Annals of Clinical Epidemiology 2024;6(2):25–32

Annals of Clinical Epidemiology

SEMINAR

Expanding the Scope: In-depth Review of Interaction in Regression Models

Akihiro Shiroshita[1,2], Norio Yamamoto[2,3], Natsumi Saka[2,4,5], Hiroshi Shiba[6], Shinji Toki[7], Mari Yamamoto[8], Eisuke Dohi[9], Yuki Kataoka[3,10,11,12]

[1] *Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine*

[2] *Scientific Research Works Peer Support Group (SRWS-PSG)*

[3] *Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University*

[4] *Department of Health Research Methods, Evidence & Impact, McMaster University*

[5] *Department of Orthopedic Surgery, Teikyo University School of Medicine*

[6] *Department of Internal Medicine, Suwa Central Hospital*

[7] *Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University School of Medicine*

[8] *Department of Rheumatology and Nephrology,Chubu Rosai Hospital*

[9] *Department of Mental Disorder Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry*

[10] *Department of Internal Medicine, Kyoto Min-Iren Asukai Hospital*

[11] *Section of Clinical Epidemiology, Department of Community Medicine, Kyoto University Graduate School of Medicine*

[12] *Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine/Public Health*

KEY WORDS

Interaction, Effect modification, Regression

1. SCENARIO

In the research lab, Dr. X and his team were investigating the interaction between gene Z and smoking in relation to the onset of diabetes. They used a logistic regression model, incorporating an interaction term, and performed the Wald test for evaluating the statistical significance of the interaction term. However, their lab boss suddenly urged them to consider marginal effects instead. He said it was introduced in JAMA and told Dr. X to read the paper [1]. Dr. X read it but could not understand why his original analysis was inappropriate. For him, we will review regression analysis and interaction term, and explain what the marginal effect is.

The article is structured as follows. Section 2 briefly explains linear regression, generalized linear regression, and nonlinear regression, encompassing a review of the interpretation of coefficients in regression analysis. Section 3 reviewed the interpretation of an interaction term in multiple linear regression and logistic regression. It highlights a notable misapprehension and offers a rationale for an alternative approach. In Section 4, we introduce the concept of marginal effects. Lastly, in Section 5, we present our systematic review concerning gene-environment interactions (GEI) to evaluate the appropriateness of interpretation of an interaction term.

2. FUNDAMENTALS OF REGRESSION ANALYSIS

The objective of scholarly investigations is to approximate population characteristics (e.g., mean systolic blood pressure). Regression analysis is employed to delineate this quantity, especially a subgroup-specific comparison. Various nomenclatures exist for regression analysis. We usually talk about linearity in the predictor X. Some investigators might assert that linearity represents a linear association of X on the original scale or, at a minimum, on alternative scales (e.g., log and logistic). However, actually, it is so named because of the linearity on coefficients β. **Table 1** summarizes a definition of the relationship between outcome Y and predictor X based on the linearity on coefficients β. In certain publications, generalized linear regression, as opposed to linear regression, is designated as "non-linear regression" since Y is not linear with respect to X on the original scale. We refrain from emphasizing distinctions in terminology. Our concentration in this article pertains to generalized linear regression other than linear regression, where Y is not linear with respect to X on the original scale but linear with respect to X on alternative scales (e.g., logit or log).

We will provide readers some examples so that you can envision each form of regression analysis.

Example 1

Simple linear regression, $Y = \beta_0 + \beta_1 X + \varepsilon_i (E[\varepsilon_i|X_i] = 0)$.
ε_i: error term = the part of Y that is not explained by X.
Simple linear regression is a model used to describe the relationship between two variables by fitting a straight line to the data points. The goal of simple linear regression is to find the best-fitting line that minimizes the sum of squared differences between the actual data points and their corresponding predicted values on the line. In this example, the slope β_1 can be interpreted as difference in mean value of Y comparing subgroups differing in their value of X by a single unit. Here, we say the association is "linear" when the association between a predictor and the outcome is constant, that is fitted line perfectly straight.

This interpretation is the same in multiple linear regression. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i (E[\varepsilon_i|X_i] = 0)$. The slope β_1 can be interpreted as difference in mean value of Y comparing subgroups differing in their value of X_1 by a single unit and the same X_2.

Example 2

Generalized linear regression (logistic regression), $logit(P(Y = 1)) = \beta_0 + \beta_1 X$.
Generalized linear regression, also known as generalized linear model (GLM), is an extension of simple/multiple linear regression that allows for a broader range of rela-

Table 1 **Terminology used for categorizing regression models**

		Estimator		Equation
Linear model	Linear regression	Least squares/maximum likelihood estimation		$E[Y|X_1 = x_1, \dots X_K = X_k] = \beta_0 + \beta_1 x_1, \dots + \beta_K X_K$
	Generalized linear regression	Maximum likelihood estimation		$g^*(E[Y|X_1 = x_1, \dots X_K = X_k]) = \beta_0 + \beta_1 x_1, \dots + \beta_K X_K$
Non-linear model	Non-linear regression	Non-linear least squares		$Y_i = f^{**}(X_i; \beta) + \varepsilon_i^{***}$

* $g(\ )$: link function.
** $f(\ )$: non-linear function.
*** ε_i: error term; $E[\varepsilon_i|X_i] = 0$

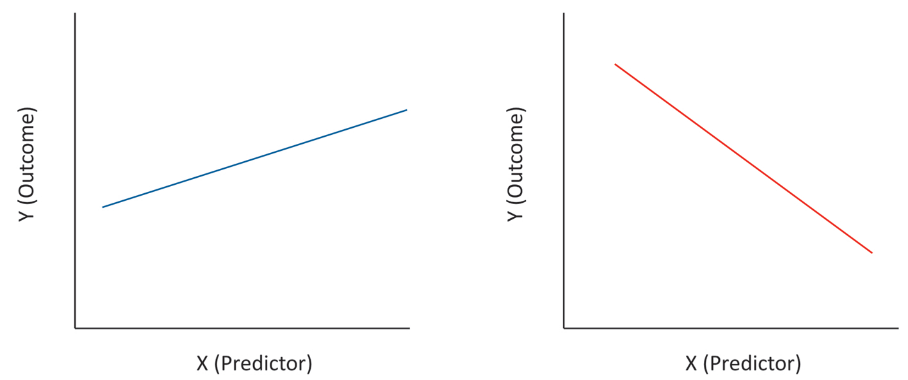

Fig. 1 **Fitted plot of a simple linear regression**

tionships between Y and X. Generalized linear regression can handle different types of response variables, such as binary, count, or categorical data, by introducing a link function and various kinds of a probability distribution of Y. It enables us to describe a more complex relationship compared to simple/multiple linear regression. A famous example is logistic regression. In the equation above, logit or log-odds is used as a link function. The distribution of Y is defined as the Bernoulli distribution. $Y \sim Bernoulli(p)$; $0 < p < 1$. We can interpret β_1 can be interpreted as the logit or log odds ratio of $Y = 1$ comparing subgroups differing in their value of X by a single unit. In other words, $\exp^{\beta_1}$ can be interpreted as the odds ratio of $Y = 1$ comparing subgroups differing in their value of X by a single unit. We should be aware that logistic regression is linear according to X on the "logit or log-odds" scale, but not on the original scale. Here, we say the association is "nonlinear" when the association between a predictor and the outcome is not constant, that is fitted line is not perfectly straight.

Example 3

Non-linear regression, $Y = \frac{\beta_0}{X + \beta_1} + \epsilon_i \left(E[\epsilon_i \,|\, X_i] = 0\right)$.

Nonlinear regression allows for much more complex and curved relationships between X and Y. The aforementioned equation is just one example of non-linear regression. This particular model, designated as the hyperbolic model, encapsulates the Michaelis-Menten kinetics, which we likely encountered long time ago (perhaps during the high school days).

3. INTERACTION

Henceforth, our attention will be devoted to generalized linear regression and interaction, especially logistic regression which is frequently employed in clinical inves-

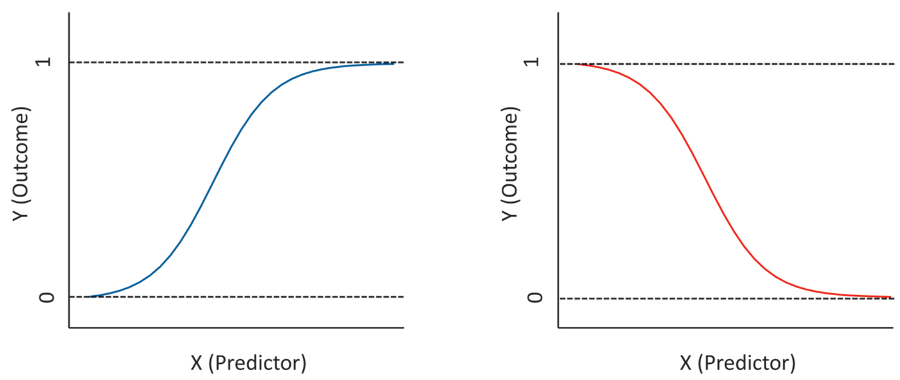

Fig. 2 Fitted plot of logistic regression

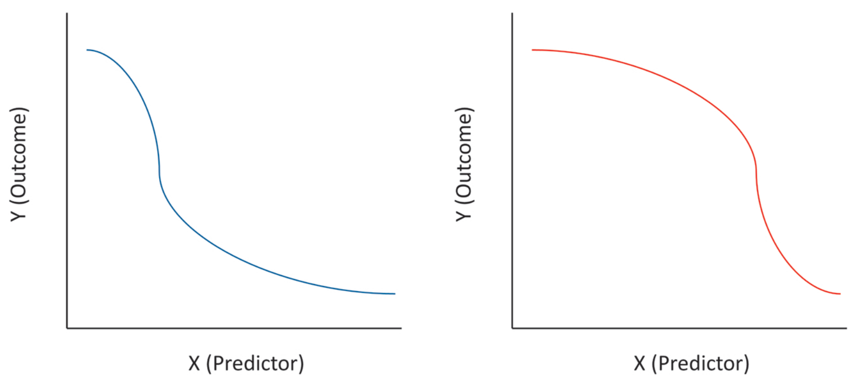

Fig. 3 Fitted plot of a non-linear regression

tigations. In this context, we assert the presence of an interaction when the relationship between one predictor is dependent upon the influence of another predictor. First, we will contemplate multiple linear regression with an interaction term, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon_i (E[\varepsilon_i|X_i] = 0)$. We might describe the relationship as "nonlinear" even though fitted lines are perfectly straight. It is because the association between a predictor and the outcome is not constant. Nonetheless, this semantic nuance does not hold an importance in the interpretation of the coefficients. β_1 can be interpreted as the association between X_1 and Y among the subgroup of $X_2 = 0$. β_3 can be interpreted as the difference in the association between X_1 and Y comparing subgroups differing in X_2 by one unit. If the p-value of β_3 is "statistically significant", we may describe there is an interaction.

Subsequently, we will deliberate on generalized linear regression with an interaction term other than multiple linear regression, $logit(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$.

Please remember logistic regression is linear according to X on the "logit or log-odds" scale, but not on the original scale. Most of the time, our principal focus is on the probability of $Y = 1$, not logit of $Y = 1$. In that situation, the model is inherently interactive because the association between X_1 and Y is not constant and can be influenced by the value of X_2 (i.e., $logit(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$). Remember in logistic regression, the change of the intercept moves the fitted curve of the association between X_1 and Y upward or downward (**Fig. 5**). Because the fitted curve is S-shaped, the change in probability of $Y = 1$ based on the change from $x_1 \rightarrow x_2$ would change. It

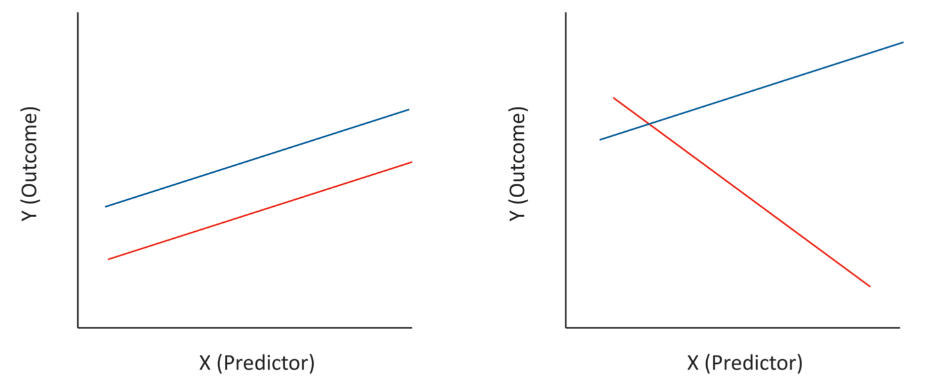

Fig. 4 **Fitted plot of multiple linear regression with an interaction term (blue: subgroup A and red: subgroup B)**

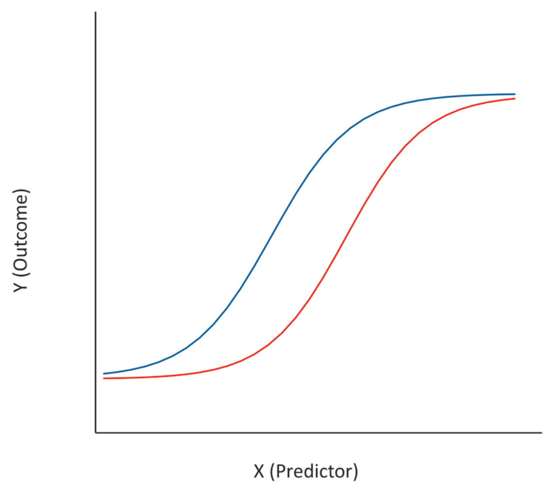

Fig. 5 **Shift of fitted curve based on intercept-only change in logistic regression**

also applies to multivariable logistic regression, $logit(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. When X_2 changes, the fitted curve will move upward/downward and the change in probability of $Y = 1$ based on the change from $x_1 \rightarrow x_2$ would change as well. In other words, the relationship between one predictor is dependent upon the influence of another predictor, which is a definition of interaction. Thus, logistic regression has inherent interaction of X_2 on the association between X_1 and $P(Y = 1)$. In this logistic regression, the coefficient on the interaction term does not necessarily indicate the presence of interaction between X_1 and Y based on X_2. Then, what is the meaning of adding an interaction term? This phenomenon is applied to other generalized linear regression such as Poisson regression and log-risk model (i.e., relative risk model or log-binomial model). It aims to improve the model fitness (**Fig. 6**). In the field of sociology and economics, researchers are recommended to avoid interpreting the coefficient of interaction terms in "non-linear" models (i.e., generalized linear regression other than multiple linear regression) [2]. In the medical field, our previous meta-epidemiological study elucidated even among the randomized controlled trials published in 10 high-Journal-Impact-Factor journals, the coefficients of non-linear regression models were not appropriately interpreted [3].

4. MARGINAL EFFECT

Instead of interpreting the coefficients of interaction terms, several alternatives have proposed recommendations [4]. One of them is the marginal effect, or the incremental change in the outcome associated with a one-unit change in a particular predictor, while holding all other variables constant (**Fig. 7**).

$$Marginal\ effect = \eta(x_k = b,\ X_{-k} = X^*) - \eta(x_k = a,\ X_{-k} = X^*)$$

x_k: a predictor of interest
X_{-k}: control variables
$\eta(\ )$: difference between two predictions from $x_k = a \rightarrow x_k = b$

It helps to elucidate the impact of each predictor on the outcome in question. Examining the second difference in the two marginal effects across a subgroup could be an appropriate way to evaluate an effect modification on the original scale (**Fig. 7**). As an example, we will share our sample R code for evaluating a marginal effect and second difference (https://github.com/AkiShiroshita/Supplement-interaction/tree/main).

Another strategy is calculating relative excess risk due to interaction (RERI). It indicates additive interaction while the regression coefficient of generalized linear model without simple/multiple linear regression indicates multiplicative interaction. While We do not provide an in-depth elucidation herein, but we recommend instructive guide authored by Tyler J. VanderWeele and colleagues [5].

5. EXAMPLE: SYSTEMATIC REVIEW ON G*E INTERACTION STUDIES

In the field of genetics, the environmental effect can differ based on the presence of individual genotype, which is called a genome-environment interaction (G*E interaction or GEI) [6]. To date, a lot of studies have evaluated and proposed GEI and it is one of the hot topics [7, 8].

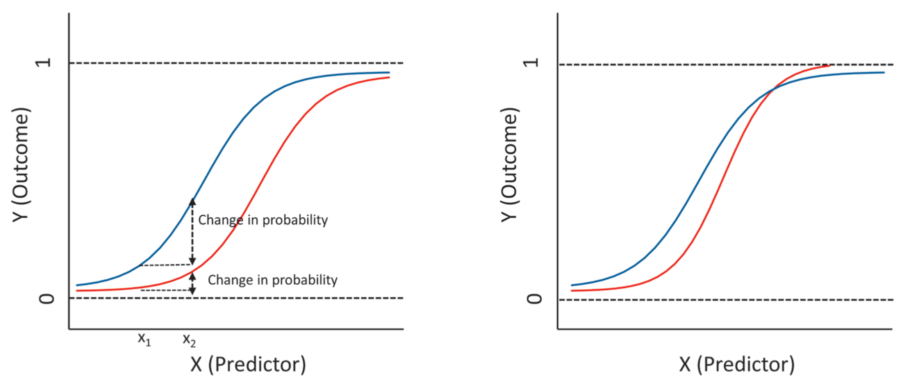

Fig. 6 **Fitted plot of logistic regression without/with an interaction term**

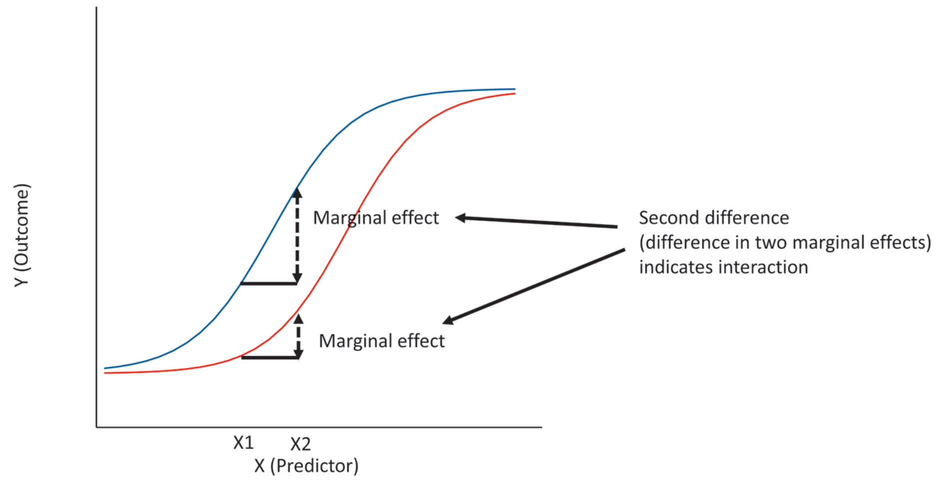

Fig. 7 **Marginal effects and second difference (blue: subgroup A and red: subgroup B)**

The outcome of interest is often a binary or categorical outcome. We comprehensively reviewed how regression models are used in analysis of GEI including Genome-Wide Association Study (GWAS).

The methodologies are expounded upon in the supplementary file with comprehensive detail. Articles were selected via MEDLINE through Ovid. Inclusion criteria encompassed full-text observational studies evaluating gene-environment interactions (GEI) utilizing regression models, irrespective of outcome types (e.g., primary, secondary, or exploratory outcomes) up until August 15, 2022. We limited our search to English language articles. Review articles and case reports were excluded. Following title and abstract screening, we randomly sampled 50 articles, after which one of the authors (HS, ST, MY, or ED) conducted full-text reviews. Another author (AS, NY, NS, or YK) corroborated the findings and determined the final inclusion of articles. When researchers in the original study employed generalized linear regression other than simple/multiple linear regression on at least one categorical outcome and interpreted the significant result of the coefficient of interaction terms as the presence of GEI, we deemed it as an "inappropriate interpretation". Conversely, when they utilized simple/multiple linear regressions and interpreted the significant result of the coefficient of interaction terms as the presence of GEI, we considered it as an "appropriate interpretation". Furthermore, when they employed generalized linear regression other than simple/multiple linear regression and assessed the presence of GEI based on alternative metrics, such as visual inspection of forest plots and marginal effects, we likewise judged it as an "appropriate interpretation". In instances where we could not evaluate their interpretation of interactions within the main text, supplements, or cited protocols, we deemed it as an "unclear description". One of the authors (HS, ST, MY, or ED) appraised the appropriateness, and the other two authors (NY, NS, or YK, and AS) confirmed it. In cases of conflict, resolution was achieved through discussion.

As a result, our Ovid search selected 2,071 studies, and after the title and abstract screening, 560 studies remained. Among them, we randomly selected 50 studies and performed a full-text review. Finally, 19 studies were included in our analysis. **Table 2** summarizes the study characteristics. We discerned an inappropriate interpretation of an interaction term among 10/19 (53%) of the included studies, which constitutes a remarkable proportion of articles. Among them, 8/10 (80%) used logistic regression. In certain investigations, multiple linear regression incorporating an interaction term was employed to assess binary outcomes (i.e., log-linear model or linear probability model) [9–11]. Nonetheless, we were concerned that these models might not fit the data well. They yield predicted probabilities beyond the zero-to-one range, and the difference between the estimate and true value would be substantial when the majority of probabilities are proximate to either zero or one. Our systematic review highlights the current situation where many researchers misinterpret an interaction term in generalized linear regression.

6. CONCLUSION

Dr. X got aware that in the context of generalized linear

Table 2 Study characteristics

Study name	Sample size	Type of outcome	Software	Model	Multiplicity adjustment	At least one significant result	Inappropriate interpretation of an interaction term
Abdulkadir 2021 [9]	678	Categorical	R	Linear regression	Benjamin-Hochberg method	Yes	Appropriate
Li 2019 [12]	1,140	Categorical	R	Logistic regression	Permutation tests	Yes	Inappropriate
Yang 2015 [13]	1,336	Categorical	SPSS	Logistic regression	Not	Yes	Inappropriate
Wu 2011 [14]	399	Categorical	SPSS	Logistic regression	Not	Yes	Inappropriate
Angstadt 2014 [15]	Over 1,800	Categorical	SAS	Logistic regression	Benjamin-Hochberg method	Yes	Inappropriate
White 2012 [16]p. 5	139	Continuous	Statistica	Generalized linear model	Bonferroni correction	Yes	Appropriate
Elam 2018 [10]	479	Categorical	Mplus	Structural equation modeling	Not	Yes	Appropriate
Tang 2020 [17]	20,155	Categorical	QUANTO software	Logistic regression	Bonferroni correction	Yes	Inappropriate
Rask-Anderson 2017 [18]	362,496	Continuous	R	Linear regression	False discovery rate	Yes	Appropriate
Aklillu 2018 [19]	163	Categorical	Arlequin	Non-linear regression	Not	Yes	Inappropriate
Schweren 2016 [20]	316	Continuous	Not described	Linear mixed effects model	Not	Yes	Appropriate
Meer 2016 [21]	539	Continuous	R	Linear mixed effects model	False-discovery rate and family-wise error correction	Yes	Appropriate
Mullins 2016 [11]	2,769	Categorical	R	Linear regression and logistic regression	Bonferroni correction	Yes	Appropriate
Bolhuis 2019 [22]	2,512	Continuous	R	Linear regression	False-discovery rate	Yes	Appropriate
Sund 2021 [23]	41,198	Categorical	STATA	Mixed-effects logistic regression	Not	Yes	Inappropriate
Lehto 2020 [24]	243,797	Categorical	STATA	Logistic regression model	Bonferroni correction	No	Inappropriate
Sarginson 2014 [25]	1,222	Categorical	R	Poisson regression model	False-discovery rate	Yes	Inappropriate
Schmidt 2006 [26]	810	Categorical	SAS	Logistic regression model	Not	Yes	Inappropriate
Tin 2015 [27]	11,663	Continuous	Metal and R	Linear regression	P-value < 5*10-8	Yes	Appropriate

regression beyond simple/multiple linear regression, a significant coefficient does not necessarily indicate the presence of interaction, as it may do in linear models. He decided to use marginal effects of gene Z based on the presence/absence of smoking, and evaluated the second difference (i.e., difference of two marginal effects). He could find a significant interaction on the probability scale and his lab boss praised his effort. Finally, his paper was published in an outstanding journal in his field. We share his descriptions on marginal effects. In the method section, "We calculated the marginal effects of gene Z in smokers and non-smokers", and test the second difference, or whether two marginal effects are equal or not". In the results section, "the association of gene Z was stronger for smokers than non-smokers (second difference = 0.082; p-value = 0.03).".

CONFLICT OF INTEREST

Nothing to declare.

ACKNOWLEDGMENT

None.

REFERENCES

1. 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.
2. Editors' Comment: A Few Guidelines for Quantitative Submissions - Sarah A. Mustillo, Omar A. Lizardo, Rory M. McVeigh, 2018. Accessed May 8, 2022. https://journals.sagepub.com/doi/full/10.1177/0003122418806282
3. Shiroshita A, Yamamoto N, Saka N, Okumura M, Shiba H, Kataoka Y. Inappropriate Evaluation of Effect Modifications Based on Categorical Outcomes: A Systematic Review of Randomized Controlled Trials. *Int J Environment Res Public Health* 2022; 19:15262.
4. Mize TD. Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociol Sci* 2019;6:81–117.
5. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiol Methods* 2014;3:33–72.
6. Gene–environment-wide association studies: emerging approaches | Nature Reviews Genetics. Accessed July 20, 2022. https://www.nature.com/articles/nrg2764
7. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry* 2011;168:1041–9. doi: 10.1176/appi.ajp.2011.11020191
8. Border R, Johnson EC, Evans LM, Smolen A, Berley N, Sullivan PF, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *Am J Psychiatry* 2019;176:376–87.
9. Abdulkadir M, Yu D, Osiecki L, King RA, Fernandez TV, Brown LW, et al. Investigation of gene-environment interactions in relation to tic severity. *J Neural Transm (Vienna)* 2021;128:1757–65.
10. Elam KK, Chassin L, Pandika D. Polygenic risk, family cohesion, and adolescent aggression in Mexican American and European American families: Developmental pathways to alcohol use. *Dev Psychopathol* 2018;30:1715–28.
11. Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med* 2016;46:759–70.
12. Li G, Wang L, Zhang K, Cao C, Cao X, Fang R, et al. FKBP5 Genotype Linked to Combined PTSD-Depression Symptom in Chinese Earthquake Survivors. *Can J Psychiatry* 2019;64:863–71.
13. Yang HY, Lu KC, Fang WH, Lee HS, Wu CC, Huang YH, et al. Impact of interaction of cigarette smoking with angiotensin-converting enzyme polymorphisms on end-stage renal disease risk in a Han Chinese population. *J Renin Angiotensin Aldosterone Syst* 2015;16:203–10.
14. Wu J, Zheng Q, Huang YQ, Wang Y, Li S, Lu DW, et al. Significant evidence of association between polymorphisms in ZNF533, environmental factors, and nonsyndromic orofacial clefts in the Western Han Chinese population. *DNA Cell Biol* 2011;30:47–54.
15. Angstadt AY, Hartman TJ, Lesko SM, Muscat JE, Zhu J, Gallagher CJ, et al. The effect of UGT1A and UGT2B polymorphisms on colorectal cancer risk: haplotype associations and gene–environment interactions. *Genes Chromosomes Cancer* 2014;53:454–66.
16. White MG, Bogdan R, Fisher PM, Muñoz KE, Williamson DE, Hariri AR. FKBP5 and emotional neglect interact to predict individual differences in amygdala reactivity. *Genes Brain Behav* 2012;11:869–78.
17. Tang H, Jiang L, Stolzenberg-Solomon RZ, Arslan AA, Freeman LEB, Bracci BM, et al. Genome-Wide Gene-Diabetes and Gene-Obesity Interaction Scan in 8,255 Cases and 11,900 Controls from PanScan and PanC4 Consortia. *Cancer Epidemiol Biomarkers Prev* 2020;29:1784–91.
18. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet* 2017;13:e1006977.
19. Aklillu E, Carrillo JA, Makonnen E, Bertilsson L, Djordjevic N. N-Acetyltransferase-2 (NAT2) phenotype is influenced by genotype-environment interaction in Ethiopians. *Eur J Clin Pharmacol* 2018;74:903–11.
20. Schweren LJS, Hartman CA, Heslenfeld DJ, Groenman AP, Franke B, Oosterlaan J, et al. Age and DRD4 Genotype Moderate Associations Between Stimulant Treatment History and Cortex Structure in Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry* 2016;55:877–85.e3.
21. van der Meer D, Hoekstra PJ, Bralten J, van Donkelaar M, Heslenfeld DJ, Oosterlaan J, et al. Interplay between stress response genes associated with attention-deficit hyperactivity disorder and brain volume. *Genes Brain Behav* 2016;15:627–36.
22. Bolhuis K, Tiemeier H, Jansen PR, Muetzel RL, Neumann A, Hillegers MHJ, et al. Interaction of schizophrenia polygenic risk and cortisol level on pre-adolescent brain structure. *Psychoneuroendocrinology* 2019;101: 295–303.
23. Sund ER, van Lenthe FJ, Avendano M, Raina P, Krokstad S. Does urbanicity modify the relationship between a polygenic risk score for depression and mental health symptoms? Cross-sectional evidence from the observational HUNT Study in Norway. *J Epidemiol Community Health* 2021;75:420–5.
24. Lehto K, Hägg S, Lu D, Karlsson R, Pedersen NL, Mosing MA. Childhood Adoption and Mental Health in Adulthood: The Role of Gene-Environment Correlations and Interactions in the UK Biobank. *Biol Psychiatry* 2020;87:708–16.
25. Sarginson JE, Deakin JFW, Anderson IM, Downey D, Thomas E, Elliott R, et al. Neuronal nitric oxide synthase (NOS1) polymorphisms interact with financial hardship to affect depression risk. *Neuropsychopharmacology* 2014;39:2857–66.
26. Schmidt S, Hauser MA, Scott WK, Postel EA, Agarwal A, Gallins P, et al. Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum Genet* 2006;78:852–64.
27. Tin A, Köttgen A, Folsom AR, Maruthur NM, Tajuddin SM, Nalls MA, et al. Genetic loci for serum magnesium among African-Americans and gene-environment interaction at MUC1 and TRPM6 in European-Americans: the Atherosclerosis Risk in Communities (ARIC) study. *BMC Genet* 2015; 16:56.