JAMA Guide to Statistics and Methods

# Marginal Effects—Quantifying the Effect of Changes in Risk Factors in Logistic Regression Models

Edward C. Norton, PhD; Bryan E. Dowd, PhD; Matthew L. Maciejewski, PhD

**Marginal effects** can be used to express how the predicted probability of a binary outcome changes with a change in a risk factor. For example, how does 1-year mortality risk change with a 1-year increase in age or for a patient with diabetes compared with a patient without diabetes? This approach can make the results more easily understood. Marginal effects often are reported with logistic regression analyses to communicate and quantify the incremental risk associated with each factor.[1,2]

In a 2013 article in *JAMA Psychiatry*, Cummings et al[3] studied factors that predicted access to outpatient mental health facilities that accept Medicaid. Their main outcome had 3 categories, which were labeled "no access," "some access," and "good access." An ordered logistic regression model was developed and results were presented as the change in the probability of each outcome for a change in certain demographic factors.

## Use of Marginal Effects

### Why Are Marginal Effects Used?

There are several ways to express the strength of the association between a risk factor and a binary outcome from a logistic regression. One popular approach is the odds ratio (OR).[4] The odds are the ratio of the probability that an outcome occurs to the probability that the outcome does not occur. The ratio of the odds for 2 groups—the OR—is often used to quantify differences between 2 different groups; eg, treatment and control groups. Another approach is the risk ratio, which is the probability that the outcome occurs in the presence of the risk factor divided by the probability that the outcome occurs in the absence of the risk factor. Risk ratios are often easier to use in clinical practice than are ORs.[4,5]

A third alternative is the marginal effect, which is the change in the probability that the outcome occurs as the risk factor changes by 1 unit while holding all the other explanatory variables constant. When the risk factor is continuous (eg, age), the change in the probability that the outcome occurs that is associated with a 1-unit change in the risk factor has been called a *marginal* effect. When the risk factor is discrete (eg, presence or absence of diabetes), the change has been called an *incremental* effect. In this article, the term *marginal effect* represents this strength of association measure in both instances.

### What Are Marginal Effects?

Of the 3 approaches, marginal effects are the most intuitive because they are expressed as the change in the predicted probability that the outcome occurs that is associated with a 1-unit change in the risk factor. Unlike ORs, it is easier to compare marginal effects across different studies because they are less sensitive to the statistical model conditions that influence the reported values of ORs.[6] Marginal effects depend on the values of the other explanatory variables and will not be the same for all members of a group.

For example, consider a linear regression analysis predicting body weight in pounds from a person's height measured in inches. If the regression coefficient is 5, it means that a 1-in increase in height is associated with a 5-lb increase in weight. In this instance, the marginal effect of the 1-unit change in the risk factor, height, is how it changes the predicted outcome, weight in pounds. This is true in linear regressions unless the predictors included in the model includes higher powers of the risk factors (eg, age and age squared) or interactions among the explanatory variables (eg, 2 explanatory variables multiplied together). In a simple linear regression (eg, without interactions between predictors), this marginal effect is constant across all values of the risk factor. For instance, a change in height from 5 ft 8 in to 5 ft 9 in has the same predicted effect as does a change from 6 ft 3 in to 6 ft 4 in. The marginal effect is also constant across all values of the other explanatory variables, such as age or presence of diabetes.

In a nonlinear model like logistic regression, the marginal effect of the risk factor is an informative way to answer the research question—how does a change in a risk factor affect the probability that the outcome occurs? In logistic regression, neither the marginal effect nor the OR is the same as the regression coefficient. Instead, the marginal effect reflects the nonlinear function on which the logistic regression model is based. Logistic regression ensures that predicted probabilities lie between 0 and 1, even for extreme values of a continuous risk factor, by modeling the relationship as a curve that fits between 0 and 1. Thus, the marginal effect of a 1-unit increase in age is not constant. The marginal effect will be small when the probability of the outcome is close to 0 or 1 and relatively large when the probability is close to 0.5. Because the values of the other covariates change the predicted probabilities, the marginal effect of any covariate depends on the value of other covariates in the model. For example, the marginal effect of a 1-unit increase in age may depend on whether the study participant is a man or a woman, even without including an interaction term between sex and age.[7] The variability in marginal effects makes intuitive sense because it is expected that the effect of a risk factor on the outcome is heterogeneous; ie, different effects for different values of the risk factor and other explanatory variables.

In logistic regression, there is no single marginal effect for the entire sample of individuals, so analysts must choose how to present marginal effects. The most common way is to report the average marginal effect across all persons in the data set, knowing that it is larger for some individuals and smaller for others. A second way is to report the marginal effect calculated at the means of all covariates. This can lead to a challenging interpretation; for instance, an estimated marginal effect of a risk factor for a person who is 50% pregnant or 20%

diabetic. A third way is to report the marginal effect for an individual with a specific set of characteristics; for example, the effect of an intervention on pregnant patients with diabetes.

An important advantage of marginal effects over ORs is that estimated marginal effects are less sensitive than ORs to inclusion of different sets of explanatory variables and estimation based on different samples of data.[4] The sensitivity of ORs and marginal effects to different model specifications and data sets was reviewed by Norton and Dowd.[6]

## What Are the Limitations of Marginal Effects?

Marginal effects vary across individuals, so it is important to present reported marginal effects in context by comparing the marginal effects with the magnitude of the baseline risk. For example, a change in probability of 1% may seem small if the baseline risk is 80% but may be large for a rare outcome (eg, baseline risk of 2%).

Care must be exercised when reporting marginal effects from case-control studies.[8] In this type of model, the sample proportions of the outcome values are not representative of the population.[5] Simple logistic models cannot provide either a meaningful marginal effect or a meaningful risk ratio from a case-control study, so ORs are the appropriate measures of association in this setting.

Until recently, it was challenging to compute marginal effects from logistic regressions and other nonlinear models such as ordered logistic, Poisson, negative binomial, and conditional logistic models. In recent years, standard statistical packages have added commands that make it easier to generate marginal effects, including the margins command in Stata and the margins package in R.

## How Should the Marginal Effects be Interpreted in Cummings et al?

Cummings et al[3] described how changes in 4 county-level characteristics would change the predicted probability of either having no access or having good access to mental health outpatient treatment facilities that accept Medicaid (see Table 2 in the article[3]). For example, an increase of 31 percentage points in the fraction of the county population living in a rural community (the standard deviation of that variable) would on average increase the probability of no access to mental health care by 27.9 percentage points (baseline risk = 34.8%) but would also increase the probability of good access by 3.4 percentage points (baseline risk = 20.2%), holding the effect of other explanatory variables constant. Such a change in rural population therefore would decrease the probability of some access, the third possible outcome, by 31.3 percentage points (27.9 + 3.4).

Marginal effects are a useful way to describe the average effect of changes in explanatory variables on the change in the probability of outcomes in logistic regression and other nonlinear models. Marginal effects provide a direct and easily interpreted answer to the research question of interest.

### ARTICLE INFORMATION

**Author Affiliations:** Department of Health Management and Policy, Department of Economics, University of Michigan, Ann Arbor (Norton); National Bureau of Economic Research, Cambridge, Massachusetts (Norton); Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis (Dowd); Durham Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Durham Veterans Affairs Health Care System, Durham, North Carolina (Maciejewski); Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina (Maciejewski); Division of General Internal Medicine, Department of Medicine, Duke University School of Medicine, Durham, North Carolina (Maciejewski).

**Corresponding Author:** Matthew L. Maciejewski, PhD, Durham Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Durham Veterans Affairs Health Care System, 508 Fulton St, Ste 600, Durham, NC 27705 (matthew.maciejewski@va.gov).

**Section Editors:** Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, *JAMA*.

**Published Online:** March 8, 2019. doi:10.1001/jama.2019.1954

**Conflict of Interest Disclosures:** Dr Maciejewski reported being supported by research career scientist award 10-391 from the Veterans Affairs Health Services Research and Development and support from the Durham Veterans Affairs Health Services Research and Development Center of Innovation (CIN 13-410); receiving grants from the National Institute on Drug Abuse and the Department of Veterans Affairs; receiving a contract from the National Committee for Quality Assurance to Duke University for research; and that his spouse owns stock in Amgen. No other disclosures were reported.

### REFERENCES

**1**. Meurer WJ, Tolles J. Logistic regression diagnostics: understanding how well a model predicts outcomes. *JAMA*. 2017;317(10):1068-1069. doi:10.1001/jama.2016.20441

**2**. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA*. 2016;316(5):533-534. doi:10.1001/jama.2016.7653

**3**. Cummings JR, Wen H, Ko M, Druss BG. Geography and the Medicaid mental health care infrastructure: implications for health care reform. *JAMA Psychiatry*. 2013;70(10):1084-1090. doi:10.1001/jamapsychiatry.2013.377

**4**. Norton EC, Dowd BE, Maciejewski ML. Odds ratios—current best practice and use. *JAMA*. 2018;320(1):84-85. doi:10.1001/jama.2018.6971

**5**. Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *BMJ Evid Based Med*. 1996;1(6):164-166. doi:10.1136/ebm.1996.1.164

**6**. Norton EC, Dowd BE. Log odds and the interpretation of logit models. *Health Serv Res*. 2018;53(2):859-878. doi:10.1111/1475-6773.12712

**7**. Karaca-Mandic P, Norton EC, Dowd B. Interaction terms in nonlinear models. *Health Serv Res*. 2012;47(1 pt 1):255-274. doi:10.1111/j.1475-6773.2011.01314.x

**8**. Irony TZ. Case-control studies: using "real-world" evidence to assess association. *JAMA*. 2018;320(10):1027-1028. doi:10.1001/jama.2018.12115