Interpreting regression models in clinical outcome studies

Citation for published version:

Digital Object Identifier (DOI):
10.1302/2046-3758.49.2000571

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Bone & Joint Research

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
EDITORIAL

Interpreting regression models in clinical outcome studies

D. F. Hamilton, M. Ghert, A. H. R. W. Simpson

From The British Editorial Society of Bone and Joint Surgery, London, United Kingdom

Measuring the outcome of an intervention is central to the practice of evidence based medicine, and most research papers evaluating patient outcomes now incorporate some form of patient-based metric, such as questionnaires or performance tests. Once an outcome has been defined, researchers typically want to know if any other factors can influence the result. This is typically assessed with regression analysis.

Regression analysis determines the relationship of an independent variable (such as bone mineral density) on a dependent variable (such as ageing) with the statistical assumption that all other variables remain fixed. The calculation of the relationship results in a theoretical straight line, and the correlation co-efficient (r) measures how closely the observed data are to the theoretically straight line that we have calculated.

In such a linear model, we can judge how well the line fits the data (‘goodness of fit’) by calculating the coefficient of determination (or square of the regression line, R^2). R^2 is a measure of the percentage of total variation in the dependant variable that is accounted for by the independent variable. An R^2 of 1.0 indicates that the data perfectly fit the linear model. Any R^2 value less than 1.0 indicates that at least some variability in the data cannot be accounted for by the model (e.g., an R^2 of 0.5 indicates that 50% of the variability in the outcome data cannot be explained by the model).

How should we interpret these findings? The authors quite correctly suggest that models such as these could be of value in discharge planning and resource utilisation by targeting the patients that most need intervention and rehabilitation. The reported R^2 for the models, however, was 0.18. Bearing in mind that R^2, the coefficient of determination, measures the percentage of the variation in the dependent variable that is explained by variation in the independent variable, taking the compliment (100 – R^2) we see that 82% of the variation in the outcome parameter assessed is unexplained by the model. The principal problem is that the variance in the population studied can strongly influence R^2 magnitude. Therefore, there is no guarantee that a high coefficient of determination is indicative of ‘goodness of fit’. Similarly there is no guarantee that a small R^2 indicates a weak relationship, given that the statistic is largely influenced by variation in the independent variable.
Therefore, there is no rule for interpreting the strength of $R^2$ in its application to clinical relevance. Useful high values of $R^2$ can be obtained with clinical data sets, however, a low $R^2$ can still provide a useful clinical model with respect to data trends, but may be low in precision. In this study there is an association between the performance tests and length of stay; and, using the equations, we can indeed predict one from the other. The accuracy of this prediction though, needs to be borne in mind when using it as a clinical tool.

Furthermore, it is not rational to compare $R^2$ across different samples, which given clinical populations, are likely to differ significantly in the variance of the independent and dependent variables.

In controlled environments, such as biomechanical tests on cadaveric bones, the variance across predictive measurements is likely to be low, and therefore $R^2$ values can be expected to lie in the 0.8 range. In clinical studies, however, $R^2$ values vary widely depending on the nature of the analysis. For example, when comparing radiographic parameters or associating surgical technical factors, values of $R^2$ are reported in the 0.2 to 0.4 range. Whereas, comparing data between separate (but intrinsically similar) outcome assessment questionnaires can yield higher values in excess of 0.7.

As such, further validation of the Poitras study using new datasets and, ideally, confirmatory analysis of the findings using a much larger sample size, would be required before their regression model could be recommended for use clinically. This does not devalue the appropriateness — or indeed ‘worthiness’ — of reporting these findings in the literature, as the important clinical tools typically start as ideas in small datasets. As with all research papers, the reader requires a basic understanding of methodology to evaluate how relevant the results are to wider practice.

References