Accepted version of manuscript published in *Pharmacoeconomics*. Cite as:

Onukwugha, E., Bergtold, J. & Jain, R. A Primer on Marginal Effects—Part I: Theory and Formulae. *PharmacoEconomics* **33,** 25–30 (2015). https://doi.org/10.1007/s40273-014-0210-6

Type of Paper:

Practical Application

Title:

A Primer on Marginal Effects – Part I: Theory and Formulae

Short title: Primer on Marginal Effects – Part I

Authors: Onukwugha E[1], Bergtold J[2], Jain R[3]

[1]220 Arch Street, Department of Pharmaceutical Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA.

[2]304G Waters Hall, Department of Agricultural Economics, Kansas State University, Manhattan, KS, 66506-4011.

[3]HealthCore, Inc., 800 Delaware Avenue 5th Floor, Wilmington, DE 19801, USA.

Abstract

Marginal analysis evaluates changes in an objective function associated with a unit change in a relevant variable. The primary statistic of marginal analysis is the marginal effect (ME). The ME facilitates the examination of outcomes for defined patient profiles while measuring the change in original units (e.g., costs, probabilities). The ME has a long history in economics; however, it is not widely used in health services research despite its flexibility and ability to provide unique insights. This paper, the first in a two-part series, introduces and illustrates the calculation of the ME for a variety of regression models often used in health services research. Part One includes a review of prior studies discussing MEs followed by derivation of ME formulas for various regression models including: linear, logistic, multinomial logit model (MLM), generalized linear model (GLM) for continuous data, GLM for count data, two-part model, sample selection (two-stage) model and parametric survival model. Prior theoretical papers in health services research reported the derivation and interpretation of ME primarily for the linear and logistic models, with less emphasis on count models, survival models, MLM, two-part model, and sample selection models. These additional models are relevant for health services research studies examining costs and utilization. Part Two of the series focuses on the methods for estimating and interpreting the ME in applied research. The illustration, discussion, and application of ME in this two-part series support the conduct of future studies applying the marginal concept.

Word count: 241

Key points for decision makers

- The marginal effect (ME) measures change using the original units of the outcome measure and is applicable to decision making focused on the consequences of a unit (i.e., marginal) change in a variable or combination of variables.
- Despite its historical availability, flexibility, and ability to yield useful insights, the ME is not widely reported in health services research studies investigating, for example, medication use, hospital stays, utilization of other health services, and survival.
- This two-part study reviews prior literature regarding the use and reporting of ME (Part I) as well as provides ME formulas (Part I), statistical code (Part II) and empirical examples (Part II) to support its use and interpretation across a variety of models commonly used in health services research.

Section I. Introduction

Decisions made ‘at the margin’ compare the additional benefit with the additional cost of the decision and in doing so can lead to optimal allocation of scarce resources. Decision-making at the margin is applicable to decisions made by patients, providers, policy makers and payers. From a patient’s perspective, the marginal concept can be applied at baseline (e.g., what is the additional benefit of initiating androgen deprivation therapy compared to opting for radiation treatment compared to the additional cost?) or sequentially in real time (e.g., what is the additional benefit of continuing chemotherapy for one more month compared to the additional cost?). From a provider’s perspective, the benefits of decision making using the marginal concept may accrue in communicating with the patients about the treatment options [1] (e.g., cost versus benefit of active treatment compared to palliative treatment) or while deciding the optimal size of their practice (e.g., cost versus benefit of adding a new patient). From a policy maker’s point of view, the marginal concept can be applied in identifying program size (e.g., the optimal size of a transitional care program based on the cost and benefit of adding a new client) or how far to extend a health benefit (e.g., the cost and benefit of formulary tier placements).

The primary statistic of marginal analysis is the marginal effect (ME). The ME facilitates the examination of various outcomes for defined patient profiles while measuring the change in original units (e.g., costs, probabilities). Marginal analysis can be investigated within the framework of a regression model that represents some outcome or decision process. This is accomplished by the estimation of the ME or the impact on the outcome of interest (dependent variable) from an incremental increase in a given factor (independent variable). The ME also forms the basis for examining heterogeneity in effect, or the difference in estimated effects across patients or groups of patients. Heterogeneity in effect is examined by calculating the interaction ME. The interaction ME measures changes in the ME associated with unit changes in a given factor (continuous) or across groups (categories).

Prior work has illustrated differences between linear models and the binary regression models with respect to the derivation and interpretation of the ME[2] and has provided formulas and LIMDEP programs[3] to facilitate the use and interpretation of the ME and interaction effect[4, 5]. The estimation and interpretation of interaction effects also has been studied for generalized linear and two-stage models[3, 6-8] but have not been adopted into health services research. While the derivation of the ME is illustrated in textbooks[9, 10] and in the literature for a subset of models, it remains relatively underused in health services research examining cost and utilization patterns. Increased use of the ME provides opportunities for gaining insight into cost and utilization patterns across specific patient profiles. These insights are unique to MEs because, unlike hazard ratios, odds ratios, relative risks, or other similarly-transformed model parameter estimates, the ME is provided in absolute, not relative terms, and measured in natural units of the dependent variable, facilitating interpretation and translation for decision making.

The current paper represents Part One of a two-part series. The paper provides the theory and the derivation of marginal and interaction effects across an expanded set of models relevant for health services research. Part Two of the series provides specific empirical examples and programming code illustrating the use and interpretation of the ME. The goal is that the practical review, derivation and implementation of MEs will support its wider use in cost and utilization studies to inform decision making at the margin (e.g., the additional benefit of chemotherapy), heterogeneity in outcomes (e.g., race differences in outcomes) or heterogeneity in effects (e.g., race differences in treatment effects).

Section II. Marginal Effect Formulas for Health Services Research

The ME is useful for numerous applications relevant to health services research. For example, the ME can quantify changes in terms of dollars, probability of readmission, number of prescriptions, number of hospital admission days, number of physician office visits, and days of survival (waiting time) due to changes in independent variables of interest. The logit or probit model is typically estimated when the dependent variable is a binary measure representing the occurrence of an event within a given interval of time (e.g., treatment receipt, hospital readmission, mortality). Generalized linear models (GLM) are often used to model continuous (e.g., costs) and count measures (e.g., physician office visits, hospital length of stay) while survival (duration) models are often utilized when the waiting time is of interest (e.g., survival analysis). Two-stage models are less common but can be usefully applied to cost and utilization studies when sample selection bias is of concern or the data generating process regarding cost or utilization patterns is best captured by modeling two independent processes. The next section provides ME formulas for models relevant to health services research including models for continuous, categorical, count, and survival data.

II.A. Marginal and interaction effect functions

Let the regression function be given by $y = f(\boldsymbol{x}; \boldsymbol{\beta})$, where y is the dependent variable; $f(.)$ is a well-defined continuously differentiable function; $\boldsymbol{x} = (x_1, \dots, x_k, \dots, x_K)$ is a vector of K independent variables (indexed by k), and $\boldsymbol{\beta}$ is a vector of parameters to be estimated. Then mathematically, the ME for the k^{th} independent variable (i.e., ME_k) is the partial derivative of $f(\boldsymbol{x}; \boldsymbol{\beta})$, with respect to x_k, presented as $g(.)$, i.e:

$$\mathrm{ME_k} = \frac{\partial f(\boldsymbol{x};\boldsymbol{\beta})}{\partial x_k} = g(\boldsymbol{x};\boldsymbol{\beta}) \tag{1}$$

If x_k is patient age measured in years, then the interpretations of $\mathrm{ME_k}$ for the case of binary, cost, and survival regression models possibly include 1) the change in the probability of 30-day in-hospital mortality associated with a 1 year increase in the age at admission among stroke-related admissions; 2) the change in hospitalization costs associated with a 1 year increase in the age at admission; 3) the change in survival months associated with a 1 year increase in the age at diagnosis, respectively. The ME given by equation (1) is a constant only when it is the linear regression model without nonlinear transformations of the variables. Of note, *linear* in the linear regression model refers to linearity of the functional form with respect to $\boldsymbol{\beta}$. For a linear regression model without non-linear transformations of variables (e.g. x_k^2), the ME for a given independent variable x_k is a constant equal to β_k. Suppose the linear regression equation includes both x_k and x_k^2. Then the ME of x_k is neither constant nor independent of the starting value of x_k. In fact, the value of the ME changes as the value of x_k changes. In nonlinear regression functions the ME are potentially functions of all the explanatory variables and estimated parameters, even when there are no interaction terms. The ME changes as the values of the independent variables change. This provides the ability to estimate ME for different values of the independent variables or, for more practical purposes, particular subgroups of the patient population of interest.

The interaction effect or interaction ME is the change in a ME given a change in another independent variable of interest. That is, the interaction effect is the change in the ME for the independent variable x_k of a regression model given a change in another independent variable x_j. Mathematically, the interaction effect for any well-defined twice continuously differentiable function of $\boldsymbol{x}$, $f(.)$, is the partial derivative of $\mathrm{ME_k}$ with respect to x_j and presented as $h(.)$:

$$\mathrm{ME}_{kj} = \frac{\partial \mathrm{ME}_k}{\partial x_j} = \frac{\partial^2 f(\boldsymbol{x};\boldsymbol{\beta})}{\partial x_k \partial x_j} = h(\boldsymbol{x};\boldsymbol{\beta}). \tag{2}$$

Thus, the interaction effect is potentially also a function of all the independent variables and the estimated parameters. Building on the example involving 30-day mortality among stroke-related admissions, suppose x_j is an indicator for hemorrhagic stroke. The 'age effect', ME_k, represents the change in the probability of 30-day mortality associated with a one year increase in age at admission and thus ME_{kj} represents the difference in the age effect comparing patients with hemorrhagic stroke to patients who did not have hemorrhagic stroke.

II.B. Marginal Effect Formulas for Common Regression Models

Marginal and interaction effects are discussed for a number of regression models relevant for health services research. The text contains some ME formulas and the Technical Appendix includes a more complete listing of ME formulas for various cases. For simplicity and without loss of generality we will assume that the regression functions examined are functions of only two independent variables, x_1 and x_2. That is, $E(y|\boldsymbol{x}) = f(\boldsymbol{x};\boldsymbol{\beta}) = f(x_1, x_2;\boldsymbol{\beta})$.

The independent variables being examined may be continuous or discrete. This distinction is important, as the marginal or interaction effect involving a discrete random variable utilizes discrete differences. That is, for an independent binary variable, say x_2, that takes two values (0 or 1), the ME for x_2 is

$$\mathrm{ME}_2 = f(x_1, x_2;\boldsymbol{\beta})|_{x_2=1} - f(x_1, x_2;\boldsymbol{\beta})|_{x_2=0} \,. \tag{3}$$

For an interaction effect, if the other independent variable of interest (i.e., x_1) is also discrete then the formula involves taking a discrete difference as in equation (3) and the interaction effect on the ME for x_2 given a change in x_1 is:

$$\text{ME}_{21} = \text{ME}_2|_{x_1=1} - \text{ME}_2|_{x_1=0}. \quad (4)$$

The independent variable x_2 in equation (4) can be either continuous or discrete, as well.

Given these potential changes in the nature of independent variables, a number of cases of the derivation of ME for each regression model are examined. These cases deal with the functional form of the regression or predictor function being examined. They include:

Case 1: Linearity in the covariates with a continuous covariate.

Case 2: Inclusion of nonlinear transformations of continuous covariates.

Case 3: Inclusion of an interaction term between continuous covariates.

Case 4: Linearity in the covariates with a discrete covariate.

Case 5: Inclusion of an interaction term with a discrete covariate.

Case 6: Interaction effects when both covariates are continuous.

Case 7: Interaction effects with a continuous and discrete covariate.

Case 8: Interaction effects for a predictor linear in the coefficients and covariates.

For cases 1 to 3, we are interested in the ME of a continuous covariate, while for cases 4 and 5 we are interested in the ME for a discrete covariate. For the last three cases, 6 to 8, we are interested in the interaction effects between both covariates. Note that the terms independent variable and covariates are synonymous and are used interchangeably. The regression model and associated ME formula for Case 1 are described below. The ME formulas for the rest of the cases are available in the Technical Appendix.

II.B.1. Linear Regression Model

A typical linear regression equation with two independent variables takes the form:

$$E(y|x_1, x_2) = f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

The ME for Case 1 takes the form:

$$\text{ME}_2\text{: } \frac{\partial f(\cdot)}{\partial x_2} = \boldsymbol{\beta_2}$$

This regression may include nonlinear transformations of the independent variable, such as x_2^2, $\ln(x_2)$, and/or an interaction term between two variables, $x_1 \cdot x_2$. This regression model would still be a linear function of the coefficients even with the inclusion of these nonlinear transformations. In a linear regression with only linear transformations of variables, *the ME is constant*.

II.B.2. Logistic Regression Model (or Logit Model)

The logistic regression model is useful for investigating an event occurring over a fixed interval of time (e.g., 30-day readmission). The model assumes that the regression function is nonlinear and bounded between 0 and 1, allowing $E(y|\boldsymbol{x})$ to capture the probability of obtaining the outcome of interest (i.e. *y* = 1) (with the other outcome serving as the reference outcome. Let the predictor (or index) function of the model be given by $\eta(x_1, x_2; \beta)$, then the Logit model takes the form:

$$E(y = 1|x_1, x_2) = \Lambda\big(\eta(x_1, x_2; \beta)\big) = \frac{1}{1 + e^{-(\eta(x_1, x_2; \beta))}}$$

where, $\Lambda(\cdot)$ is the cumulative distribution function of the standard logistic distribution. The use of $\eta(x_1, x_2; \beta)$ signifies that η is a function of independent variables. It is important to note that $E(y = 1|x_1, x_2)$ is non-linear in the coefficients, β. Let $\eta(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, consistent with Case 1. Then ME_2 for Case 1 is given by:

$$\text{ME}_2\text{: } \frac{\partial \Lambda(\cdot)}{\partial x_2} = \frac{\partial \Lambda(\cdot)}{\partial \eta} * \frac{\partial \eta}{\partial x_2} = \boldsymbol{\Lambda(x_1, x_2)\big(1 - \Lambda(x_1, x_2)\big)\beta_2}$$

II.B.3. Multinomial Logistic Regression Model

The Logit model is a special case of a more general model known as the multinomial logistic regression model. The multinomial model can accommodate multiple nominal outcomes such as hospital discharge categories (e.g. home, skilled nursing facility, against medical advice) or patient groupings based on utilization (e.g., classes of medication). The conditional mean function is of interest here and includes a set of probability functions. Let the predictor for outcome *j* be $\eta_j(x_1, x_2; \beta)$ for *j* = 0,1,…,*J* outcomes. Then:

$$E(y = j|x_1, x_2, x_3) = \text{Prob}(y = j|x_1, x_2, x_3) \equiv h(\cdot) = \frac{\exp\,(\eta_j(x_1, x_2; \beta))}{1 + \sum_{k=1}^{J} \exp\,(\eta_j(x_1, x_2; \beta))},$$

where *j* = 0 is the reference outcome such that $\boldsymbol{\beta}_0 = 0$, making $\eta_0(x_1, x_2; \beta) = 1$, resulting in the denominator in the above function. Let $\eta_j(x_1, x_2; \beta) = \beta_{0,j} + \beta_{1,j}x_1 + \beta_{2,j}x_2$ for *j* = 1,…,*J*, consistent with Case 1. Then ME_2 for Case 1 is given by:

$$\text{ME}_{2,j}\text{: } \frac{\partial h(\cdot)}{\partial x_2} = \boldsymbol{h_j(x_1, x_2)\left[\beta_{2,j} - \textstyle\sum_{k=0}^{J} h_k(x_1, x_2)\beta_{2,k}\right]}..$$

II.B.4 Generalized Linear Model (GLM) with Log Link Function

The conditional mean for the GLM model with log link takes the form:

$$E(y|x_1, x_2) = \exp\big(\eta(x_1, x_2; \beta)\big),$$

where $\eta(x_1, x_2; \beta)$ is the predictor function. This regression allows for the estimation of a model where the dependent variable has a skewed distribution (e.g. cost data, length of hospital stay). Let $\eta(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, consistent with Case 1. Then ME_2 for Case 1 is given by:

$$\text{ME}_2\text{: } \frac{\partial exp(\eta)}{\partial x_2} = \mathbf{exp}(\boldsymbol{\eta})\,\boldsymbol{\beta_2}$$

II.B.5 Count Data Models

When the dependent variable (y) involves counts of events (e.g. number of visits by a patient to a physician or days of stay during a hospital admission), the underlying regression model may be modeled using a Poisson or a negative binomial distribution.

Poisson Model: The Poisson regression model is modeled using the Poisson distribution: $\boldsymbol{P}(y|x_1,x_2) = \left[e^{-\lambda_i}\lambda_i^y\right]/y!$, where $\lambda_i = E(y|x_1,x_2) = Var(y|x_1,x_2) = \exp\big(\eta(x_1,x_2;\beta)\big)$ and $\eta(x_1,x_2;\beta)$ is the predictor function. The regression model assumes equidispersion (i.e., the conditional mean is equal to the conditional variance) which may be too restrictive[9]. Let $\eta(x_1,x_2;\beta) \ = \ \beta_0 + \beta_1 x_1 + \beta_2 x_2$, consistent with Case 1. Then ME_2 for Case 1 is given by:

$$ME_2: \ \frac{\boldsymbol{\partial exp(\eta)}}{\boldsymbol{\partial x_2}} \ = \ \mathbf{exp}(\boldsymbol{\eta})\,\boldsymbol{\beta_2}$$

Additional details and derivations for the negative binomial model are available in the Technical Appendix.

II.B.6 Survival Models

Following Wooldridge[10], we specify a survival model using the hazard function:

$$\lambda(t;\boldsymbol{x}) = f(t|\boldsymbol{x})/S(t|\boldsymbol{x}),$$

where t is a duration of time; $f(t|\boldsymbol{x})$ is the conditional probability that the event of interest will occur at time t; $S(t|\boldsymbol{x})$ is the conditional survival function and describes the proportion of individuals at risk of the event of interest; and $\boldsymbol{x}$ is a vector of explanatory variables. The conditional hazard function $\lambda(t;\boldsymbol{x})$ (sometimes shown as $h(t;\boldsymbol{x})$) examines how the risk of an event changes over time [11]. A common approach for incorporating independent variables is to use proportional hazards models.

The proportional hazards model takes the form:

$$\lambda(t;\boldsymbol{x}) = \kappa(\boldsymbol{x})\lambda_0(t),$$

where $\lambda_0(t)$ is the baseline hazard. A common parameterization is to let $\kappa(\boldsymbol{x}) = \exp\left(\eta(x_1, x_2; \beta)\right)$, where $\eta(x_1, x_2; \beta)$ is a predictor function. The ME associated with a given continuous covariate is given by:

$$\frac{\partial\lambda(t;\boldsymbol{x})}{\partial\boldsymbol{x}} = \exp\left(\eta(x_1, x_2; \beta)\right)\frac{\partial\eta(x_1,x_2;\beta)}{\partial\boldsymbol{x}} \times \lambda_0(t).$$

II.B.7 Two-Part Regression Model

Two-part models can be useful for investigating continuously measured outcomes (e.g. cost, utilization) where a significant proportion of individuals report zero costs while others report positive costs. The process generating the zeroes is considered independent of the process generating continuous values.[12]. Consider a two part model that consists of a probability and distribution involving strictly positive values, i.e.

$$H(.) = \Pr(y = 1) \times E(y|y > 0)$$

where $\Pr(y = 1) = \frac{\exp(\eta)}{1+\exp(\eta)}$ and $E(y|y > 0) = \exp(\eta)$. The regression function is then given by: $f(x_1, x_2; \boldsymbol{\beta}) = \frac{exp(\eta)}{1+exp(\eta)} \times exp(\eta) = \frac{exp(\eta)}{1+exp(-\eta)}$, where $\eta = \eta(x_1, x_2; \boldsymbol{\beta})$ is the predictor function. To the extent that there is overlap between variables included in the first and second parts, the ME function will consist of two parts: 1) the ME associated with binary regression models such as the logit model and; 2) the ME associated with the GLM model. Let $\eta(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, consistent with Case 1, then ME2 for Case 1 is given by:

$$\text{ME}_2\text{:}\ \left(\frac{\partial \boldsymbol{f}(.)}{\partial \boldsymbol{x_2}}\right) = \frac{\boldsymbol{exp(\eta)}}{\boldsymbol{1+exp(-\eta)}}\,(\boldsymbol{\beta_2}) - \frac{\boldsymbol{exp(\eta)}}{[\boldsymbol{1+exp(-\eta)}]^2}\,\boldsymbol{exp(-\eta)}\,(\boldsymbol{\beta_2})$$

$$= \boldsymbol{f(x_1, x_2; \beta)}\,\boldsymbol{\beta_2}\left[\boldsymbol{1} + \frac{\boldsymbol{1}}{\boldsymbol{1 + exp(\eta)}}\right].$$

II.B.8 Sample Selection Model

Sample selection models are useful for investigating continuously measured outcomes (e.g. cost, utilization) in a subsample of patients where the subgroup with zero values are not necessarily 'true zeroes' but may result from non-observable responses. A 'selectivity correction' known as the inverse mills ratio (IMR) is included in the primary outcome equation. When the outcome measure follows a lognormal distribution (e.g., cost data, length of hospital stay), the conditional mean is log linear. The illustration is developed for the log linear case and can be simplified to the linear case. The Heckman model with a log linear dependent variable[13, 14] is specified below:

$$\ln y = X'\beta + \beta_\lambda \lambda(W'\gamma) + \varepsilon$$

$$\varepsilon_i \sim distr.(iid), E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2$$

The expression $\lambda(W'\gamma)$ has associated coefficient, β_λ, and represents the IMR. The IMR is an estimated function of a set of independent variables, W, and associated coefficients, γ. As with the two-part model, the ME function may consist of two parts. In general form for a given continuous variable, x_k, the ME$_k$ [7, 9]is given by:

$$\frac{\partial E(y)}{\partial x_k} = \frac{\partial\left(e^{X'\beta} e^{\beta_\lambda \lambda(\cdot)} E(e^\varepsilon|x)\right)}{\partial x_k}$$

$$= e^{X'\beta}\, e^{\beta_\lambda \lambda(\cdot)} \left(\beta_k E(e^\varepsilon|x) + \gamma_k \beta_\lambda(-\lambda(\cdot)^2 - \gamma x_k \lambda(\cdot) E(e^\varepsilon|x)) + \frac{\partial E(e^\varepsilon|x)}{\partial x_k}\right)$$

II.C. Standard Errors for Statistical Inference

To estimate confidence intervals or conduct significance testing using the marginal effects, standard errors must be estimated. These standard errors will be asymptotic. The estimation of standard errors for marginal effects begins with estimation of the asymptotic covariance matrix for the marginal effects, which requires the (asymptotic) covariance matrix of the coefficient estimates, Ω. There are several methods for estimating asymptotic standard errors[15]. These include:

(i) The delta method[9], which assumes coefficient estimates from the model are asymptotically multivariate normal. The delta method uses the derivatives of the marginal effects with respect to the estimated coefficients to estimate the asymptotic covariance matrix of the marginal effects as: $\widehat{Var}(ME_i) = \nabla ME_i^T(\Omega)\nabla ME_i$, where ∇ME_i is the gradient of the ME with respect to the coefficients being estimated.;

(ii) The method of Krinsky and Robb[9, 16], which ia a Monte Carlo method for estimating $\widehat{Var}(ME)$ that invokes the law of large numbers. It is usually assumed that the coefficient estimates from the estimated regression model are asymptotically multivariate normal.;

(iii) The bootstrap method[17, 18], which utilizes pseudo-random samples (by sampling with replacement) to estimate the parameters of interest, assuming the original data set used to estimate the model is an estimate of the population or the pseudo-population.; and

(iv) The delete-*d* Jackknife method[17], which uses pseudo-random samples to estimate the parameters of interest. The pseudo-random samples are generated from the original dataset by randomly removing a specified percentage or fraction of the data (i.e. sampling without replacement).

Section III. Conclusion

The marginal effect facilitates the examination of changes in clinical and cost outcomes for defined patient profiles while measuring changes in the same units (e.g., costs, probability) as the objective function. Part One of the two-part series provides an overview of the theory supporting marginal effect derivations and illustrates the derivation of the marginal effect for various models applicable to health services research. Part Two of the two-part series discusses confidence interval estimation, interprets the average marginal effect and marginal effect at the mean, as well as provides SAS and STATA programming code associated with the empirical examples presented. The goal is to encourage broader application and accurate interpretation of the marginal effect concept and measures in situations where information regarding movement along a primary objective function is of interest.

Acknowledgements

The authors are grateful for the research assistance provided by Richard Wooldridge and Dinci Pennap. The authors are grateful to S. Kofi Ampaabeng, PhD for his review of and feedback on the Technical Appendix. All remaining errors are the authors' responsibility.

The authors have no conflicts of interest to declare.

The review and interpretation of findings are the sole responsibility of the authors. Eberechukwu Onukwugha contributed to the conceptual framework, drafted manuscript sections, revised the manuscript with input from all co-authors and serves as the overall guarantor for the work. Jason Bergtold drafted manuscript sections and reviewed, commented on and edited all drafts of the manuscript. Rahul Jain contributed to the conceptual framework, drafted manuscript sections, reviewed, commented on and edited all drafts of the manuscript.

Funding

No funding was received for the preparation of this paper.

References

1. Smith TJ, Hillner BE. Explaining marginal benefits to patients, when "marginal" means additional but not necessarily small. Clin Cancer Res. Dec 15;16(24):5981-6.
2. Ai C, Norton EC. Interaction Terms in Logit and Probit Models. Economics Letters. 2003;80:123-9.
3. Karaca-Mandic P, Norton EC, Dowd B. Interaction terms in nonlinear models. Health Serv Res. Feb;47(1 Pt 1):255-74.
4. Hoetker G. The use of logit and probit models in strategic management research: Critical issues. Strategic Management Journal. 2007;28(4):331-43.
5. Anderson S, Newell RG. Simplified marginal effects in discrete choice models. Economics Letters. 2003;81:321-6.
6. Frondel M, Vance C. On Interaction Effects: The Case of Heckit and Two-Part Models. Journal of Economics and Statistics. 2013;233(1):22-38.
7. Vance C. Marginal effects and significance testing with Heckman's sample selection model: a methodological note. Applied Economics Letters. 2009;16(14):1415-9.
8. Berry WD, Golder M, Milton D. Improving Tests of Theories Positing Interaction. Journal of Politics. 2012;74(3):1653-71.
9. Greene W. Econometric Analysis. 7th edition ed. Upper Saddle River, NJ: Prentice Hall; 2012.
10. Wooldridge J. Econometric analysis of cross section and panel data Cambridge: MIT Press; 2002.
11. Ishak KJ, Kreif N, Benedict A, Muszbek N. Overview of parametric survival analysis for health-economic applications. Pharmacoeconomics. Aug;31(8):663-75.
12. Jones A. Heath Econometrics. First edition. ed. Amsterdam: Elsevier Science B. V.; 2000.
13. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics. 1998;17(3):283-95.
14. Dow WH, Norton EC. Choosing Between and Interpreting the Heckit and Two-part models for corner solutions. Health Services and Outcomes Research Methodology. 2003;4(1):5-18.
15. Dowd BE, Greene WH, Norton EC. Computation of standard errors. Health Serv Res. 2014 Apr;49(2):731-50.
16. Krinsky I, Robb L. On approximating the statistical properties of elasticities. The review of economics and statistics. 1986;68(4):715-9.
17. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall; 1993.
18. Martinez WL, Martinez AR. Computational statistics handbook with MATLAB. Boca Raton, FL: Chapman and Hall/CRC; 2002.