

READERS' TOOLBOX

Understanding Research Methods

Determining Associations and Estimating Effects with Regression Models in Clinical Anesthesia

Kazuyoshi Aoyama, M.D., Ph.D., Ruxandra Pinto, Ph.D., Joel G. Ray, M.D., M.Sc., Andrea Hill, Ph.D., Damon C. Scales, M.D., Ph.D., Robert A. Fowler, M.D.C.M., M.S. (ANESTHESIOLOGY 2020; 133:500–9)

Determining Associations and Estimating Effects with Regression Models in Clinical Anesthesia

Estimating how patient factors influence clinical outcome while addressing bias and confounding

To infer associations or establish a likelihood of causation:

- Address the many forms of bias
- Consider and limit confounders

To determine associations between variables, one typically uses

- Linear regression for continuous variables
- Logistic regression for binary variables

Building a regression model has important steps:

- Appropriate variables
- At least 5 to 10 events per variable
- No overly influential data points
- Independent observations

Regression models can estimate the impact of patient factors on clinical outcomes of interest. Researchers must consider bias, confounding, and key assumptions when selecting and assessing a model.

Aoyama, et al. ANESTHESIOLOGY, February 2020.

ANESTHESIOLOGY
Trusted Evidence: Discovery to Practice®

Copyright © 2020 Wolters Kluwer Health. All rights reserved. Published by Lippincott, Williams & Wilkins.

SUMMARY

There are an increasing number of “big data” studies in anesthesia that seek to answer clinical questions by observing the care and outcomes of many patients across a variety of care settings. This Readers’ Toolbox will explain how to estimate the influence of patient factors on clinical outcome, addressing *bias* and *confounding*. One approach to limit the influence of confounding is to perform a clinical trial. When such a trial is infeasible, observational studies using robust regression techniques may be able to advance knowledge. *Logistic* regression is used when the outcome is binary (e.g., intracranial hemorrhage: yes or no), by modeling the natural log for the odds of an outcome. Because outcomes are influenced by many factors, we commonly use *multivariable logistic regression* to estimate the unique influence of each factor. From this tutorial, one should acquire a clearer understanding of how to perform and assess *multivariable logistic regression*.

(ANESTHESIOLOGY 2020; 133:500–9)

Clinical Scenario

You have just received a STAT call for anesthesia to your emergency room. You are asked to administer general anesthesia to a 33-year-old pregnant woman who requires an urgent cesarean section delivery after transfer from a remote hospital. The patient’s level of consciousness decreases, and there is concurrent fetal distress. The woman has a history of obesity (body mass index, 32.5 kg/m²), pregnancy-induced hypertension (ambulatory blood pressure, 146/92 mmHg), and gestational diabetes mellitus (HbA1c, 48 mmol/mol). You perform a rapid sequence induction, endotracheally intubate her, and stabilize her systems while the obstetric surgical team performs a cesarean section. A postoperative computerized tomography scan reveals intracranial hemorrhage. Both mother and neonate are transferred to intensive care units for further management.

Rationale for Investigating Factors Associated with Outcomes

This patient’s experience and both maternal and neonatal outcomes motivates the exploration of a clinical question:

Image: Jorge A. Galvez, M.D., M.B.I., Terri Navarette, and Allan F. Simpaio, M.D., M.B.I.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal’s Web site (www.anesthesiology.org).

Submitted for publication July 12, 2018. Accepted for publication May 19, 2020. Published online first on July 1, 2020. From the Department of Anesthesia and Pain Medicine, Hospital for Sick Children, Toronto, Ontario, Canada (K.A.); the Program in Child Health Evaluative Sciences, Peter Gilgan Centre for Research and Learning, Hospital for Sick Children Research Institute, Toronto, Ontario, Canada (K.A.); the Department of Critical Care Medicine (R.P., A.H., D.C.S., R.A.F.) and the Sunnybrook Research Institute (K.A., R.P., A.H., D.C.S., R.A.F.), Sunnybrook Health Science Center, Toronto, Ontario, Canada; the Keenan Research Centre of the Li Ka Shing Knowledge Institute (J.G.R.) and the Department of Obstetrics and Gynecology, St. Michael’s Hospital, Toronto, Ontario, Canada (J.G.R.); and the Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada (K.A., J.G.R., D.C.S., R.A.F.).

Copyright © 2020, the American Society of Anesthesiologists, Inc. All Rights Reserved. Anesthesiology 2020; 133:500–9. DOI: 10.1097/ALN.0000000000003425

Box 1. What to Look for in Research Using This Method

When assessing associations and effects derived from regression analyses in clinical research articles, ask the following questions:

- How reliable are the associations?
- Are the associations generalizable to other populations?
- Does exposure or intervention precede outcomes?
- Is there a dose–response or an exposure–response relationship?
- Are the associations relevant based on current scientific evidence?
- How are the potential biases, including confounding, considered and adjusted?
- Are the variables of the regression models selected based on the study hypothesis?
- How is the model's fit assessed?
- Are the number of variables in the regression models appropriate?
- Is there any interaction among the variables of the regression models?
- Is there any hierarchical structure related to the associations?
- How meaningful and confident are the findings of the regression models?
- Does the study adhere to consensus reporting guidelines?

what factors are associated with life-threatening conditions of pregnant or postpartum women? For the highlighted patient, did she have preexisting comorbidities contributing to her illness? What factors other than patient characteristics were influential and potentially led to the delay of initial assessment and treatment, for example, the distance between the hospital and her residence? Did the hospital have the resources to manage her deteriorating condition? Did resuscitation choices influence outcomes? Ultimately, one wants to understand the potentially preventable or reversible factors, and how one could have intervened upon them to improve maternal and perinatal neonatal outcomes.

In this Readers' Toolbox, we will explore the mechanisms and techniques to estimate the potential influence of a myriad different "systems" factors (*i.e.*, patient, practitioner, hospital, and healthcare region factors) on outcomes of our patients (fig. 1). Drawing on the clinical scenario presented, we will address techniques to estimate mainly the influence of *patient factors* on outcome. Although we will briefly explain how to consider *multiple levels of influence*, including *variability* in practice among clinicians (*e.g.*, differences in primary training, volume, or experience in their own practice) or hospitals, a detailed examination of this is beyond the scope of this Readers' Toolbox. From this Readers' Toolbox, a clearer understanding will be provided of how to perform and assess a multivariable regression analyses to advance understanding of common clinical questions (box 1).

In clinical and health services research, investigators ideally wish to appreciate, measure, and appropriately account for all factors potentially related to clinical outcomes. As such, the adaptation of these associations to clinical anesthesia may enable prediction of clinical outcomes of our patients with more accuracy and certainty.

Determining Associations and the Challenges in Estimating Causation

Although some clinical events may be caused by "bad luck" (that is, random chance), certain patient, hospital, or process of care factors can influence whether an event happens or not. Although some such influential factors (called "predictors," "exposures," or "independent variables") may be associated with an event ("outcome" or "dependent variable"), the association is often insufficient to attribute to a causal relation between the two.^{1,2} In health services research, one common goal is to scientifically explore the potential of certain exposure variables to "cause" events of interest. For this purpose, a number of causation concepts exist to help explore the question of causation (Supplemental Digital Content, eText 1, <http://links.lww.com/ALN/C408>).¹⁻⁴

Several factors must be considered to even attempt to establish an association rather than causation. First, one must address the many potential forms of bias, defined as any systematic error occurring in the design or in the conduct of clinical research (box 2). *Information bias* is due to a systematic error in the assessment of a variable. *Selection bias* implies that the study sample is not representative of the population intended to be analyzed^{5,6} and can be used to describe the selection of certain patients who subsequently receive treatments based on characteristics that differ from patients in the general population. *Sampling bias* is systematic error caused by a nonrandom sampling of a population that sometimes occurs if a portion of the entire population is chosen as potential participants for a study. *Indication bias* can occur when the certain indication for choosing a particular intervention (*e.g.*, status of health insurance) also impacts the outcome (*e.g.*, tendency of healthier outcome). Last, other variables may be related to the exposure of interest that have their own unique association(s) with the

Box 2. Bias in Clinical Research

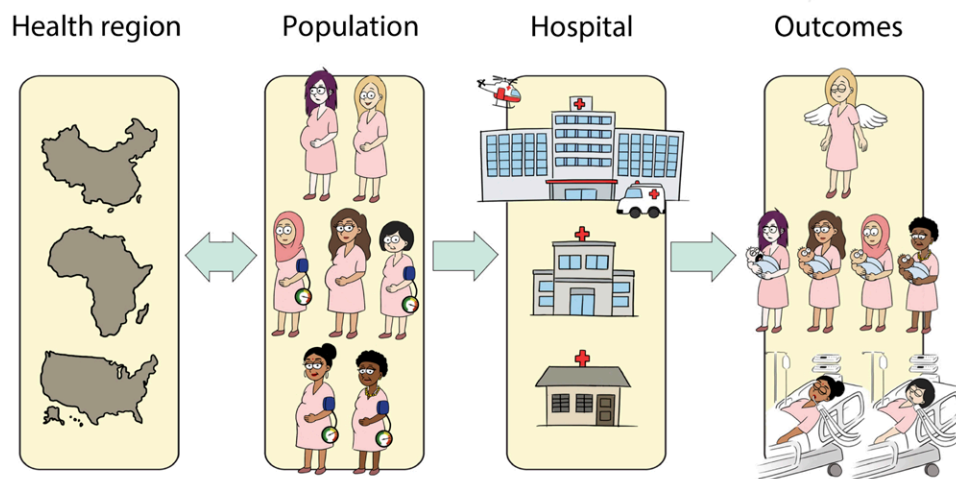
Bias: a systematic error in the design, selection of patients or recruitment (*e.g.*, selection bias), data collection (*e.g.*, measurement bias), or analysis resulting in an incorrect estimation of the true effect of one variable upon another

Confounding: a variable that influences both the dependent variable and independent variable *causing a spurious or altered association*

Effect modification: occurs when a variable *modifies an actual association* of an independent variable on a dependent variable



How do I determine the influences of multiple factors, at multiple levels, that may influence outcome of interest?



This is a conceptual framework that helps us to visualize associations between the outcome of interest and factors from different levels.



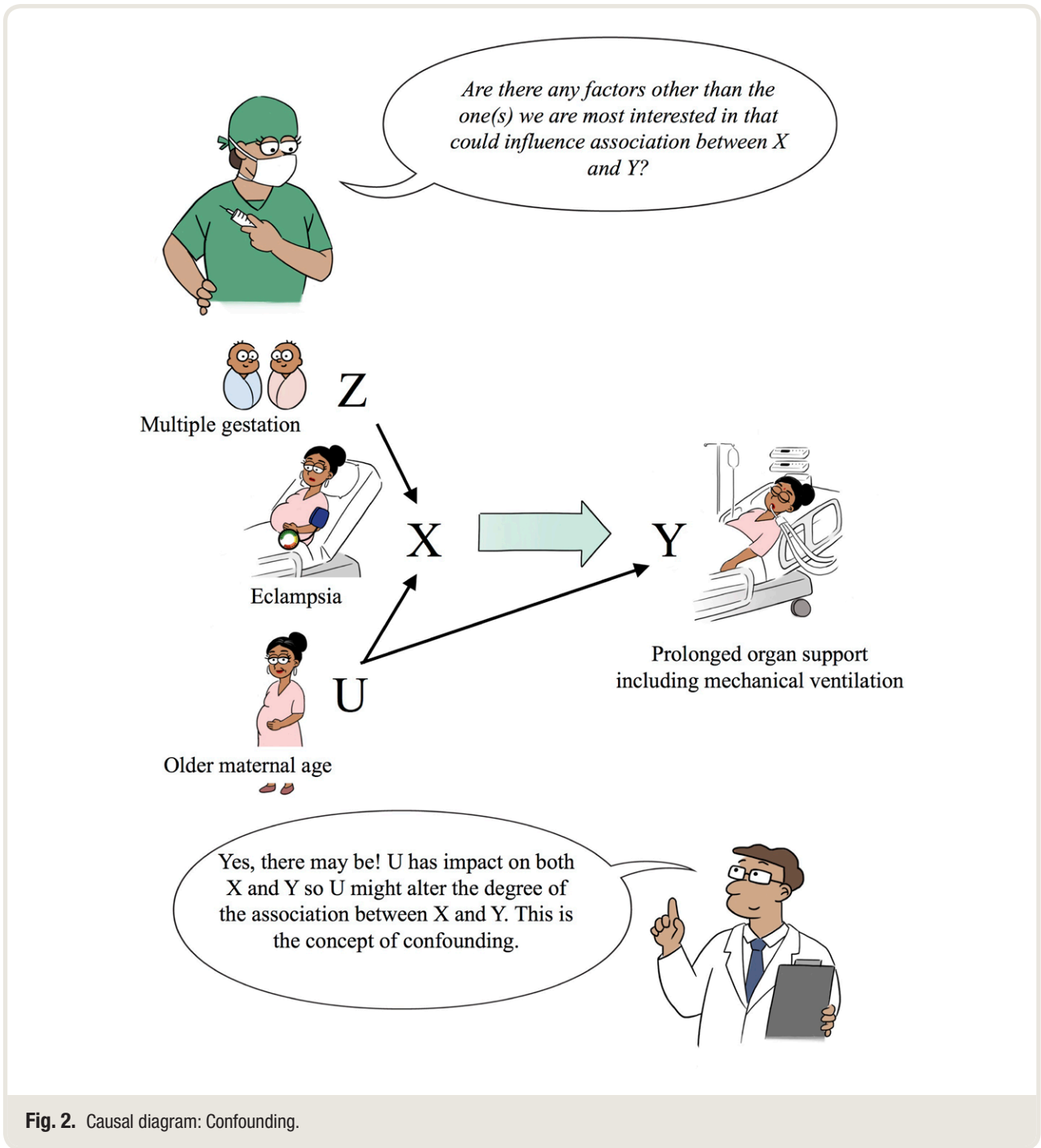
Downloaded from <http://asa2.silverchair.com/anesthesiology/article-pdf/133/3/500/514826/20200900.0-00011.pdf> by guest on 17 April 2024

Fig. 1. Conceptual framework of multilevel factors.

outcome of interest. *Confounding* occurs when two factors, themselves related, are associated with the same outcome or effect, and the measure of association of one variable with the outcome is distorted by this other, confounding variable (fig. 2).⁷

One approach to limit the influence of confounding is to use an experimental design such as a randomized clinical trial. In such a design, individuals are randomly allocated to particular groups of sufficient size so that such potential

confounding variables are also randomly, but evenly, distributed among the groups. Some of these confounding variables will be prespecified and measured, whereas others may remain unknown and unmeasured. Regardless of measured or unmeasured confounding variables, when between- or among-group comparisons are made, the influence of these potential confounding factors “cancel” each other out with only the allocation to treatment being different between groups. Aside from minimizing the chance of bias



Downloaded from <http://asa2.silverchair.com/anesthesiology/article-pdf/133/3/500/514826/20200900.0-00011.pdf> by guest on 17 April 2024

Fig. 2. Causal diagram: Confounding.

and confounding, there are related constructs that may help assess associations (box 2 and Supplemental Digital Content, eText 1, <http://links.lww.com/ALN/C408>).^{2,4,6}

To provide a standard framework for designing and reporting observational studies, there are a number of guidelines for specific study designs to help authors and readers understand the information required to rationalize and justify putative associations.⁸⁻¹⁰ These guidelines include Strengthening the Reporting of Observational

Studies in Epidemiology (STROBE) statement guidelines for reporting observational studies, the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement, and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD), to name a few.⁸⁻¹⁰ Note that the development of all analysis plans is encouraged before accessing the data set(s) to reduce the potential impact of biases.¹¹ Box 3 summarizes other approaches to address

potential confounding, such as matching and stratification.¹² In this Readers' Toolbox, we will focus on tackling the issue of establishing associations and the challenges of confounding at the patient level with a focus on the most commonly used approach called *regression*.¹³ Note that the regression described here is not the foundation of establishing causality but is rather one approach to assess the association between an exposure variable and an outcome of interest.

Establishing Associations: Regression

Regression analyses assist in estimating the magnitude and direction of the association between variables and clinical or statistical significance. A common first approach is to examine the relationship between an independent variable (or exposure variable) and a dependent variable (or outcome variable). This is called a univariate regression equation, which refers to when there is a single exposure. The association of many independent variables with the dependent variable of interest can be assessed simultaneously in a "multivariable" or multiple regression equation. We employ regression to evaluate the *association* between an exposure and an outcome of interest, accounting for the influence of (measured) confounding (box 3). Note that this is different from the development of *prediction models* using regression where the aim is to identify all of the factors that best predict whether an outcome will occur.

Regression can also be classified based on the nature of the outcome of interest. When the outcome is continuous (e.g., blood pressure, measured in mm Hg), *linear* regression is typically used. This is where the regression assumes a linear relation between the outcome and the exposure variable. When the outcome is binary (e.g., intracranial hemorrhage or not), *logistic* regression is typically used by modeling the natural log for the odds of an outcome. Although *linear* regression can be used for a binary outcome in certain situations,¹⁴ a detailed discussion of that approach is beyond the scope of this Readers' Toolbox. When the outcome is "counts," Poisson regression may be employed. For cases in which the outcome is continuous and a linear relationship may not be appropriate to use, other types of regression that take into account the structure of the data may be more appropriate. "Robust" regression is when there is concern about extreme or outlying data points, and non-parametric regression is when the distribution of outcomes is not normal, for example in the case where the distribution is not on a bell curve.^{13,15} Ideally the type of regression model is selected in direct response to the nature of the relationship between the outcome and the exposure and distribution of the raw data that will be modeled.¹³

Linear Regression

Suppose one wishes to explore an association between pregnancy week (that is, gestational age of the fetus) as the "exposure" variable and maternal degree of proteinuria as

the "outcome" variable, because the pregnancy-induced hypertension of the patient in the case scenario often causes proteinuria. Because the outcome can be considered a continuous variable (measured in grams/day), a simple linear regression might be applied to obtain an estimate of the magnitude of association between gestational age and proteinuria.

Now suppose one also wants to include the influence of obesity (measured as body mass index in kg/m²) because there may also be a dose-response relationship between body mass index and protein loss in the urine. In this case, a *multivariable* regression analysis might be used to include body mass index together with gestational age. In this regression model, you would assume a straight-line (*linear*) relationship between one continuous exposure variable (i.e., gestational age) and the continuous outcome measure (i.e., proteinuria), keeping the other exposure variable constant (i.e., body mass index; fig. 1).

Univariate Logistic Regression

Suppose one is interested in the outcome of intracranial hemorrhage during pregnancy, a binary outcome rather

Box 3. Methods of Addressing Potential Confounding

In Design

- **Randomization:** In randomized trials, the distribution of measured and unmeasured confounding should be similarly distributed between groups if the sample size is sufficiently large.
- **Matching:** In observational studies, forcing comparison groups to have similar distributions of critical characters can achieve the adjustment of measured confounding. Inevitably, matching reduces the sufficient sample size of comparisons.

In Analysis

- **Stratification:** Stratifying data based on measured confounders can minimize the effect of measured confounding, but stratification is challenging when multiple confounders exist (which is usually the case).
- **Regression:** Regression analysis can incorporate multiple measured potential confounders to attempt to identify the independent influence of a variable on the outcome of interest. Regression models vary in the degree to which they fully explain the outcome of interest.
- **Propensity scores:** Somewhat similar to matching, propensity scores attempt to address confounding by accounting for the covariates that predict receiving the treatment and creating more similar groups of patients who differ only on the exposure of interest. The propensity score is the probability that a patient would receive the treatment – a way to simulate the randomization of patients to two groups of exposures.⁹ Multivariable regression models can compute the probability, where the treatment is the outcome, and the covariates are predictors.

than a continuous outcome, and having this life-threatening condition or not, in relation to preexisting systolic blood pressure (in mm Hg). Other factors might also influence the outcome, such as preexisting diabetes (measured by HbA1c) and obesity (in kg/m²). In this case, the relationship between the variables is not linear because the binary outcome does not vary the “degrees” of the issue along a straight line, so a logistic regression model is more appropriate. A *univariate logistic regression* includes one exposure variable only, whereas a *multivariable logistic regression* includes multiple exposure variables. We will explore both of these in the context of the example presented above. The patient in our clinical scenario may have a risk of developing an intracranial hemorrhage because of her pregnancy-induced hypertension. As such, there may be an association between intracranial hemorrhage (here a binary outcome of “yes” or “no”) and systolic blood pressure (a continuous independent variable in mm Hg). The intracranial hemorrhage as an outcome can be coded as 0 (“no”) or 1 (“yes”). However, systolic blood pressure is a continuous variable, which makes developing an association where the outcome is 0 or 1 a challenge, because there is not a straight-line relationship between this variable and the binary outcome. Plotting 0/1 values on the y axis and systolic blood pressure on the x axis and fitting it with a regression line (Supplemental Digital Content, fig. 2, <http://links.lww.com/ALN/C408>) shows that a patient with systolic blood pressure of less than 80 mmHg will have a predicted probability of intracranial hemorrhage of less than 0. This situation is clinically very unlikely. Further, the patient in the clinical scenario with a systolic blood pressure of 146 mmHg will have a predicted probability of intracranial hemorrhage greater than 1, and although clinically possible, it is likely that not all patients have this probability.

Because this analysis provides situations that are not realistic clinically, one can use a logistic regression model and take the *log odds* of the outcome. The “odds” of the event happening (intracranial hemorrhage in this case) is the probability of the event divided by the probability of no event ($p / [1 - p]$). By applying the natural log to the odds, the probability is transformed from (0,1) to $(-\infty, +\infty)$, which is the fitted red curve for the logistic regression with systolic blood pressure in the model (Supplemental Digital Content, fig. 2, <http://links.lww.com/ALN/C408>). Although logistic regression is the most common type of regression relationship in the medical literature, there are other models that might be a more accurate reflection of the real relationship, for example, probit regression and linear probability modeling.^{14,16} In addition, when considering marginal effects, which are defined as how the predicted probability of a binary outcome changes with a change in a risk factor, selection of linear *versus* logistic regression matters little.¹⁷ Logistic regression is more popular in part because the coefficients can be interpreted in terms of odds ratios and can imply the probability outside of 0 to 1.

Multivariable Logistic Regression

Multivariable logistic regression is used when more than one factor, called *covariates*, may have an association with the outcome of interest in addition to the main exposure of interest. In the clinical scenario presented, this would be the case when it is believed that HbA1c of diabetes and body mass index of obesity (measured as BMI in kg/m²) should be included because of their potential dose–response relation with intracranial hemorrhage and hence their importance as *covariates* for the exposure variable of interest, the systolic blood pressure.

In logistic regression, coefficients have a simple interpretation. For example, logistic regression may be used to answer the questions about the odds ratio associated with a unit increase in systolic blood pressure leading to intracranial hemorrhage. By exponentiating both sides of association above while controlling for the other variables in the model, the odds ratio associated with systolic blood pressure can be computed. With an odds ratio of 1.20, each additional unit (for example, 10 mmHg) of systolic blood pressure increases the odds of intracranial hemorrhage by 20%, when the other variables in the model are controlled for. Note that negative coefficients translate into odds ratios less than 1, whereas positive coefficients translate into odds ratios larger than 1. A distinction between “multivariable,” when there is more than one *independent variable* in a regression model, must be made with “multivariate,” when there is more than one *outcome* in a regression model. Importantly, odds ratios are neither probabilities nor risk ratios, but the probability of event divided by the probability of no event. Another limitation of odds ratios is that because they are conditional on the sample and the model specification, odds ratios should not be compared across different studies with different samples. More details of practical aspect of odds ratios can be found elsewhere.¹⁸

How to Select Variables for a Regression Model

This is a topic of great debate. The most important and primary principle of variable selection is that variables chosen to examine for associations should be based on the research question and a sense of what is clinically or biologically plausible, based upon existing literature or clinical experience. In terms of selecting the maximum number of variables in a model, one rule of thumb is that there should be 5 to 10 outcome events for each variable added to the model. If 50 patients experienced an intracranial hemorrhage in the clinical scenario data set, then one might be justified in exploring associations of 5 to 10 variables with that outcome. One of the downsides of having too many variables in a model is that the model becomes too complex and too tailored to an individual data set, both to the individual data points and to the noise in the data set. Although the goal is often to have results

representative or generalizable to the population, when the model is “overfitted” to a specific data set, this will usually not be the case.

How to Determine the Importance and Confidence of Variables Included in a Regression Model

One mechanism for this is to decide which and how many variables are believed to be associated with the outcome. In a multivariable regression equation, statistical software aids in learning the coefficients and *P* values.

Another commonly used method is to perform a *stepwise regression*, for example, when unsure about which variables to investigate.¹⁹ This refers to an *automated* selection algorithm performed by statistical software according to certain preset rules. Variables are entered and/or removed from a regression model, sequentially and one by one on the basis of their relative *statistical* significance. There are three common types of stepwise regression: (1) forward selection, (2) backward selection, and (3) stepwise selection (which is a combination of forward and backward selection). A major challenge to this automated approach is that the clinical context and relevance of the relationships are not known to the statistical software, so it may be misleading or come up with putative and statistically significant associations that do not make clinical sense. For example, clinically relevant factors may be dropped from the model if the statistical significance is below a prespecified threshold, which may impact the most clinically valid estimate of the association between an exposure and outcome. Given the large data sets available, hypothesis-driven variable selection is often preferred.

How to Assess Relationships between the Exposure Variables in a Regression Model

Sometimes two variables that one might think are “different” and decide to include in a model are actually very similar (for example, body mass index and body weight). When one exposure variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy, this is referred to as “collinearity” (*i.e.*, one variable is collinear with the other). When collinearity exists to a large degree, it is hard for statistical software to know how to best assign coefficients to each variable. The amount of collinearity between all pairs of variables in a collinearity matrix can be assessed. If two variables are collinear with values greater than 0.5 to 0.8, one might consider choosing only one of the pair for inclusion in the regression equation.

Although collinearity means correlation among variables in a regression model, if two variables *interact*, the relationship between each of the interacting variables and the outcome variable depends on the value of the third interacting variable.²⁰ This relationship makes it more difficult to predict the consequences of changing the value of a variable. For example, in the clinical scenario, there are

two variables of hypertension and obesity that may increase the risk of intracranial hemorrhage. One may be concerned about an *interaction* between hypertension and obesity, and if that is the case, it can be determined whether the measure of association from one variable is influenced by another variable by creating and adding an *interaction variable* (Hypertension*Obesity) into the model. If the coefficient of this interaction variable is clinically important and statistically significant, it can be concluded that the relationship between intracranial hemorrhage and hypertension may depend on the presence of obesity (or *vice versa*).

Key Questions to Consider When Developing or Interpreting Regression Analyses

Logistic regression needs to be developed based on four key assumptions, which are slightly different from the assumptions used for linear regression (Supplemental Digital Content, eText 2, <http://links.lww.com/ALN/C408>): (1) the model adequately fits the data – assessed by the Hosmer–Lemeshow statistical goodness-of-fit test and C-statistics; with an appropriate fit indicating that the model adequately predicts the outcome; (2) the model is not overspecified or overfitted – as a conservative rule of thumb, the number of variables should be at least 5 to 10 events per variable; otherwise the model may be less precise through producing biased estimates and under- or overestimated variance; (3) there are no overly influential observations – influential data points may artificially skew the regression line especially in small data sets; there are many different statistics *via* plots to visually identify pattern of the data; and (4) the observations need to be *independent* of each other – with the same caveats as linear regression (Supplemental Digital Content eText 2, <http://links.lww.com/ALN/C408>).^{21,22} A more detailed discussion of these and other considerations for assumptions to develop regression models is detailed well in recent reviews.^{13,21,23,24}

How to Assess Performance (Goodness-of-Fit) of a Regression Model

The coefficient of determination, or R^2 , is a goodness-of-fit measure for a linear regression model. R^2 is a statistical measure of how close the data are fitted to the regression line, interpreted as the amount of variation in the dependent variable that is explained by the independent variable and expressed as a percentage (R^2 can vary from 0 to 1). A useful guide is that an R^2 value between 0 and 0.5 indicates a poor fit, between 0.51 and 0.70 indicates a moderate fit, and between 0.71 and 1.00 indicates a good fit. Note that one could have statistically *significant* results with a poor fit, and it is important to know that statistical significance does not imply a good fit.

The C-statistic is a goodness-of-fit measure for a logistic regression model that investigates how well the model distinguishes the outcome and referred to as the *discrimination*

of a model. The C-statistic is also referred to as the area under the receiver operating characteristic curve (abbreviated as AUC). The closer the C-statistic is to 1, the better the model is at distinguishing an outcome. The set of values described above for R^2 also applies to the C-statistic. In certain circumstances of uncommon and rare outcomes (defined as those that occur less than 10% of the time) that are often faced in the field of anesthesia, the area under the precision-recall curve (abbreviated as AUPRC) reflects better discrimination than the area under the receiver operating characteristic curve.²⁵ The Hosmer–Lemeshow test is another goodness-of-fit measure for a logistic regression model that investigates how close values expected by the model are to the observed values, also known as the *calibration* of a model. When the Hosmer–Lemeshow test is not statistically significant, it indicates that the numbers of events is not significantly different from that expected by the model, demonstrating that the overall model fit is good.

How to Assess Your Confidence in the Results of a Regression Model

Regression coefficients of the exposures of interest demonstrate the relationship between an exposure and the response. In a linear regression model, the coefficient value represents the mean change in the response given per one-unit change in the exposure. As described previously, in a logistic regression, exponentiating the coefficient of the exposure provides the odds ratio associated with the exposure. Negative coefficients translate into odds ratios less than 1, whereas positive coefficients translate into odds ratios greater than 1. Regression coefficients should be accompanied by 95% CI (also written as 95% *Cis*), which provides the range of values within which the coefficient is likely to reside. Narrower CI values denote greater precision of the point estimate (*i.e.*, the coefficient or odds ratio calculated). In either linear or logistic regression, if the 95% CI of the coefficient of the exposure is across 0, the exposure is not statistically significant, at least at that level of *confidence*.

Why Clinically Meaningful Difference Is Preferred to Statistical Difference

In contemporary medical clinical research, effect sizes and CI are preferred over the use of *P* values.^{26,27} Effect sizes represent the magnitude of an effect from either an exposure or intervention, and CI values represent the certainty or precision around the magnitude of the effect.²⁶ *P* values represent the probability of the observed finding, measuring the statistical inference of the observed finding based on the study hypothesis. *P* values indicate *statistical difference*, which indicates whether the findings are statistically significant, based on the prespecified value (conventionally set at 0.05). However, using effect sizes with CI, one may observe the clinical impact of the observed finding. Effect sizes are reported in various ways, including mean differences, risk

ratios, odds ratios, and variations of correlation. *Clinically meaningful difference* (also called *minimal clinically important difference*) expressed as effect size is an estimation that surpasses a minimally important cutoff point based on clinical perspective. This measure provides the patient with the smallest effect that is beneficial in the presence of an intervention or the absence of exposure.²⁷ Clinically meaningful difference is vital to interpret study findings in a clinical context. It is essential to predetermine the clinically meaningful difference when developing a study protocol so as to calculate the sample size required to show the difference and choose the appropriate candidate variables in regression models.¹¹ Sample size for a study is determined by effect sizes, significance level (conventionally set at 0.05), and power (conventionally set at 0.8), with power being a probability of finding a real effect.

When evaluating regression models, the following questions are important to consider in the interpretation of the analyses: (1) Were important confounders adjusted in the study design and the models? (2) Was the sample size of the study prespecified based on effect sizes, significance level, and power? (3) Was the number of covariates in the model appropriate in light of the number of events in the study? (4) Other than statistical difference, what is the clinically meaningful difference (*i.e.*, magnitude) of the findings, and are the findings large enough to change practice? (5) What is the certainty (*i.e.*, precision) of the findings? (6) If there is no significant statistical difference, is this truly due to no effect or due to a lack of precision? (7) Were sensitivity analyses prospectively defined and robust enough to assure the validity of the findings?

What to Do When There Are Multiple “Levels” of Exposures That May Influence Outcomes

When data collection is longitudinal in nature and contains multiple, repeated observations per participant over time, it is highly likely that these measures will not be independent of one another.^{28,29} For example, in a woman with preeclampsia in whom blood pressure is repeatedly measured over the course of her pregnancy, an elevated blood pressure observed once is likely to be high again at another time. This leads to a problem where the standard error of the estimates in a regression model can be underestimated, and the precision can be overestimated. Longitudinal studies need to account for the possible correlation between the repeated measurements within an individual as described elsewhere.³⁰

When there is a *hierarchical* structure to the data, in which observations are seen within individuals and those individuals are clustered together in groups, individuals within a cluster may be more similar to each other than those individuals in different clusters (fig. 1).^{28,29,31} For example, blood pressure among patients in a high-risk pregnancy outpatient program is likely to be higher than blood pressure readings for patients in a midwifery

Box 4. Where to Find More Information on This Topic

For more information on determining associations and estimating effects with regression analyses in clinical research:

1. Rothman KJ, Greenland S, Lash TL: *Modern Epidemiology*, 3rd edition. Riverwoods, Wolters Kluwer, 2012
2. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB: *Designing Clinical Research*, 4th edition. Philadelphia, Lippincott Williams & Wilkins, 2013
3. Hernán MA: A definition of causal effect for epidemiological research. *J Epidemiol Community Health* 2004; 58:265–71
4. Rothman KJ, Greenland S: Causation and causal inference in epidemiology. *Am J Public Health* 2005; 95:144–50
5. Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C: Bias in clinical research. *Kidney Int* 2008; 73:148–53
6. Jager KJ, Zoccali C, MacLeod A, Dekker FW: Confounding: What it is and how to deal with it. *Kidney Int* 2008; 73:256–60
7. Slinker BK, Glantz SA: Multiple linear regression: Accounting for multiple simultaneous determinants of a continuous dependent variable. *Circulation* 2008; 117:1732–7
8. Bewick V, Cheek L, Ball J: Statistics review 14: Logistic regression. *Crit Care* 2005; 9:112–8
9. Austin PC, Tu JV, Alter DA: Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: Should we be analyzing cardiovascular outcomes data differently? *Am Heart J* 2003; 145:27–35
10. Merlo J, Chaix B, Yang M, Lynch J, Råstam L: A brief conceptual tutorial of multilevel analysis in social epidemiology: Linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005; 59:443–9

practice. This effect can lead to an underestimation of the influence of correlated error in a regression model and an overestimation of the association of an exposure variable with the outcome of interest.

Individual patients with pregnancy-induced hypertension have their own individual characteristics (*e.g.*, weight, age, parity) that influence an outcome, but influential factors may also exist at the hospital level (*e.g.*, a blood pressure clinic, a specific approach to blood pressure control, or a high-risk obstetrician). Together, different variables are structured within a more complex hierarchy that might better represent real-world care.

Ignoring issues of nonindependence may result in an incorrect estimation of the magnitude of a relation between a list of variables and an outcome. The issues may be due to temporally related data or data that are clustered (hierarchical).³² Fixed effects analysis is one way to analyze temporally related (repeated measures) data to account for unmeasured time-invariant factors.³³ The challenge of correcting regression modeling to account for such a structure (*i.e.*, use of mixed models) is described elsewhere.^{29,34}

Conclusions

This Readers' Toolbox has provided explanations of how to estimate associations of patient factors with clinical outcome of interest, considering *bias* and *confounding*, as well as how to determine and quantify the degree of associations using common regression models. The importance of choosing appropriate variables for a regression model is highlighted, underscoring the need to ensure that the baseline assumptions of a selected model are correct and to assess model performance and your confidence in the results as derived from regression analyses (box 4).

Acknowledgments

The authors thank and acknowledge Dr. Brian P. Kavanagh (Anesthesia and Critical Care Medicine, Hospital for Sick Children, University of Toronto) for his contribution to conceiving this article.

Research Support

Supported by operating grant No. 342397 from the Canadian Institutes of Health Research (Ottawa, Ontario, Canada; to Drs. Aoyama, Pinto, Ray, Scales, and Fowler) and by a fellowship from the Canadian Institutes of Health Research (to Dr. Aoyama).

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Aoyama: 555 University Avenue, #2211, Toronto, Ontario M5G 1X8, Canada. kazu.aoyama@utoronto.ca. ANESTHESIOLOGY's articles are made freely accessible to all readers on www.anesthesiology.org, for personal use only, 6 months from the cover date of the issue.

References

1. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA: Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol* 2002; 155:176–84
2. Hill AB: President's address: The environment and disease. 1964:295–300
3. Hernán MA: A definition of causal effect for epidemiological research. *J Epidemiol Community Health* 2004; 58:265–71
4. Rothman KJ, Greenland S: Causation and causal inference in epidemiology. *Am J Public Health* 2005; 95:144–50
5. Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C: Bias in clinical research. *Kidney Int* 2008; 73:148–53
6. Hernán MA, Hernández-Díaz S, Robins JM: A structural approach to selection bias. *Epidemiology* 2004; 15:615–25

7. Jager KJ, Zoccali C, Macleod A, Dekker FW: Confounding: What it is and how to deal with it. *Kidney Int* 2008; 73:256–60
8. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* 2007; 370:1453–7
9. Velentgas P, Dreyer N, Nourjah P, Smith S, Torchia M: Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide, Rockville, MD, Agency for Healthcare Research and Quality, 2013
10. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 2015; 162:W1–73
11. Eisenach JC, Kheterpal S, Houle TT: Reporting of observational research in ANESTHESIOLOGY: The importance of the analysis plan. *ANESTHESIOLOGY* 2016; 124:998–1000
12. Glass TA, Goodman SN, Hernán MA, Samet JM: Causal inference in public health. *Annu Rev Public Health* 2013; 34:61–75
13. Slinker BK, Glantz SA: Multiple linear regression: Accounting for multiple simultaneous determinants of a continuous dependent variable. *Circulation* 2008; 117:1732–7
14. Hellevik O: Linear versus logistic regression when the dependent variable is a dichotomy. *Qual Quant* 2009; 43:59–74
15. Chen C: Robust regression and outlier detection with the ROBUSTREG procedure. SAS Institute Inc., Cary, North Carolina, 2002
16. Sun E, Mello MM, Rishel CA, Vaughn MT, Kheterpal S, Saager L, Fleisher LA, Damrose EJ, Kadry B, Jena AB; Multicenter Perioperative Outcomes Group (MPOG): Association of overlapping surgery with perioperative outcomes. *JAMA* 2019; 321:762–72
17. Norton EC, Dowd BE, Maciejewski ML: Marginal effects: Quantifying the effect of changes in risk factors in logistic regression models. *JAMA* 2019; 321:1304–5
18. Norton EC, Dowd BE: Log odds and the interpretation of logit models. *Health Serv Res* 2018; 53:859–78
19. Austin PC, Tu JV: Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol* 2004; 57:1138–46
20. Kleinbaum DG: Epidemiologic methods: The “art” in the state of the art. *J Clin Epidemiol* 2002; 55:1196–200
21. Bewick V, Cheek L, Ball J: Statistics review 14: Logistic regression. *Crit Care* 2005; 9:112–8
22. Hosmer D, Lemeshow S, Sturdivant R: Applied Logistic Regression. New York, Wiley, 2013
23. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49:1373–9
24. Berk R, MacDonald JM: Overdispersion and Poisson regression. *J Quant Criminol* 2008; 24:269–84
25. Ozenne B, Subtil F, Maucort-Boulch D: The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 2015; 68:855–9
26. Houle TT: Importance of effect sizes for the accumulation of knowledge. *ANESTHESIOLOGY* 2007; 106:415–7
27. Angst F, Aeschlimann A, Angst J: The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017; 82:128–36
28. Austin PC, Tu JV, Alter DA: Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: Should we be analyzing cardiovascular outcomes data differently? *Am Heart J* 2003; 145:27–35
29. Austin PC, Goel V, van Walraven C: An introduction to multilevel regression models. *Can J Public Health* 2001; 92:150–4
30. Diggle PJ, Heagerty P, Liang K-Y, Zeger SL: Analysis of Longitudinal Data. Oxford, Oxford University Press, 2013
31. Aoyama K, Pinto R, Ray JG, Hill AD, Scales DC, Lapinsky SE, Hladunewich M, Seaward GR, Fowler RA: Variability in intensive care unit admission among pregnant and postpartum women in Canada: A nationwide population-based observational study. *Crit Care* 2019; 23:381
32. Dunlop D: Regression for longitudinal data: A bridge from least squares regression. *Am Stat* 1994; 48:299–303
33. Gunasekara FI, Richardson K, Carter K, Blakely T: Fixed effects analysis of repeated measures data. *Int J Epidemiol* 2014; 43:264–9
34. Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, Bruckner T, Satariano WA: To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* 2010; 21:467–74