

Interpreting Poisson Regression Models in Dental Caries Studies

Alex Man Him Chau Edward Chin Man Lo May Chun Mei Wong Chun Hung Chu

Faculty of Dentistry, University of Hong Kong, Hong Kong SAR, China

Keywords

Data mining · Dentistry · Epidemiology · Oral health · Regression analysis · Statistical data interpretation · Statistical model

Abstract

Oral epidemiology involves studying and investigating the distribution and determinants of dental-related diseases in a specified population group to inform decisions in the management of health problems. In oral epidemiology studies, the hypothesis is typically followed by a cogent study design and data collection. Appropriate statistical analysis is essential to demonstrate the scientific association between the independent factors and the target variable. Analysis also helps to develop and build a statistical model. Poisson regression and its extensions have gained more attention in caries epidemiology than other working models such as logistic regression. This review discusses the fundamental principles and basic knowledge of Poisson regression models. It also introduces the use of a robust variance estimator with a focus on the “robust” interpretation of the model. In addition, extensions of regression models, including the zero-inflated model, hurdle model, and negative binomial model, and their interpretation in caries studies are reviewed. Principles of model fitting, including goodness-of-fit measures, are also discussed. Clinicians and researchers should pay attention to the statistical context of the models used and interpret the models to improve the oral and general health of the communities in which they live.

© 2018 S. Karger AG, Basel

Epidemiology involves studying and investigating the distributions and determinants of a disease in a specified population group to make informed decisions in controlling health problems [Last and International Epidemiological Association, 2001]. Oral epidemiologists are no exception to this definition. In oral epidemiology studies, the hypothesis is typically followed by a cogent study design and data collection. Statistical analysis is needed to demonstrate the scientific association between the independent factors and the target variables. Point and interval estimates of important variables often serve as the primary outcome of a study and guide the study design such as sample size determination [Petrie et al., 2002b]. Oral epidemiologists are also interested in any factors that show an association with the primary outcome. Socio-economic factors, for example, are often included in caries research. Oral health status has been shown to be associated with social gradients in many studies across the world [Sheiham et al., 2011]. The relationship between health outcomes and these independent factors may be described in many approaches. A bivariate relationship may be presented in tables, charts, curves, and even pictograms, and there are many visualization tools available [Chen et al., 2008]. However, multivariate relationships are more difficult to visualize on a piece of paper or a computer monitor.

Statistical models serve as a language and means of presentation to demonstrate the multivariate relationship of a study by considering all relevant (but not necessarily significant) factors, all at the same time. It is necessary for researchers to understand the assumptions of construct-

ing statistical models in order to validly interpret them. They should be familiar with the nature of such models for a critical review of the evidence generated. In fact, many syllabi of a postgraduate clinical programme involve critical appraisal of academic articles, including the use of proper statistical analysis and modelling. Consequently, the proper application and interpretation of statistical models are essential for both researchers and readers to elicit the complex relationship between the target health outcomes and the covariates.

Statistical models should be accurate enough for the task of the researcher [Krzanowski, 1998]; hence, models should serve to reach the objectives of the oral epidemiological studies. In statistical terms, models may serve 3 main objectives [Shmueli, 2010]. Descriptive modelling aims at summarizing data, focuses on the measurement level and looks for associations between variables. Explanatory modelling focuses on the construct level of a theoretical model and looks for causal relationships. Predictive modelling aims at predicting future observations from the input values. When a researcher applies a statistical model in a caries study, it is important to match the objectives on both the statistical and the dental sides.

Poisson regression and its extensions have gained more attention in caries epidemiological studies than other working models such as logistic regression. In this paper, the basic knowledge on regression modelling is reviewed. This paper also discusses the use of the robust variance estimator, with a focus on the “robust” property of the model. Extensions including the zero-inflated model, hurdle model, and negative binomial (NB) are introduced. The paper also discusses measures of goodness of fit and their interpretation in the model selection process.

Poisson Regression Models in Generalized Linear Modelling

In generalized linear modelling, if y is a dependent variable (number of men with dental caries) and x is a collection of independent variables such that $x_1, x_2 \dots x_k$ are k different independent variables, a multivariate linear regression model is defined as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (1)$$

where $\beta_0, \beta_1 \dots \beta_k$ are coefficients of their corresponding independent variables $x_1, x_2 \dots x_k$, and ε is the error term, because this model is almost impossible to obtain all of the y values exactly by considering only β and x values. There are many ways to obtain an estimator of all the

β values. One common way is to minimize the squared value of the error term, i.e. ε^2 , which is known as the ordinary least square method. Apart from specifying the model, there are several important assumptions that a linear regression model must hold for a valid inference of the β . These assumptions may be stated in different ways, but essentially these 4 classical descriptions include most of the practical situations: (1) linearity and additivity between x and y , (2) homoscedasticity ε , (3) normally distributed ε against y , and (4) independence of ε and y (free from autocorrelation).

It is well known that these assumptions may not always be satisfied. For example, the absolute risk of caries of a person should never exceed 1. A person who has a high risk (determined by all independent variables in the model) may have an estimated caries risk greater than 1. This is called the convergence problem, and it may suggest, at least partly, a violation of assumption 1. To satisfy the assumption requirements, some pretreatment on the variables may be done. One possible method is to transform y into another variable, say z , so that a linear regression modelling in z satisfies all of the required assumptions. The corresponding coefficients ($\beta_1, \beta_2 \dots \beta_k$), which are originally modelled for z , have to be transformed back properly to a model for y for interpretation.

The method of transformation from y to z is called the link function. In Poisson regression, the link function is $\log(y)$. Hence:

$$z = \log(y) \quad (2)$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon. \quad (3)$$

To estimate the β values, a process called maximum likelihood estimation (MLE) may be used. The concept behind MLE is to answer the question: “Given the distribution of y and the values of all independent variables (x), what are the most likely β coefficients?” Suppose we collected samples from 200 participants in a study of the number of carious teeth in a man, denoted by y , following a Poisson distribution. Patient level factors, such as age, gender, and monthly income, were collected as independent variables (x). MLE required us to write down, then maximize, a mathematical expression to describe the probability of obtaining such a sample by using a multivariate Poisson distribution, which was derived from equation 13 shown below. Since all the samples were collected at the same time, the required expression was a multiplication of all 200 expressions.

One reason for the popularity of Poisson regression over logistic regression is direct modelling of the relative risk (RR) [Zou, 2004]. Logistic regression uses the link

function of $\log(p/1 - p)$, the log odds ratio which is approximately equal to the RR only if p is small [Schmidt and Kohlmann, 2008]. As we can see from above, the link function allows direct modelling of $\log(\text{RR})$ using Poisson regression. The probabilistic interpretation, however, does not make direct medical sense because the “rate” of having new caries in a particular time period is $\log(\text{RR})$. Derivation of Poisson distribution from the event rate concept is illustrated in the Appendix.

A Closer Look at Poisson Regression with Robust Variance Estimator

The use of Poisson regression with robust variance estimator (PRRVE) in oral epidemiological studies has gained popularity in recent years. The prevalence ratios of caries and fluorosis have been modelled against socio-economic factors [Do et al., 2014]. Through the use of hierarchical PRRVE, oral health-related quality of life scores were found to be associated with number of untreated caries and missing teeth [Feldens et al., 2016]. The prevalence ratio of dental pain was also associated with ethnicity and low socio-economic class [Boeira et al., 2012]. In addition, risk ratio values of 2 caries diagnostic criteria were compared using dmft count and caries prevalence [Mendes et al., 2010]. Other published articles have shown that the dependent factors modelled with PRRVE included tooth pain experience [Ferraz et al., 2014] and sense of coherence level (high or low) [Lage et al., 2017].

The rationale of PRRVE can be traced back to a seminal paper published half a century ago [Huber, 1967]. The paper stated that MLE of β coefficients is consistent even when the working distribution is not true, meaning that MLE is close to the true mean when the sample size is large enough, regardless of what distribution is being worked with. This concerned the mean of the estimator (point estimation). The variance of the estimator (interval estimation) was, however, investigated much later. Instead of using the moment-generating method to derive the exact expression for variance, deriving the variance estimator from another mathematical expression, called the Fisher information, may adjust the confidence interval to be asymptotically valid even when the working distribution is not true [Royall, 1986].

A modified Poisson regression approach was proposed about 2 decades after Royall’s work [Zou, 2004]. Zou used a logarithmic link function to transform the RR into a generalized linear modelling model, using Poisson distribution as the working model. While RR is actually

binomially distributed, the Poisson regression model is actually misspecified. Using Huber’s [1967] and Royall’s [1986] result, the asymptotic mean and variance estimators were derived [Zou, 2004]. A comparative study on different ways of modelling prevalence ratios showed that PRRVE provides correct estimates, superior to logistic regression [Barros and Hirakata, 2003].

The indiscriminating use of the robust technique has not been without controversy. Freedman [2006] warned that model misspecification might exist in factor and model levels. Specifically, the term “robustness” in PRRVE refers to the robustness against selecting the wrong working model, but not against including the wrong factors in the model. Freedman provided a simple example showing that the exclusion of a correct independent variable will lead to incorrect estimation of the remaining parameters in the model which is not alleviated by the adoption of the robust variance estimation technique [Freedman, 2006]. Barros and Hirakata [2003] also warned that PRRVE suffers from a convergence problem and exhibited problems at the extreme values of the independent variables, especially if the model includes any continuous independent variables.

Zero-Inflated Poisson Model, Hurdle Model, and NB Model

Some of the dependent variables in dental studies, such as the distribution of caries experience scores measured in dmft or DMFT in a group of samples, involves a significant proportion of zeros. Poisson regression, as seen in the Appendix, naturally has an equal mean and variance, hence the Poisson regression model is not flexible enough to fit a high number of zeros. There are many types of extensions from the Poisson model to accommodate this problem. Zero-inflated Poisson (ZIP) regression is one of the most popular models in oral epidemiological studies. More specifically, a new parameter (ω) is introduced to entertain the high proportion of zeros:

$$P(Y = 0) = \omega + (1 - \omega)e^{-\lambda}. \quad (4)$$

$$P(Y > 0) = (1 - \omega) \frac{e^{-\lambda} \lambda^y}{y!}. \quad (5)$$

The mean (μ) is equal to $(1 - \omega)\lambda$ and the variance

$$\mu + \left(\frac{\omega}{1 - \omega} \right) \mu^2 \quad (6)$$

[Ridout et al., 1998]. Estimating the new parameter ω may need a different set of covariates independently from those estimating λ , or possibly using the same covariates. By considering

$$\log\left(\frac{\omega}{1-\omega}\right) = -\tau \log(\lambda), \quad (7)$$

ω can be estimated by the same set of covariates as λ , and τ is referred to the scale parameter [Lambert, 1992]. In a ZIP model, the probability of y being zero came from 2 sources, namely the parameter ω and the zero part from the Poisson distribution $(1-\omega)e^{-\lambda}$. Some researchers have interpreted the former source as “structural zeros” and the latter as “sampling zeros” [Hu et al., 2011]. Other terms have included “excessive zeros” after the estimated zeros from the working distribution [Preisser et al., 2016] and “non-susceptible group” and “susceptible group” in the context of modelling health data [Preisser et al., 2012]. Interpreting ZIP models may be confusing when dental terminology is applied, such as defining “at-risk population” and “caries severity” [Preisser et al., 2012].

Hurdle models provide an alternative for flexibility by modelling all the zeros separately from the non-zeros. A hurdle level is determined first, usually at zero. All sample data are then categorized based on whether the dependent variables exceed the hurdle level, followed by constructing a zero-truncated working model for those past the hurdle. Without the loss of generality, the term “hurdle model” in this paper refers to the zero hurdle model. For example, the Poisson hurdle model [Ridout et al., 1998] may be represented as:

$$P(Y=0) = \pi_0 \quad (8)$$

$$P(Y>0) = \frac{(1-\pi_0)e^{-\lambda}\lambda^y}{(1-e^{-\lambda})y!}. \quad (9)$$

Apart from zero-inflated and hurdle models, the mean of the Poisson distribution, λ , may be modelled by another distribution instead of being a constant. Gamma distribution is a popular choice. Instead of modelling the data directly with gamma distribution, the mean of a Poisson distribution may be modelled by a gamma distribution with the result being called NB distribution (see the Appendix). The basic concept of gamma distribution is modelling the time required for the α -th event to happen in a Poisson process with a rate of β per unit time. NB distribution models the number of selected samples, answering the question “what is the probability that, when the x -th sample is taken from the population, the exactly

k -th desired sample is selected?” NB distribution has a mean of $\alpha\beta$ and variance of $\alpha\beta^2$, allowing NB to be more flexible than the Poisson distribution. The formula used for regression differs slightly from the distribution by including the mean term in the model specification [Hilbe, 2011]. The reason for using other mixture models, such as the β -binomial distribution, becomes clear. These models also have zero-inflated and hurdle versions for further flexibility to fit the data. Finally, a model that may have other-than-zero inflation part, with an unlimited number and types of models is called a generic mixture model, and it provides an outstanding flexibility to fit the data [Gilthorpe et al., 2009].

Justification and Interpretation of the Models in Caries Researches

At least 2 levels can be discussed when a researcher is justifying a model – model level and factor level. On the model level, one of the most traditional goodness-of-fit measures, the coefficient of determination R^2 , describes how likely the independent variable is to explain the dependent variable in a linear relationship in multivariate linear regression. It is numerically equal to the fraction of variance explained [Miller and Miller, 2015]. A number of pseudo- R^2 measures have been developed for other types of regression, such as Nagelkerke R^2 [Nagelkerke, 1991] for logistic regression. Pseudo- R^2 measures do not usually represent the fraction of variance explained [Mittlböck and Schemper, 1996], but they may still be interpreted to explain the proportion of variation [Nagelkerke, 1991].

Goodness-of-fit measures developed from information theory have gained more attention on model selection in oral epidemiological studies published in recent years. The use of information theory in statistical modelling should be studied for proper application and interpretations. Suppose there are many messages to be sent from one location to another (A to B). Each message is written in a language with the number of available “codes” equal to n . Suppose it is not possible to directly transmit the message but only strings of binary digits, 0 or 1, known as “bits.” The mean number of bits (string length) required to transmit English messages ($n = 26$), is therefore $\log_2 26$ or approximately 4.7 or 5 to the nearest larger integer. One natural choice of coding policy may be a one-to-one arrangement, meaning letter A is encoded as 00001 and Z as 11010, for instance. It is also known that not all letters are used with equal frequency. To reduce

the string length, the coding system may allow a shorter string for frequently used letters without messing up other codes. Let p_i be the probability of the i -th letter being used in a message, Shannon [1948] defined information entropy as:

$$H = - \sum_{i=1}^n p_i \log_2 p_i, \quad (10)$$

which is also equal to the minimal mean string length required. Assume that the true values of p are not known, but another set of approximately equal probabilities can be obtained somehow (for example, by sampling some messages), denoted by q . Obviously, a coding policy using q will result in longer strings for the messages in long run. The mean string length using the coding system with q is equal to:

$$- \sum_{i=1}^n p_i \log_2 q_i, \quad (11)$$

which is defined as cross entropy [Good, 1963]. The Akaike information criterion can be understood as the criterion that the choice of parameters of a model should minimize the expected cross entropy of the MLE relative to the “true” model of the data [Akaike, 1998]. Other information criteria may be applied in different modelling situations, and they usually extend the Akaike information criterion in their respective contexts. For example, the Bayesian information criterion was developed from behaviours of Bayes estimators [Schwarz, 1978]. The quasi-likelihood information criterion was developed for generalized estimating equations, in which no “true” model existed [Pan, 2001]. These information criteria describe how close the estimated values are from the statistical model to the observed. Concerning Shmueli’s [2010] objectives of statistical modelling, these information criteria measures should be interpreted as the descriptive power of the models given the data set collected by the researchers.

In caries researches, however, goodness-of-fit measures are not routinely reported. Among 15 papers on childhood dental caries modelled with ZIP or zero-inflated NB that were published in or before 2011, over half did not report on goodness of fit [Preisser et al., 2012]. We also noticed that none of the studies mentioned in the PRRVE section above reported any model fit measures. Dental researchers are generally less interested in model fit and model structures, but more so in factors in overall effects [Preisser et al., 2012]. The probability interpretation of the constant rate across time in Poisson regression (see the Appendix) and NB regression, regardless of the Poisson- γ

mixture or waiting time perspective, is seldom discussed in the dental literature. One recent study proposed that hurdle models were easier to interpret than zero-inflated models [Hofstetter et al., 2016]. Another study proposed that the “spell” parameter in the generalized NB model may match the biological development of dental caries [Vergnes et al., 2017]. These 2 studies are examples of the struggle to find a good common ground for both statistical model reasoning and dental interpretation.

On the factor level, independent variables may be either included or excluded through a prescribed process, which may be known as subset selection [Hastie et al., 2009]. Backward, forward, and stepwise methods are common in oral epidemiological studies. This method has been criticized for letting the computer select the relevant factors without considering their scientific values [Greenland, 2000]. It is also interesting that the stopping criteria for determining the final model are arbitrary. For example, some studies chose the final model by including all independent variables with their p values less than a predetermined value, such as 0.20 [Marques et al., 2017], or maximizing some model fit criteria, such as the adjusted R^2 value [Petrie et al., 2002a] or χ^2 statistic in logistic regression [Lee and Koval, 1997]. Researchers should note that there is no guarantee that the subset selection procedure will yield the best model from all possible combinations of factors [Hastie et al., 2009], but only a “reasonably good” model.

The best-fit model is not always interpreted as the best model. Using Shmueli’s [2010] classification, the objectives of many caries researches included in this review may be described as descriptive. Explanatory models using factor analysis and structural equation models are not uncommon. Predictive models are more commonly applied in patient level caries risk assessment, with techniques including the neural network and support vector machine, but are less common in community level studies. To serve these objectives, it is tempting to use more advanced, more flexible, more complicated models for a better fit of the data in hand. Restricted cubic spline models [Liang et al., 2016], for example, allow a more flexible model shape. On the other hand, some researchers prefer more parsimonious models. Principal component analysis [Shaffer et al., 2013], for example, aims at reducing independent factors by creating new variables which capture the features of these factors better. While adopting a better-fit model is not wrong per se, researchers must pay attention to the possible overfit problem when more advanced and complex modelling techniques are applied. If the interpretation of a complex model does not match

with its statistical meaning within its mathematical structure, the added complexity could be misused for its flexibility, and problems including overfit may arise [Hastie et al., 2009].

Conclusion

There are many working distributions and modelling techniques available for modelling caries distribution in a community. This paper presents PRRVE, zero-inflated models, and hurdle models, and acknowledges their assumptions, pros, and cons which the applied statisticians must pay attention to. Some techniques for model selection are presented, on both the model level (such as goodness of fit) and factor level (such as subset selection). The interpretation of statistical models in caries research is largely emphasized on the factor level and seldom on the model level. With more advanced statistical models developed from time to time, researchers must be cautious about possible statistical and dental misinterpretation.

Appendix

Assume that we have to describe a probability model as follows: n men were sampled and examined in an oral epidemiological study. Assume that each man has the same probability p of having at least 1 carious tooth. The probability of a man being caries free is therefore $1 - p$, and the number of caries-free teeth is $n - x$. The probability that there are exactly x men in our sample having at least 1 carious tooth can be said to follow a binomial distribution:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}. \quad (12)$$

The mean and the variance of these distributions are np and $np(1 - p)$, respectively.

With an assumption of a small p , equation 12 can be approximated to be a Poisson distribution:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad (13)$$

where $\lambda = np$. Although there is no consensus on how small p can be, some introductory statistics text may recommend $p < 0.05$ if $n > 20$ [Miller and Miller, 2015]. Since p is small, $1 - p$ tends to be equal to 1. Therefore, the mean and variance of the Poisson distribution are both equal to λ , which can also be proved using a mathematical derivation such as the moment-generating function. The parameter λ in the Poisson distribution can be interpreted as the number of events happening per unit time, meaning that an event happening in an expected rate of λ .

Now we turn our interest to the time required until k events have happened. The required probability mass function is:

$$f(t) = 1 - P(X \leq k - 1) = 1 - \sum_{x=0}^{k-1} \frac{(\lambda t)^x e^{-\lambda t}}{x!}. \quad (14)$$

Differentiating both sides with respect to t results in the probability density function:

$$f(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \quad (15)$$

where $\Gamma(k) = (k - 1)!$. This expression is said to follow gamma distribution. There are at least 2 ways to define gamma distribution:

$$g(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha}. \quad [\text{Miller and Miller, 2015}] \quad (16)$$

or,

$$g(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}. \quad [\text{Dobson and Barnett, 2008}] \quad (17)$$

Hence, the probability density function of the Poisson waiting time follows gamma distribution with $\alpha = k$ and $\beta = 1/\lambda$ if the first definition is adopted, or the same α but $\beta = \lambda$ if the second definition is adopted. The first definition was adopted in this paper for the following derivation.

If λ of a Poisson distribution is modelled by a gamma distribution with parameters α and β , the joint probability density distribution function is:

$$P(X = x) g(\lambda) = \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) \left(\frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha) \beta^\alpha} \right). \quad (18)$$

The required marginal probability density distribution function:

$$\begin{aligned} P(X = x) &= \int_0^\infty \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) \left(\frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}}{\Gamma(\alpha) \beta^\alpha} \right) d\lambda \\ &= \binom{x + \alpha - 1}{x} \left(\frac{\beta}{1 + \beta} \right)^x \left(1 - \frac{\beta}{1 + \beta} \right)^\alpha, \end{aligned} \quad (19)$$

which is said to be an NB distribution with mean $\mu = \alpha\beta$ and variance $\alpha\beta + \alpha\beta^2 = \mu + \mu^2/\alpha$.

So far we have only considered the probability density functions of Poisson, gamma and NB distributions. The regression models for those distributions are different, but more complicated and less intuitive. One point to note for the derivation of an NB regression model is that the gamma distribution is assumed to be 1 parameter with mean equal to 1, meaning $\alpha\beta = 1$ and variance = $\alpha\beta^2 = \beta(\alpha\beta) = \beta$, or $\beta = \kappa = 1/\theta$ in Greene [2008] and $\beta = \nu$ in Hilbe [2011], with the name “scale parameter.”

Disclosure Statement

All authors and co-authors have no conflicts of interest to declare.

Author Contributions

All authors participated in designing and writing the paper.

References

- Akaike H: Information theory and an extension of the maximum likelihood principle; in Parzen E, Tanabe K, Kitagawa G (eds): *Selected Papers of Hirotugu Akaike*. New York, Springer, 1998, pp 199–213.
- Barros AJ, Hirakata VN: Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol* 2003;3:21.
- Boeira GF, Correa MB, Peres KG, Peres MA, Santos IS, Matijasevich A, Barros AJ, Demarco FF: Caries is the main cause for dental pain in childhood: findings from a birth cohort. *Caries Res* 2012;46:488–495.
- Chen C-H, Härdle W, Unwin A: *Handbook of Data Visualization*. Berlin, Springer, 2008.
- Do LG, Miller J, Phelan C, Sivanewaran S, Spencer AJ, Wright C: Dental caries and fluorosis experience of 8–12-year-old children by early-life exposure to fluoride. *Community Dent Oral Epidemiol* 2014;42:553–562.
- Dobson AJ, Barnett AG: Exponential family and generalized linear models; in Dobson AJ, Barnett AG (eds): *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC Texts in Statistical Science Series. Boca Raton, CRC Press, 2008, pp 45–58.
- Feldens CA, Ardenghi TM, Dos Santos Dullius AI, Vargas-Ferreira F, Hernandez PA, Kramer PF: Clarifying the impact of untreated and treated dental caries on oral health-related quality of life among adolescents. *Caries Res* 2016;50:414–421.
- Ferraz NK, Nogueira LC, Pinheiro ML, Marques LS, Ramos-Jorge ML, Ramos-Jorge J: Clinical consequences of untreated dental caries and toothache in preschool children. *Pediatr Dent* 2014;36:389–392.
- Freedman DA: On the so-called “Huber sandwich estimator” and “robust standard errors.” *Am Stat* 2006;60:299–302.
- Gilthorpe MS, Frydenberg M, Cheng Y, Baelum V: Modelling count data with excessive zeros: the need for class prediction in zero-inflated models and the issue of data generation in choosing between zero-inflated and generic mixture models for dental caries data. *Stat Med* 2009;28:3539–3553.
- Good IJ: Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann Math Statist* 1963;34:911–934.
- Greene W: Functional forms for the negative binomial model for count data. *Econom Lett* 2008;99:585–590.
- Greenland S: When should epidemiologic regressions use random coefficients? *Biometrics* 2000;56:915–921.
- Hastie T, Tibshirani R, Friedman JH: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ed 2. New York, Springer, 2009.
- Hilbe JM: *Negative Binomial Regression*. Cambridge, Cambridge University Press, 2011.
- Hofstetter H, Dusseldorp E, Zeileis A, Schuller AA: Modeling caries experience: advantages of the use of the hurdle model. *Caries Res* 2016;50:517–526.
- Hu MC, Pavlicova M, Nunes EV: Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *Am J Drug Alcohol Abuse* 2011;37:367–375.
- Huber PJ: The behavior of maximum likelihood estimates under nonstandard conditions; in LeCam LM, Neyman J (eds): *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, University of California Press, 1967.
- Krzanowski WJ: *An Introduction to Statistical Modelling*. London, Arnold, 1998.
- Lage CF, Fulgencio LB, Corrêa-Faria P, Serra-Negra JM, Paiva SM, Pordeus IA: Association between dental caries experience and sense of coherence among adolescents and mothers. *Int J Paediatr Dent* 2017;27:412–419.
- Lambert D: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992;34:1–14.
- Last JM; International Epidemiological Association: *A Dictionary of Epidemiology*, ed 4. New York, Oxford University Press, 2001.
- Lee KI, Koval JJ: Determination of the best significance level in forward stepwise logistic regression. *Commun Stat Simul Comput* 1997;26:559–575.
- Liang JJ, Zhang ZQ, Chen YJ, Mai JC, Ma J, Yang WH, Jing J: Dental caries is negatively correlated with body mass index among 7–9 years old children in Guangzhou, China. *BMC Public Health* 2016;16:638.
- Marques FP, Tórrès LHDN, Bidinotto AB, Hilgert JB, Hugo FN, De Marchi RJ: Incidence and predictors of edentulism among south Brazilian older adults. *Community Dent Oral Epidemiol* 2017;45:160–167.
- Mendes FM, Braga MM, Oliveira LB, Antunes JL, Ardenghi TM, Bonecker M: Discriminant validity of the International Caries Detection and Assessment System (ICDAS) and comparability with World Health Organization criteria in a cross-sectional study. *Community Dent Oral Epidemiol* 2010;38:398–407.
- Miller I, Miller M: John E. Freund's *Mathematical Statistics with Applications*. London, Pearson, 2015.
- Mittlbock M, Schemper M: Explained variation for logistic regression. *Stat Med* 1996;15:1987–1997.
- Nagelkerke NJD: A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–692.
- Pan W: Akaike's information criterion in generalized estimating equations. *Biometrics* 2001;57:120–125.
- Petrie A, Bulman J, Osborn J: Further statistics in dentistry. 6. Multiple linear regression. *Br Dent J* 2002a;193:675.
- Petrie A, Bulman JS, Osborn JF: Further statistics in dentistry. 4. Clinical trials 2. *Br Dent J* 2002b;193:557–561.
- Preisner JS, Das K, Benecha H, Stamm JW: Logistic regression for dichotomized counts. *Stat Methods Med Res* 2016;25:3038–3056.
- Preisner JS, Stamm JW, Long DL, Kincade ME: Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. *Caries Res* 2012;46:413–423.
- Ridout M, Demétrio CG, Hinde J: Models for count data with many zeros. *Proceedings of the XIXth International Biometric Conference*, Cape Town, 1998, vol 19, pp 179–192.
- Royall RM: Model robust confidence intervals using maximum likelihood estimators. *Int Stat Rev* 1986;54:221–226.
- Schmidt CO, Kohlmann T: When to use the odds ratio or the relative risk? *Int J Public Health* 2008;53:165–167.
- Schwarz G: Estimating the dimension of a model. *Ann Stat* 1978;6:461–464.
- Shaffer JR, Polk DE, Feingold E, Wang X, Cuenco KT, Weeks DE, DeSensi RS, Weyant RJ, Crout R, McNeil DW, Marazita ML: Demographic, socioeconomic, and behavioral factors affecting patterns of tooth decay in the permanent dentition: principal components and factor analyses. *Community Dent Oral Epidemiol* 2013;41:364–373.
- Shannon CE: A Mathematical theory of communication. *Bell System Tech J* 1948;27:379–423.
- Sheiham A, Alexander D, Cohen L, Marinho V, Moyses S, Petersen PE, Spencer J, Watt RG, Weyant R: Global oral health inequalities: task group – implementation and delivery of oral health strategies. *Adv Dent Res* 2011;23:259–267.
- Shmueli G: To explain or to predict? *Stat Sci* 2010;25:289–310.
- Vergnes JN, Boucher JP, Lelong N, Sixou M, Nabet C: Discrete distribution based on compound sum to model dental caries count data. *Caries Res* 2017;51:68–78.
- Zou G: A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702–706.