

(Check for updates

EDITORIAL The clinician's guide to interpreting a regression analysis

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2022

Eye (2022) 36:1715-1717; https://doi.org/10.1038/s41433-022-01949-z

INTRODUCTION

When researchers are conducting clinical studies to investigate factors associated with, or treatments for disease and conditions to improve patient care and clinical practice, statistical evaluation of the data is often necessary. Regression analysis is an important statistical method that is commonly used to determine the relationship between several factors and disease outcomes or to identify relevant prognostic factors for diseases [1].

This editorial will acquaint readers with the basic principles of and an approach to interpreting results from two types of regression analyses widely used in ophthalmology: linear, and logistic regression.

LINEAR REGRESSION ANALYSIS

Linear regression is used to quantify a linear relationship or association between a continuous response/outcome variable or dependent variable with at least one independent or explanatory variable by fitting a linear equation to observed data [1]. The variable that the equation solves for, which is the outcome or response of interest, is called the dependent variable [1]. The variable that is used to explain the value of the dependent variable is called the predictor, explanatory, or independent variable [1].

In a linear regression model, the dependent variable must be continuous (e.g. intraocular pressure or visual acuity), whereas, the independent variable may be either continuous (e.g. age), binary (e.g. sex), categorical (e.g. age-related macular degeneration stage or diabetic retinopathy severity scale score), or a combination of these [1].

When investigating the effect or association of a single independent variable on a continuous dependent variable, this type of analysis is called a simple linear regression [2]. In many circumstances though, a single independent variable may not be enough to adequately explain the dependent variable. Often it is necessary to control for confounders and in these situations, one can perform a multivariable linear regression to study the effect or association with multiple independent variables on the dependent variable [1, 2]. When incorporating numerous independent variables, the regression model estimates the effect or contribution of each independent variable while holding the values of all other independent variables constant [3].

When interpreting the results of a linear regression, there are a few key outputs for each independent variable included in the model:

 Estimated regression coefficient—The estimated regression coefficient indicates the direction and strength of the relationship or association between the independent and

Received: 8 January 2022 Revised: 17 January 2022 Accepted: 18 January 2022 Published online: 31 January 2022

dependent variables [4]. Specifically, the regression coefficient describes the change in the dependent variable for each one-unit change in the independent variable, if continuous [4]. For instance, if examining the relationship between a continuous predictor variable and intra-ocular pressure (dependent variable), a regression coefficient of 2 means that for every one-unit increase in the predictor, there is a two-unit increase in intra-ocular pressure. If the independent variable is binary or categorical, then the oneunit change represents switching from one category to the reference category [4]. For instance, if examining the relationship between a binary predictor variable, such as sex, where 'female' is set as the reference category, and intra-ocular pressure (dependent variable), a regression coefficient of 2 means that, on average, males have an intra-ocular pressure that is 2 mm Hg higher than females.

- Confidence Interval (CI)—The CI, typically set at 95%, is a measure of the precision of the coefficient estimate of the independent variable [4]. A large CI indicates a low level of precision, whereas a small CI indicates a higher precision [5].
- 3. *P* value—The *p* value for the regression coefficient indicates whether the relationship between the independent and dependent variables is statistically significant [6].

LOGISTIC REGRESSION ANALYSIS

As with linear regression, logistic regression is used to estimate the association between one or more independent variables with a dependent variable [7]. However, the distinguishing feature in logistic regression is that the dependent variable (outcome) must be binary (or dichotomous), meaning that the variable can only take two different values or levels, such as '1 versus 0' or 'yes versus no' [2, 7]. The effect size of predictor variables on the dependent variable is best explained using an odds ratio (OR) [2]. ORs are used to compare the relative odds of the occurrence of the outcome of interest, given exposure to the variable of interest [5]. An OR equal to 1 means that the odds of the event in one group are the same as the odds of the event in another group; there is no difference [8]. An OR > 1 implies that one group has a higher odds of having the event compared with the reference group, whereas an OR < 1 means that one group has a lower odds of having an event compared with the reference group [8]. When interpreting the results of a logistic regression, the key outputs include the OR, CI, and p-value for each independent variable included in the model.

CLINICAL EXAMPLE

Sen et al. investigated the association between several factors (independent variables) and visual acuity outcomes (dependent variable) in patients receiving anti-vascular endothelial growth factor therapy for macular oedema (DMO) by means of both linear and logistic regression [9]. Multivariable linear regression demonstrated 1716

that age (Estimate -0.33, 95% CI -0.48 to -0.19, p < 0.001) was significantly associated with best-corrected visual acuity (BCVA) at 100 weeks at alpha = 0.05 significance level [9]. The regression coefficient of -0.33 means that the BCVA at 100 weeks decreases by 0.33 with each additional year of older age.

Multivariable logistic regression also demonstrated that age and ellipsoid zone status were statistically significant associated with achieving a BCVA letter score >70 letters at 100 weeks at the alpha = 0.05 significance level. Patients \geq 75 years of age were at a decreased odds of achieving a BCVA letter score >70 letters at 100 weeks compared to those <50 years of age, since the OR is less than 1 (OR 0.96, 95% CI 0.94 to 0.98, p = 0.001) [9]. Similarly, patients between the ages of 50-74 years were also at a decreased odds of achieving a BCVA letter score >70 letters at 100 weeks compared to those <50 years of age, since the OR is less than 1 (OR 0.15, 95% CI 0.04 to 0.48, p = 0.001) [9]. As well, those with a not intact ellipsoid zone were at a decreased odds of achieving a BCVA letter score >70 letters at 100 weeks compared to those with an intact ellipsoid zone (OR 0.20, 95% CI 0.07 to 0.56; p = 0.002). On the other hand, patients with an ungradable/guestionable ellipsoid zone were at an increased odds of achieving a BCVA letter score >70 letters at 100 weeks compared to those with an intact ellipsoid zone, since the OR is greater than 1 (OR 2.26, 95% CI 1.14 to 4.48; p = 0.02) [9].

The narrower the CI, the more precise the estimate is; and the smaller the p value (relative to alpha = 0.05), the greater the evidence against the null hypothesis of no effect or association.

CONCLUSION

Simply put, linear and logistic regression are useful tools for appreciating the relationship between predictor/explanatory and outcome variables for continuous and dichotomous outcomes, respectively, that can be applied in clinical practice, such as to gain an understanding of risk factors associated with a disease of interest.

Sofia Bzovsky¹, Mark R. Phillips \textcircled{D}^2 , Robyn H. Guymer $\textcircled{D}^{3,4}$, Charles C. Wykoff^{5,6}, Lehana Thabane $\textcircled{D}^{2,7}$, Mohit Bhandari $\textcircled{D}^{1,2}$, Varun Chaudhary $\textcircled{D}^{1,2}$ and on behalf of the R.E.T.I.N.A. study group*

¹Department of Surgery, McMaster University, Hamilton, ON, Canada. ²Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, ON, Canada. ³Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia. ⁴Department of Surgery, (Ophthalmology), The University of Melbourne, Melbourne, VIC, Australia. ⁵Retina Consultants of Texas (Retina Consultants of America), Houston, TX, USA. ⁶Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA. ⁷Biostatistics Unit, St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada. * A list of authors and their affiliations appears at the end of the paper. [⊠]email: vchaudh@mcmaster.ca

REFERENCES

- 1. Schneider A, Hommel G, Blettner M. Linear Regression. Anal Dtsch Ärztebl Int. 2010;107:776–82.
- Bender R. Introduction to the use of regression models in epidemiology. In: Verma M, editor. Cancer epidemiology. Methods in molecular biology. Humana Press; 2009:179–95.
- 3. Schober P, Vetter TR. Confounding in observational research. Anesth Analg. 2020;130:635.

- Schober P, Vetter TR. Linear regression in medical research. Anesth Analg. 2021; 132:108–9.
- 5. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry. 2010; 19:227–9.
- Thiese MS, Ronna B, Ott U. P value interpretations and considerations. J Thorac Dis. 2016;8:E928–31.
- 7. Schober P, Vetter TR. Logistic regression in medical research. Anesth Analg. 2021; 132:365–6.
- Zabor EC, Reddy CA, Tendulkar RD, Patil S. Logistic regression in clinical studies. Int J Radiat Oncol Biol Phys. 2022;112:271–7.
- Sen P, Gurudas S, Ramu J, Patrao N, Chandra S, Rasheed R, et al. Predictors of visual acuity outcomes after anti-vascular endothelial growth factor treatment for macular edema secondary to central retinal vein occlusion. Ophthalmol Retin. 2021;5:1115–24.

R.E.T.I.N.A. STUDY GROUP

Varun Chaudhary^{1,2}, Mohit Bhandari^{1,2}, Charles C. Wykoff^{5,6}, Sobha Sivaprasad⁸, Lehana Thabane^{2,7}, Peter Kaiser⁹, David Sarraf¹⁰, Sophie J. Bakri¹¹, Sunir J. Garg¹², Rishi P. Singh^{13,14}, Frank G. Holz¹⁵, Tien Y. Wong^{16,17}, and Robyn H. Guymer^{3,4}

AUTHOR CONTRIBUTIONS

SB was responsible for writing, critical review and feedback on manuscript. MRP was responsible for conception of idea, critical review and feedback on manuscript. RHG was responsible for critical review and feedback on manuscript. CCW was responsible for critical review and feedback on manuscript. LT was responsible for critical review and feedback on manuscript. LT was responsible for critical review and feedback on manuscript. WB was responsible for conception of idea, critical review and feedback on manuscript. VC was responsible for conception of idea, critical review and feedback on manuscript.

COMPETING INTERESTS

SB: Nothing to disclose. MRP: Nothing to disclose. RHG: Advisory boards: Bayer, Novartis, Apellis, Roche, Genentech Inc.-unrelated to this study. CCW: Consultant: Acuela, Adverum Biotechnologies, Inc, Aerpio, Alimera Sciences, Allegro Ophthalmics, LLC, Allergan, Apellis Pharmaceuticals, Bayer AG, Chengdu Kanghong Pharmaceuticals Group Co. Ltd. Clearside Biomedical, DORC (Dutch Ophthalmic Research Center), EyePoint Pharmaceuticals, Gentech/Roche, GyroscopeTx, IVERIC bio, Kodiak Sciences Inc, Novartis AG, ONL Therapeutics, Oxurion NV, PolyPhotonix, Recens Medical, Regeron Pharmaceuticals, Inc, REGENXBIO Inc, Santen Pharmaceutical Co, Ltd, and Takeda Pharmaceutical Company Limited; Research funds: Adverum Biotechnologies, Inc, Aerie Pharmaceuticals, Inc, Aerpio, Alimera Sciences, Allergan, Apellis Pharmaceuticals, Chengdu Kanghong Pharmaceutical Group Co, Ltd, Clearside Biomedical, Gemini Therapeutics, Genentech/Roche, Graybug Vision, Inc, GyroscopeTx, Ionis Pharmaceuticals, IVERIC bio, Kodiak Sciences Inc, Neurotech LLC, Novartis AG, Opthea, Outlook Therapeutics, Inc, Recens Medical, Regeneron Pharmaceuticals, Inc, REGENXBIO Inc, Samsung Pharm Co, Ltd, Santen Pharmaceutical Co, Ltd, and Xbrane Biopharma AB-unrelated to this study. LT: Nothing to disclose. MB: Research funds: Pendopharm, Bioventus, Acumed-unrelated to this study. VC: Advisory Board Member: Alcon, Roche, Bayer, Novartis; Grants: Bayer, Novartis—unrelated to this study.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Varun Chaudhary.

Reprints and permission information is available at http://www.nature.com/ reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ON BEHALF OF THE R.E.T.I.N.A. STUDY GROUP

Varun Chaudhary ^{1,2 A}, Mohit Bhandari ^{1,2}, Charles C. Wykoff^{5,6}, Sobha Sivaprasad⁸, Lehana Thabane ^{2,7}, Peter Kaiser⁹, David Sarraf¹⁰, Sophie J. Bakri¹¹, Sunir J. Garg¹², Rishi P. Singh^{13,14}, Frank G. Holz¹⁵, Tien Y. Wong^{16,17} and Robyn H. Guymer ^{3,4}

⁸NIHR Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, UK. ⁹Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ¹⁰Retinal Disorders and Ophthalmic Genetics, Stein Eye Institute, University of California, Los Angeles, CA, USA. ¹¹Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA. ¹²The Retina Service at Wills Eye Hospital, Philadelphia, PA, USA. ¹³Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ¹⁴Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. ¹⁵Department of Ophthalmology, University of Bonn, Bonn, Germany. ¹⁶Singapore Eye Research Institute, Singapore, Singapore. ¹⁷Singapore National Eye Centre, Duke-NUD Medical School, Singapore, Singapore.