

Author Response: The Coefficient of Determination: What Determines a Useful R^2 Statistic?

We read with great interest the letter by Drs Crabb, Saunders, and Russell¹ regarding our recent publication.² In brief, they commented on the limitations and low accuracy of our model, which might make it unsuitable for use in clinical practice. We appreciate their thoughtful comments and recognize that our study is limited by the low R^2 statistic (coefficient of determination) of our model, in which the dependent variable “rates of VF change” was predicted by a set of baseline clinical independent variables using a linear model. This limitation is carefully described, elucidated, and extensively discussed in our manuscript, in which we have recommended caution when interpreting the results of our risk calculator. Nevertheless, two aspects of our analyses need to be heightened: (1) The coefficient of determination and statistical significance of the model (P value) should be examined in combination. Our model was statistically significant ($P < 0.001$), which means that our model significantly fits the data better than using the average rate of visual field change of the sample; (2) R^2 values are used to measure the goodness of fit of a linear model for the “reference” sample. It is the performance of the model in a “validation” sample (in our study we used the c-index statistic) that actually assesses its appropriateness.³⁻⁵

We employed a relatively simple statistical method that, among its disadvantages, is based on pooled (averaged) data and assumes a linear relationship between dependent and independent variables. We welcome Crabb and colleagues’ more elegant statistical approach aiming to calculate the potential range of R^2 values based on a different population—applying multiple permutations of this sample—and generating simulated reference models. The probability of gaining a better statistic than our model (R^2 equal to 0.13) was close to 35%, or, from a different perspective, 65% of the models would have a R^2 statistic lower than ours. This shows that our results were within the area of the distribution curve where most values were observed under the assumptions of their simulated model. One should be reminded, however, that the simulated model by Crabb and colleagues employed two variables as predictors, as opposed to 10 predictors in our model. Moreover, deriving reference and validation samples from the same population—as the authors did—is problematic.

We agree that the use of our calculator to guide clinical practice warrants further improvement and that the actual prediction of future visual field progression based on visual field mean deviation (MD) is partly limited by the substantial proportion of unexplained variance in the multiple linear regression model. Dr Crabb and colleagues mentioned that there is no “magic” number to define whether a linear regression coefficient of determination is “good” or “bad.” A low coefficient of determination, despite explaining a small proportion of variance, can still fit the data satisfactorily when compared with the “mean” and hence is a good alternative to be used for prediction.^{4,5} In other words, despite a low R^2 statistic, a statistically significant “ F test” for the multiple regression model that leads to a large F statistic (= model mean square/residual mean square) suggests that our linear model fits the data for predicting rates of visual field change better than the sample “mean.” Another point to be kept in mind is that in spite of the wide use of the ordinary

least square (OLS) method to calculate rates of visual field change (slopes), one should be reminded that OLS is a Best Linear Unbiased Estimator (“BLUE”) if certain assumptions hold. The assumption of homoscedasticity—which accepts that the error term has a constant variance—is violated when applied to visual field testing since there is increased variability of sensitivity as disease progresses, as recently confirmed by Dr Crabb’s group.⁶ Again, this is a limitation not only of our study, but also of other studies that aim to measure rates of visual change in glaucoma using the OLS method.

We were unable to find a measure of goodness of fit as expressed by R^2 values for the risk calculator generated from the Ocular Hypertension Treatment Study (OHTS) as they employed a statistical method (i.e., survival analysis) that does not use the OLS method. Hence, we are unable to say whether our model is “better” or “worse” than other studies in the literature. The key message, however, is that from a clinical—and not purely statistical—perspective, the definition of “appropriateness” is vague, subjective, and far from consensual.

Given the dramatic burdens of visual field loss in glaucoma and how this can affect vision-related quality of life, we believe that our model should be taken into consideration by clinicians and patients as it is the only currently available tool to predict future functional outcomes given a set of variables easily collected during baseline clinical examination. Our group is currently working on adding the 95% confidence intervals for predicted rates of change and risk (%) of progression as a means to remind those who use the calculator that our predictions can vary significantly but still provide a range of values so that clinicians can be more confident when interpreting the results for individual patients.

The objective stratification of risk of visual field progression by combining baseline clinical variables can potentially be useful in clinical management. Research that aims to provide a quantitative estimate of risk will facilitate glaucoma diagnosis and monitoring. This information may facilitate clinical judgment when added to the diagnostic and therapeutic repertoire and experience of clinicians.

As far as the laus of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

Albert Einstein (1879–1955)

Carlos Gustavo De Moraes^{1,2}

Jeffrey M. Liebmann²

Robert Ritch^{1,3}

David S. Greenfield⁴

Mitra Sebi⁴

Yun S. Chung⁴

¹Einhorn Clinical Research Center, New York Eye and Ear Infirmary, New York, New York; the ²Department of Ophthalmology, New York University School of Medicine, New York, New York; the ³Department of Ophthalmology, New York Medical College, Valhalla, New York; and the ⁴Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Palm Beach Gardens, Florida.

E-mail: gus.moraes@nyumc.org

References

1. Saunders L, Russell R, Crabb D. The coefficient of determination: what determines a useful R^2 statistic? *Invest Ophthalmol Vis Sci.* 2012;53:6830–6832.

2. De Moraes CG, Schi M, Greenfield DS, et al. A validated risk calculator to assess risk and rate of visual field progression in treated glaucoma patients. *Invest Ophthalmol Vis Sci.* 2012;53:2702-2707.
 3. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15:361-387.
 4. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Straight-line regression analysis. In: Kleinbaum DG, Kupper LL, Nizam A, Muller KE, eds. *Applied Regression Analysis and Multivariable Methods.* 4th ed. Belmont, CA: Thomson Learning, Inc.; 2008:41-90.
 5. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. Testing hypotheses in multiple regression. In: Kleinbaum DG, Kupper LL, Nizam A, Muller KE, eds. *Applied Regression Analysis and Multivariable Methods.* 4th ed. Belmont, CA: Thomson Learning, Inc.; 2008:139-163.
 6. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci.* 2012;53:5985-5990.
- Citation: *Invest Ophthalmol Vis Sci.* 2013;54:37-38.
doi:10.1167/iovs.12-11354