continuous variable can mathematically be described in terms of the covariance of the variables.[7] Covariance is similar to variance, but whereas variance describes the variability of a single variable, covariance is a measure of how 2 variables vary together.[7] However, covariance depends on the measurement scale of the variables, and its absolute magnitude cannot be easily interpreted or compared across studies. To facilitate interpretation, a Pearson correlation coefficient is commonly used. This coefficient is a dimensionless measure of the covariance, which is scaled such that it ranges from −1 to +1.[7]

Figure 1 shows scatterplots with examples of simulated data sampled from bivariate normal distributions with different Pearson correlation coefficients. As illustrated, *r* = 0 indicates that there is no linear relationship between the
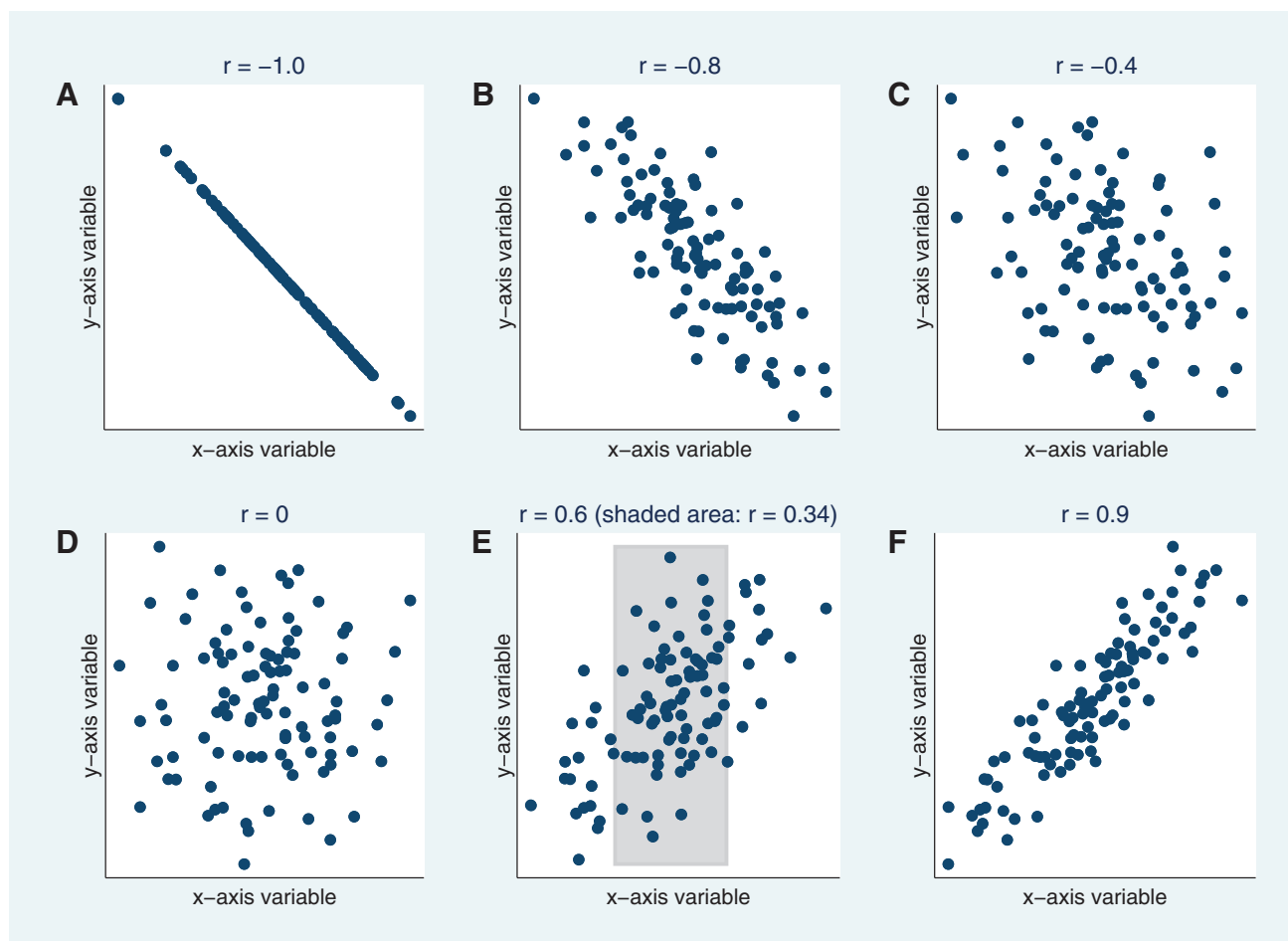
Figure 1. A–F, Scatter plots with data sampled from simulated bivariate normal distributions with varying Pearson correlation coefficients (*r*). Note that the scatter approaches a straight line as the coefficient approaches −1 or +1, whereas there is no linear relationship when the coefficient is 0 (D). E shows by example that the correlation depends on the range of the assessed values. While the coefficient is +0.6 for the whole range of data shown in E, it is only +0.34 when calculated for the data in the shaded area.

variables, and the relationship becomes stronger (ie, the scatter decreases) as the absolute value of *r* increases and ultimately approaches a straight line as the coefficient approaches −1 or +1.

A perfect correlation of −1 or +1 means that all the data points lie exactly on the straight line, which we would expect, for example, if we correlate the weight of samples of water with their volume, assuming that both quantities can be measured very accurately and precisely. However, such absolute relationships are not typical in medical research due to variability of biological processes and measurement error.

Assumptions of a Pearson Correlation

Assumptions of a Pearson correlation have been intensely debated.[8–10] It is therefore not surprising, but nonetheless confusing, that different statistical resources present different assumptions. In reality, the coefficient can be calculated as a measure of a linear relationship without any assumptions.

However, proper inference on the strength of the association in the population from which the data were sampled (what one is usually interested in) does require that some assumptions be met:[9–11]

1. As is actually true for any statistical inference, the data are derived from a random, or at least representative, sample. If the data are not representative of the population of interest, one cannot draw meaningful conclusions about that population.
2. Both variables are continuous, jointly normally distributed, random variables. They follow a bivariate normal distribution in the population from which they were sampled. The bivariate normal distribution is beyond the scope of this tutorial but need not be fully understood to use a Pearson coefficient.

Two typical properties of the bivariate normal distribution can be relatively easily assessed, and researchers should check approximate compliance of their data with these properties:

i. Both variables are normally distributed. Methods to assess this assumption have recently been reviewed in this series of basic statistical tutorials.[12]
ii. If there is a relationship between jointly normally distributed data, it is always linear.[13] Therefore, if the data points in a scatter plot seem to lie close to some curve, the assumption of a bivariate normal distribution is violated.

There are several possibilities to deal with violations to this assumption. First, variables can often be transformed to approach a normal distribution and to linearize the relationship between the variables.[12] Second, in contrast to a Pearson correlation, a Spearman correlation (see below) does not require normally distributed data and can be used to analyze nonlinear monotonic (ie, continuously increasing or decreasing) relationships.[14]

3. There are no relevant outliers. Extreme outliers may have undue influence on the Pearson correlation coefficient. While it is generally not legitimate to simply exclude outliers,[15] running the correlation analysis with and without the outlier(s) and comparing the coefficients is a possibility to assess the actual influence of the outlier on the analysis. For data with relevant outliers, Spearman correlation is preferred as it tends to be relatively robust against outliers.[14]

4. Each pair of x–y values is measured independently from each other pair. Multiple observations from the same subjects would violate this assumption.[11] The way to deal with such data depends on whether we are interested in correlations within subjects or between subjects as reviewed previously.[16,17]

Interpretation of the Correlation Coefficient

Several approaches have been suggested to translate the correlation coefficient into descriptors like "weak," "moderate," or "strong" relationship (see the Table for an example).[3,18] These cutoff points are arbitrary and inconsistent and should be used judiciously. While most researchers would probably agree that a coefficient of <0.1 indicates a negligible and >0.9 a very strong relationship, values in-between are disputable. For example, a correlation coefficient of 0.65 could either be interpreted as a "good" or "moderate" correlation, depending on the applied rule of thumb. It is also quite capricious to claim that a correlation coefficient of 0.39 represents a "weak" association, whereas 0.40 is a "moderate" association.

Rather than using oversimplified rules, we suggest that a specific coefficient should be interpreted as a measure of the strength of the relationship in the context of the posed scientific question. Note that the range of the assessed values should be considered in the interpretation, as a wider range of values tends to show a higher correlation than a smaller range (Figure 1E).[19]

The observed correlation may also not necessarily be a good estimate for the population correlation coefficient, because samples are inevitably affected by chance. Therefore, the observed coefficient should always be accompanied by a confidence interval, which provides the range of plausible values of the coefficient in the population from which the data were sampled.[20]

In the study by Nishimura et al,[1] the authors report a correlation coefficient of 0.42 for the relationship between the infused crystalloid volume and the amount of interstitial fluid leakage, so there appears to be a considerable association between the 2 variables. However, the 95% confidence interval, which ranges from 0.03 to 0.70, suggests that the results are also compatible with a negligible ($r = 0.03$) and

Table. Example of a Conventional Approach to Interpreting a Correlation Coefficient

Absolute Magnitude of the Observed Correlation Coefficient	Interpretation
0.00–0.10	Negligible correlation
0.10–0.39	Weak correlation
0.40–0.69	Moderate correlation
0.70–0.89	Strong correlation
0.90–1.00	Very strong correlation

Several stratifications (with different cutoff points) have been previously published.

hence clinically unimportant relationship. On the other hand, the data are also compatible with a quite strong association ($r = 0.70$). Data with such a wide confidence interval do not allow a definitive conclusion about the strength of the relationship between the variables.

Researchers typically also aim to determine whether their result is "statistically significant." A t test is available to test the null hypothesis that the correlation coefficient is zero.[13] Note that the P value derived from the test provides no information on how strongly the 2 variables are related. With large datasets, very small correlation coefficients can be "statistically significant." Therefore, a statistically significant correlation must not be confused with a clinically relevant correlation. For further information on how results of hypothesis tests and confidence intervals should be interpreted, we refer the reader to previous tutorials in *Anesthesia & Analgesia*.[20,21]

Coefficient of Determination

The correlation coefficient is sometimes criticized as having no obvious intrinsic interpretation,[6] and researchers sometimes report the square of the correlation coefficient. This R^2 is termed the "coefficient of determination." It can be interpreted as the proportion of variance in 1 variable that is accounted for by the other.[6]

The correlation coefficient of 0.42 reported by Nishimura et al[1] corresponds to a coefficient of determination (R^2) of 0.18, suggesting that about 18% of the variability of the amount of interstitial fluid leakage can be "explained" by the relationship with the amount of infused crystalloid fluid. As more than 80% of the variability is yet unexplained, there must be 1 or more other relevant factors that are related to interstitial leakage.

In interpreting the coefficient of determination, note that the squared correlation coefficient is always a positive number, so information on the direction of a relationship is lost. The landmark publication by Ozer[22] provides a more complete discussion on the coefficient of determination.

Pearson Correlation Versus Linear Regression

Due to similarities between a Pearson correlation and a linear regression, researchers sometimes are uncertain as to which test to use. Both techniques have a close mathematical relationship, but distinct purposes and assumptions.

Linear regression will be covered in a subsequent tutorial in this series. Briefly, simple linear regression has only 1 independent variable (x) and 1 dependent variable (y). It fits a line through the data points of the scatter plot, which allows estimates of y values from x values.[23] However, the regression line

itself provides no information about how strongly the variables are related. In contrast, a correlation does not fit such a line and does not allow such estimations, but it describes the strength of the relationship. The choice of a correlation or a linear regression thus depends on the research objective: strength of relationship versus estimation of y values from x values.

However, additional factors should be considered. In a Pearson correlation analysis, both variables are assumed to be normally distributed. The observed values of these variables are subject to natural random variation. In contrast, in linear regression, the values of the independent variable (x) are considered known constants.[23] Therefore, a Pearson correlation analysis is conventionally applied when both variables are observed, while a linear regression is generally, but not exclusively, used when fixed values of the independent variable (x) are chosen by the investigators in an experimental protocol.

To illustrate the difference, in the study by Nishimura et al,[1] the infused volume and the amount of leakage are observed variables. However, had the investigators chosen different infusion regimes to which they assigned patients (eg, 500, 1000, 1500, and 2000 mL), the independent variable would no longer be random, and a Pearson correlation analysis would have been inappropriate.

SPEARMAN RANK CORRELATION

In the previously mentioned study by Kim et al,[2] the scatter plot of OGFR expression and cell growth does not seem compatible with a bivariate normal distribution, and the relationship appears to be monotonic but nonlinear. Spearman rank correlation can be used for an analysis of the association between such data.[14]

Basically, a Spearman coefficient is a Pearson correlation coefficient calculated with the ranks of the values of each of the 2 variables instead of their actual values (Figure 2).[13] A Spearman coefficient is commonly abbreviated as ρ (rho) or "r_s." Because ordinal data can also be ranked, use of a Spearman coefficient is not restricted to continuous variables. By using ranks, the coefficient quantifies strictly monotonic relationships between 2 variables (ranking of the data converts a nonlinear strictly monotonic relationship to a linear relationship, see Figure 2). Moreover, this property makes a Spearman coefficient relatively robust against outliers (Figure 3).

Analogous to Pearson coefficient, a Spearman coefficient also ranges from −1 to +1. It can be interpreted as describing anything between no association ($\rho = 0$) to a perfect monotonic relationship ($\rho = -1$ or $+1$). Analogous considerations as described above for a Pearson correlation also apply to the interpretation of confidence intervals and P values for a Spearman coefficient.

PITFALLS AND MISINTERPRETATIONS

Correlations are frequently misunderstood and misused.[4,5] It is important to note that an observed correlation (ie, association) does not assure that the relationship between 2 variables is causal. Ice cream sales increase as the temperature increases during summer, and so does the sales of fans. Hence, fan sales tend to increase along with ice cream sales, but this positive correlation does not justify the conclusion that eating ice cream causes people to buy fans. While the fallacy is easily detected in this example, it might be tempting to conclude that infusion of large amounts of crystalloid fluid causes fluid leakage into the interstitium. While it is certainly possible that a causal relationship exists, we would not be justified to conclude this based on a correlation analysis. The distinction between association and causation is discussed in detail in a previous tutorial.[24]

Correlations also do not describe the strength of agreement between 2 variables (eg, the agreement between the readings from 2 measurement devices, diagnostic tests, or observers/raters).[25] Two variables can exhibit a high degree of correlation but can at the same time disagree
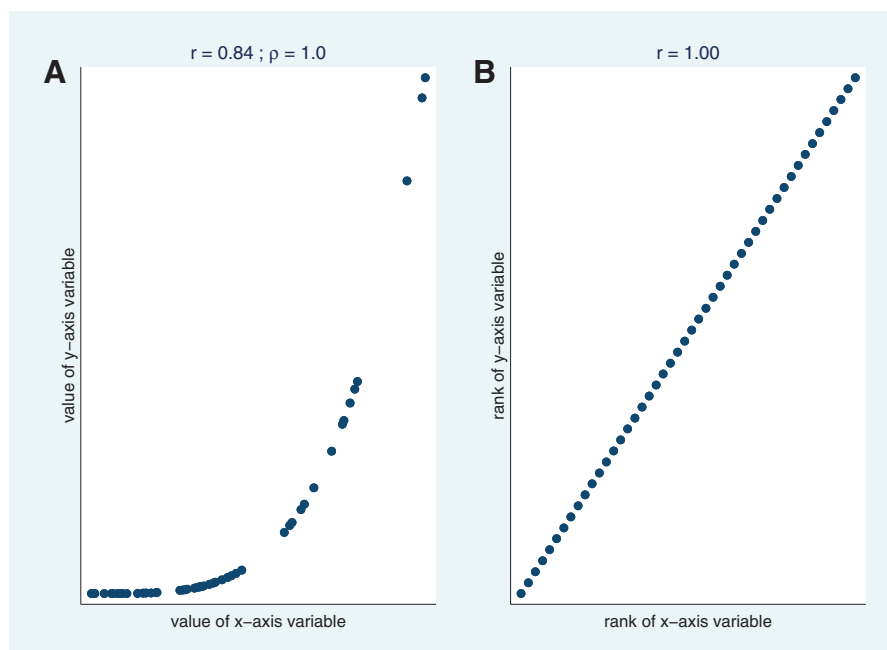

Figure 2. A, A strictly monotonic curve with a Pearson correlation coefficient (r) of +0.84. Also in the left-side flat part, the curve is continuously slightly increasing. After ranking the values of both variables from lowest to highest, the ranks show a perfect linear relationship (B). Spearman rank correlation is Pearson correlation calculated with the data ranks instead of their actual values. Hence, Spearman coefficient (ρ) of +1.0 in A corresponds to Pearson correlation of +1.0 in B.

Figure 3. Constructed examples to illustrate that the relationship between data should also be assessed by visual inspection of plots, rather than relying only on correlation coefficients. A, A correlation coefficient close to 0 does not necessarily mean that the *x* axis and the *y* axis variable are not related. In fact, the graph suggests a strong quadratic relationship. B–D, Pearson correlation coefficient (*r*) is +0.84, just as in Figure 2A, yet the actual relationship between the data is quite different in each panel. Note that the outlier in B has a relevant influence on Pearson coefficient as excluding this extreme value yields a perfect linear relationship, whereas it has virtually no influence on Spearman coefficient (with the outlier, ρ is very close to +1, without the outlier, it is exactly +1). C, The relationship is neither linear nor monotonic, and neither Pearson nor Spearman coefficient captures the sinusoid relationship. D, Sampled from a bivariate normal distribution.
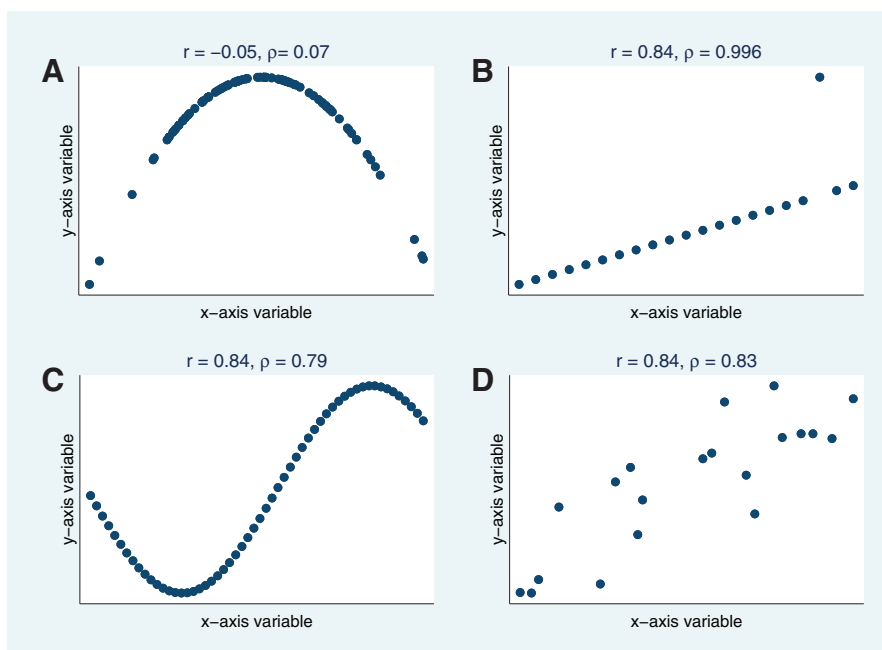
substantially, for example if 1 technique measures consistently higher than the other.

Another misconception is that a correlation coefficient close to zero demonstrates that the variables are not related. Correlation should be used to describe a linear or monotonic association, but this does not exclude that researchers might deliberately or inadvertently misuse the correlation coefficient for relationships that are not adequately characterized by correlation analysis (eg, quadratic relationship as in Figure 3A). Very different relationships can result in similar correlation coefficients (Figures 2A and 3B–D). Therefore, researchers are well advised not only to rely on the correlation coefficient but also to plot the data for a visual inspection of the relationship.[26] Graphing data are generally a good first step before performing any numerical analysis.

CONCLUSIONS

Correlation coefficients describe the strength and direction of an association between variables. A Pearson correlation is a measure of a linear association between 2 normally distributed random variables. A Spearman rank correlation describes the monotonic relationship between 2 variables. It is (1) useful for nonnormally distributed continuous data, (2) can be used for ordinal data, and (3) is relatively robust to outliers. Hypothesis tests are used to test the null hypothesis of no correlation, and confidence intervals provide a range of plausible values of the estimate.

Researchers should avoid inferring causation from correlation, and correlation is unsuited for analyses of agreement. Visual inspection of scatter plots is always advisable, as correlation fails to adequately describe nonlinear or nonmonotonic relationships, and different relationships between variables can result in similar correlation coefficients. ◼

DISCLOSURES
Name: Patrick Schober, MD, PhD, MMedStat.
Contribution: This author helped write and revise the article.
Name: Christa Boer, PhD, MSc.
Contribution: This author helped write and revise the article.
Name: Lothar A. Schwarte, MD, PhD, MBA.
Contribution: This author helped write and revise the article.
This manuscript was handled by: Thomas R. Vetter, MD, MPH.
Acting EIC on final acceptance: Thomas R. Vetter, MD, MPH.

REFERENCES
1. Nishimura A, Tabuchi Y, Kikuchi M, Masuda R, Goto K, Iijima T. The amount of fluid given during surgery that leaks into the interstitium correlates with infused fluid volume and varies widely between patients. *Anesth Analg*. 2016;123:925–932.
2. Kim JY, Ahn HJ, Kim JK, Kim J, Lee SH, Chae HB. Morphine suppresses lung cancer cell proliferation through the interaction with opioid growth factor receptor: an in vitro and human lung tissue study. *Anesth Analg*. 2016;123:1429–1436.
3. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24:69–71.
4. Porter AM. Misuse of correlation and regression in three medical journals. *J R Soc Med*. 1999;92:123–128.
5. Schober P, Bossers SM, Dong PV, Boer C, Schwarte LA. What do anesthesiologists know about p values, confidence intervals, and correlations: a pilot survey. *Anesthesiol Res Pract*. 2017;2017:4201289.
6. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. *Am Stat*. 1988;42:59–66.
7. Wackerly DD, Mendenhall III W, Scheaffer RL. Multivariate probability distributions. In: *Mathematical Statistics with Applications*. 7th ed. Belmont, CA: Brooks/Cole, 2008:223–295.
8. Nefzger MD, Drasgow J. The needless assumption of normality in Pearson's r. *Am Psychol*. 1957;12:623–625.
9. Binder A. Considerations of the place of assumptions in correlational analysis. *Am Psychol*. 1959;14:504–501.
10. Kowalski CJ. On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *J R Stat Soc*. 1972;21:1–12.
11. Bland JM, Altman DG. Correlation, regression, and repeated data. *BMJ*. 1994;308:896.
12. Vetter TR. Fundamentals of research data and variables: the devil is in the details. *Anesth Analg*. 2017;125:1375–1380.
13. Kutner MH, Nachtsheim CJ, Neter J, Li W. Inferences in regression and correlation analysis. In: *Applied Linear Statistical Models (International Edition)*. 5th ed. Singapore: McGraw-Hill/Irvin, 2005:40–99.

14. Caruso JC, Cliff N. Empirical size, coverage, and power of confidence intervals for Spearman's Rho. *Educ Psychol Meas*. 1997;57:637–654.

15. Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. *Korean J Anesthesiol*. 2017;70:407–411.

16. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: part 1–correlation within subjects. *BMJ*. 1995;310:446.

17. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: part 2–correlation between subjects. *BMJ*. 1995;310:633.

18. Overholser BR, Sowinski KM. Biostatistics primer: part 2. *Nutr Clin Pract*. 2008;23:76–84.

19. Bland JM, Altman DG. Correlation in restricted ranges of data. *BMJ*. 2011;342:d556.

20. Schober P, Bossers SM, Schwarte LA. Statistical significance versus clinical importance of observed effect sizes: what do *P* values and confidence intervals really represent? *Anesth Analg*. 2018;126:1068–1072.

21. Mascha EJ, Vetter TR. Significance, errors, power, and sample size: the blocking and tackling of statistics. *Anesth Analg*. 2018;126:691–698.

22. Ozer DJ. Correlation and the coefficient of determination. *Psychol Bull*. 1985;97:307–315.

23. Kutner MH, Nachtsheim CJ, Neter J, Li W. Simple linear regression. In: *Applied Linear Statistical Models (International Edition)*. 5th ed. Singapore: McGraw-Hill/Irvin, 2005:2–39.

24. Vetter TR. Magic mirror, on the wall-which is the right study design of them all? Part II. *Anesth Analg*. 2017;125:328–332.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.

26. Anscombe FJ. Graphs in statistical analyses. *Am Stat*. 1973;29:17–21.