The Coefficient of Determination: What Determines a Useful \( R^2 \) Statistic?

We were very interested to read the recent study reported by De Moraes and colleagues1 describing the development of a validated risk calculator for visual field (VF) progression in patients with glaucoma. An accurate risk calculator, available at the point of glaucoma diagnosis, would potentially be beneficial for clinicians. De Moraes et al. have, thus, developed two models using the retrospective New York Glaucoma Progression Study to assess the probability of patient progression and the expected rate of this progression, validating their model using patient data from the Diagnostic Advanced Imaging in Glaucoma Study. We now take the opportunity to expand on the uncertainty touched on by the authors regarding an adjusted \( R^2 \) of 0.13 associated with their rate calculator and, in particular, demonstrate that a model with a statistic of this magnitude is inappropriate for use as a predictive tool.

The well-known \( R^2 \) statistic, or the (multiple) coefficient of determination, pertains to the proportion of variance in the response variable explained by a fitted model relative to simply taking the mean of the response. In other words, it describes how well the model fits the data. An \( R^2 \) close to 1 implies an almost perfect relationship between the model and the data, whereas an \( R^2 \) close to 0 implies that just fitting the mean is equivalent to the model fitted. Often, more than one variable is available to explain an outcome (multivariate model); in this situation, as variables are added, the \( R^2 \) will increase even if the variable is not important. The adjusted \( R^2 \) attempts to correct for this by penalizing for increasing the number of variables, so is often preferred in comparing models. Unfortunately, there are no set criteria as to what universally represents a “good” \( R^2 \) value, so the only way of assessing such a statistic is via comparison with another predictive model. To this end, we designed an alternative, purely illustrative rate calculator using only the patient age (which has been shown to be related to rate of VF loss5) and the patient’s first two VF values. The De Moraes calculator used nine different variables, including peak intraocular pressure, central corneal thickness, the presence of an optic disc hemorrhage, the presence of exfoliation or beta-zone parapapillary atrophy, and whether the patient had or did not have glaucoma surgery.

To construct our model 68,099 anonymized VFs collected from 8252 anonymized patients visiting the Glaucoma service at Moorfields Eye Hospital between 1997 and 2009 using the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA; 24-2 test pattern, Goldmann size III stimulus and SITA Standard testing algorithm) were used. Data were examined in accordance with the Declaration of Helsinki. Patients with fewer than four VFs (per eye), once the first VF was removed to account for learning effects, were excluded. Furthermore, the interval between the first and second VFs was restricted to the range 3 to 13 months and this interval was not allowed to exceed 40% of the overall follow-up time. VF tests with false positive or false negative rates above 30% or fixation losses greater than 20% were discarded. Where both eyes of a patient were eligible, an eye was selected at random, leaving 875 patients (875 eyes) for investigation. The demographics of our study were comparable to the reference data set in the De Moraes study, but their validation data set contained patients with lower magnitude and variability of damage, as can be seen in the Table.

The “true” rates of progression for each patient were calculated using ordinary least squares regression of mean deviation (MD) over time, the same method as that used in the De Moraes study. In our model we included the difference between MDs in the second and first VFs divided by the time interval separating them. The “true” rates of progression were then regressed against the baseline age of patients and their VF status across two visits as a basic model from which adjusted \( R^2 \) values were generated. To facilitate comparisons, the study sample was split into a reference data set and a validation data set comprising exactly the same numbers of patients as included in the De Moraes study (i.e., 587 patients in the reference data set and 62 patients in the validation data set). To gain a distribution of values for the adjusted \( R^2 \) statistic, the 875 patients were randomly sampled without replacement 100,000 times into reference and validation data sets, to attain 100,000 adjusted \( R^2 \) values for each model (i.e., for the reference data set, 587 patients were selected at random from the 875 patients in our complete data set, and 62 patients were sampled from the remaining 288 patients, a process that was repeated 100,000 times). The distribution of adjusted \( R^2 \) values for the 100,000 reference models can be seen in Figure A. The median adjusted \( R^2 \) is 0.10, whereas the reported \( R^2 \) for the De Moraes calculator lies at the 87th percentile (0.13). However, given that the reported \( R^2 \) could actually have taken any value between 0.125 and 0.135, the possibility of getting this statistic by chance in our data set could, in fact, be as high as 20%. Figure B shows the adjusted \( R^2 \) statistic yielded when the reference model is fitted to the validation data set. The median adjusted statistic here is 0.08, but the spread of this distribution should be noted; it was possible to simulate an adjusted \( R^2 \) statistic as high as 0.59 (due to the small sample size). The probability of gaining a better statistic than that of the De Moraes model (\( R^2 = 0.11 \)) was close to 35%. We selected one reference model from our distribution in Figure A with an \( R^2 \) value similar in magnitude to that of the De Moraes rate calculator. The fit of this model can be seen in Figure C, whereas Figure D shows the effect of applying this model to a sample validation data set (once again sampled to match the \( R^2 \) in the model reported by De Moraes et al.); the 95% limits of agreement are shown by the dotted lines in Figure D, and are more reflective of the likely range of differences between the

---

Table. A Comparison of the Demographics of Data Sets Used in the De Moraes Study and Our Sample

<table>
<thead>
<tr>
<th></th>
<th>New York Glaucoma Progression Study</th>
<th>Advanced Imaging for Glaucoma Study</th>
<th>Moorfields Glaucoma Service Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>587</td>
<td>62</td>
<td>875</td>
</tr>
<tr>
<td>Age at baseline (y)</td>
<td>64.9 ± 13.0</td>
<td>67.4 ± 8.3</td>
<td>62.7 ± 13.0</td>
</tr>
<tr>
<td>Baseline mean deviation (MD) (dB)</td>
<td>−7.1 ± 5.1</td>
<td>−3.7 ± 4.4</td>
<td>−7.0 ± 5.3</td>
</tr>
<tr>
<td>Follow-up time (y)</td>
<td>6.4 ± 1.7</td>
<td>4.0 ± 0.9</td>
<td>5.8 ± 1.7</td>
</tr>
</tbody>
</table>

---

1. De Moraes et al. (2012)
estimated and actual rates of progression than the 95% confidence interval for the average difference (indicated by the dashed lines) reported in the abstract of the De Moraes et al. study. Figures C and D clearly demonstrate the inadequacy of our model, designed to mirror that of the De Moraes study, for predicting rates of VF loss in spite of statistical significance. We have tried to create a context for the De Moraes calculator's ability to estimate rate of VF loss, but realize there are several limitations in our comparison. For a start, the "validation" data set used was not a different data set from a different clinical center and these patients had worse average MD than those used in validating the De Moraes calculator. MDs of greater severities have been shown to be more variable than healthy ones, which suggests that rates of loss will be predicted less accurately in patients with more advanced VF damage. Thus, the characteristics of the De Moraes validation data set may lead to more favorable results, because, like ours (see Fig. D), their calculator more accurately models patients with slower rates of VF loss. An advantage of the De Moraes calculator is that, unlike our model, it has used measurements that can be taken at the first visit, although the inclusion of glaucoma surgery as a baseline predictive variable is controversial in this context. Furthermore, given the large model coefficient associated with surgery it would be interesting to see how their rate calculator would perform without this information.

$R^2$ statistics are often well understood and correctly interpreted, but can also be misleading, given that the precision of the statistic is dependent on sample size and the coefficient is commonly presented without confidence intervals or limits of tolerance. Without a sense of comparison the adjusted $R^2$ statistic is limited in its usefulness and it is apparent that there is, as yet, no reference standard for testing rate calculators such as this. Moreover, Figures C and D suggest that rate calculators with small $R^2$ values are inadequate for accurately predicting rates of loss, especially in patients with

---

**Figure.** The upper plots are histograms showing the distribution of adjusted $R^2$ values from 100,000 (A) simulated reference models and (B) simulated reference models fitted to simulated validation data sets. The black bars represent the potential $R^2$ values found by De Moraes et al., given their published results. (C) A plot of the estimated progression rate of patients in the selected reference data set against their "true" rate of progression. The solid line represents exact correspondence between the estimated and actual rates of progression (i.e., the line of unity). The $R^2$ is equal to 0.14 for this model. (D) A Bland-Altman plot comparing the progression rates of a selected validation data set to the rates estimated by model C ($R^2 = 0.12$). The 95% mean confidence interval (dashed lines) is 0.14 to $-0.11$ dB per year. However, the more informative 95% limits of agreement (dotted lines) range from $-1.0$ to $1.02$. As in the model reported in the study by De Moraes et al., the fit is worse at larger rates of progression.
fast progression, who are the most at risk of visual impairment. Finally, it is very important to emphasize that our illustrative rate calculator is not a serious attempt at introducing an alternative modeling strategy and should not be used to estimate rate of VF loss, yet it still provided predictive accuracy similar to that of the De Moraes calculator. De Moraes and colleagues should be commended for their novel attempt at developing a statistical model for predicting progression in patients with treated glaucoma and especially for attempting to validate it using independent patient data. However, the conclusion that must be drawn is that the limitations and low accuracy of their model make it unsuitable for clinical practice.

Luke Saunders
Richard Russell
David Crabb

1Department of Optometry and Visual Science, City University London, United Kingdom; 2National Institute for Health Research, Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital, National Health Service Foundation Trust, and University College London Institute of Ophthalmology, London, United Kingdom. E-mail: david.crabb.1@city.ac.uk

Supported in part by the UK National Institute for Health Research (NIHR) Health Services Research Programme (project number 10/2000/68). David Crabb’s research laboratory at City University in London is supported in part by unrestricted funding from Allergan Ltd.

References


Citation: Invest Ophthalmol Vis Sci. 2012;53:6830–6832. doi:10.1167/iovs.12-10598