

RESEARCH ARTICLE

BAYESIAN MODELLING OF TUBERCULOSIS PREVALENCE IN SOUTH AFRICA. AN OVER-DISPERSION STUDIES

Davies Abiodun Obaromi^{a*}, Samuel Olorunfemi Adams^b^aDepartment of Biostatistics, Federal University of Health Sciences, Otukpo, Nigeria.^bDepartment of Statistics, University of Abuja, Abuja, Nigeria*Corresponding Author Email: samuel.adams@uniabuja.edu.ng

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 06 March 2023
Revised 21 April 2023
Accepted 03 May 2023
Available online 09 May 2023

ABSTRACT

Introduction: To account for over-dispersion in the impact of population density, number of schools, and average household size on tuberculosis prevalence in South Africa's Eastern Cape Province, this study compared robust count regression models with a Poisson regression model. **Methods:** The prevalence of tuberculosis in the province of Eastern Cape in South Africa in 2022 served as the basis for this study's data. Various models, including the Poisson regression model; Summed up Poisson and Negative Binomial Relapse were used within the sight of over-scattering. **Results:** It was seen that Poisson Relapse models couldn't deal with over-scattering present in the South Africa's Eastern Cape Region Tuberculosis dataset. Modeling with Negative Binomial Regression (NBR) and Generalized Binomial Regression (GBR) was used to solve the over-dispersion issue because both approaches can accommodate the dispersion parameter. Based on both selection criteria, the Negative Binomial model performed slightly better than the Generalized Poisson model. The findings indicated that population density has a positive and significant impact on tuberculosis prevalence in South Africa, that the number of schools in the province has a negative but significant impact, and that the average household has no significant impact. **Conclusions:** The Eastern Cape Province of South Africa's government ought to intensify its efforts to combat the spread and prevalence of tuberculosis by identifying, isolating, and treating tuberculosis early in the population and in schools.

KEYWORDS

Poisson Regression, Negative Binomial Regression, Generalized Poisson Regression, over-dispersion, Tuberculosis, Probability of direction (Pd), Deviance Information Criteria (DIC)

1. INTRODUCTION

It is discovered that, *M. tuberculosis* complex (MTB) mycobacteria are the cause of tuberculosis, a disease that can spread through contact. Droplets in the air are primarily how it spreads from person to person (Davies et al., 2014). Pulmonary tuberculosis is the most common type and has the greatest epidemiological impact. Patients with pulmonary tuberculosis either excrete or do not excrete detectable amounts of the bacilli during breathing, coughing, and other activities. The characteristic of tuberculosis is that 90% of infected individuals never develop active TB disease, remaining latently infected with *M. tuberculosis*. On the other hand, the infection can be fatal if it causes active TB disease: More than HIV/AIDS, tuberculosis (TB) is the single leading cause of death worldwide, with an estimated 10 million people developing active TB in 2019 and 1.4 million dying from TB (WHO, 2020). Untreated active TB is thought to last 3 years on average, or about 1.5 years for smear-positive TB and 5 years for smear-negative TB. In both recent studies, the case fatality rate is approximately 70% for smear-positive TB and approximately 20% for smear-negative TB (Tiemersma et al., 2011; Ragonnet et al., 2021). In a few epidemiological or clinical examinations, the reaction of interest typically comprises of counts, for example, the quantity of cells that show blemished proof of separation or the quantity of rehashed diseases experienced by a subject. Every value that is recorded is a non-negative integer. Multiple linear regression methods may allow for the analysis of counted observed data in some instances. Because it frequently offers an appropriate representation for the variability seen in count data, the Poisson distribution frequently plays a role in the analysis that is related to that of the normal distribution in multiple linear regressions. The

Poisson regression is the most often used model for count regression data, as is well known. One of the key properties of the Poisson distribution is that the variance is equal to the mean; nevertheless, in practice, we commonly see over-dispersed data with variances bigger than the mean (Rodriguez, 2013; Adebayo and Asuzu et al., 2015; Adams et al., 2020). Over-dispersion can be referred to as an additional variance when the model's systematic structure is accurate.

The presence of over-dispersion poses a particular challenge for the analysis of spatially grouped data, such as counts of events for census tracts with populations that are socially, economically, and demographically diverse, or where the event has a tendency to cluster at the tract scale, possibly as a result of an infectious process. Over-dispersion results from the stochastic component of the model's definition. In a variety of real-world scenarios where we seek to simulate count data, we may have varying observation times for our counts (Richard and MacDonald, 2008; Adams and Godwin, 2022). When the theoretical variance predicted by a chosen model exceeds the marginal variance of a characteristic, over-dispersion has occurred. Count data may have over- or under-dispersion in practice, and the variance may be bigger or smaller than the mean. Reparameterizations of the variance function are thus a different strategy for dealing with over-dispersion. This calls for the employment of alternative models, such as mixture models built from Poisson, which have predictable impacts on their probabilistic structures, to represent and capture various sorts of dispersions in count data. Some of these mixture models can also account for under-dispersion, but this is extremely uncommon. What might happen if the potential for over-dispersion in count data is not taken into consideration when it is found?

Quick Response Code



Access this article online

Website:
www.jhcdc.com.my

DOI:
10.26480/jhcdc.02.2023.28.32

First, because the model's standard errors will be inaccurate and unfitting, they may significantly underestimate the significance of each individual regression parameter (Xekalaki, 2014).

Additionally, the excessively large changes in deviance associated with model terms will lead to the selection of models that are overly complex. Although type I errors related to testing whether regression coefficients are zero or not are underestimated, standard error estimates for regression parameters are not, which is a problem when it comes to covariates that are close to the significance level. According to statistics, disregarding over-dispersion is most likely to have little impact on point estimates of the regression parameters (although some tiny sample bias may exist), as the maximum likelihood estimator is consistent. The Poisson model is kept when there is over-dispersion, but additional ad hoc models are added to account for the variance.

The presence of over-dispersion may lead an analyst to develop a model that is more complex than necessary and that overestimates the variation explained if the goal is to build a parsimonious model (John, 1997; John et al., 1998). These statistical problems are comparable to those that occur during model fitting when model residuals exhibit spatial autocorrelation (Haining, 1990; Haining, 2003). It is also feasible to utilize different distributions that could result in over-dispersion when over-dispersion occurs. Although there are other regression models, Poisson regression, negative binomial regression, and zero-inflated models are better suited for counted response measurements like disease counts in a region (Lee and Nelder, 2000).

As a result, the goal of this study is to compare robust count regression model that utilizes a Poisson regression model to account for over-dispersion in the impact of population density, number of schools and average household size on Tuberculosis prevalence in South Africa's Eastern Cape Province.

2. LITERATURE REVIEW

In a large population-based sample, it was demonstrated that COVID-19 death rates differ by age, gender, education level, and civil status (Drefahl et al., 2020). According to studies that were carried out in India, England, and the United States (Lenard et al., 2020; Williamosns et al., 2020; Laxminarayan et al., 2020), males had a higher rate of COVID-19 case fatalities. COVID-19 incidence, prevalence, and death rates varied by region, according to other studies (Panthee et al., 2020), most likely due to epidemiologic and population factors, clinical practices, and public health policies. The variations in coronavirus fatalities have been topographical, with an error between newly created and established districts. Coronavirus is inevitably fatal for older people in developed countries like the United States and Italy (Stokes et al., 2020; Istituto, 2020). In non-industrial countries like India, Brazil, and Iran, the majority of deaths occurred in patients between the ages of 50 and 70 (De Souza, 2020). Palestine's vulnerability, delicateness, financial flexibility, and poverty affected the Coronavirus response. The four essential suppliers of wellbeing administrations in the Palestinian Domains are the Service of Wellbeing (MOH), UNRWA, non-legislative associations (NGOs), and the confidential area, with MOH and UNRWA giving the most consideration. The Ministry of Health and non-governmental organizations provide the majority of secondary services, while the private sector provides the majority of tertiary care. During the pandemic, the MOH was responsible for the care and treatment of Coronavirus cases as well as the location of cases, following contacts, screening high-risk groups, and case location. All data pertaining to COVID-19 have been collected by the central MOH surveillance department.

3. METHODS

3.1 Epidemiological Information Sources

Based on survey and notification data for TB in the Eastern Cape Province, this is a retrospective secondary data source. All of the used data comes from the electronic tuberculosis register (ETR) records of TB patients from the twenty-four health sub-districts of the province, including the two metropolitan municipalities. The data collected was for the year 2022. The Eastern Cape Socio Economic Consultative Council's (ECSECC 2022) publication of all local municipality reports served as the source for the demographic variables and indicators.

3.2 Data Evaluation

An additional parameter can be used to model-over-dispersion in relation to the Poisson model. We focus on the generalized Poisson (GP) distribution and the negative binomial (NB) distribution presented by (Nikpouraghdam et al., 2020). Both by adding a third parameter, both

models enable independent modeling of the mean and variance. The Poisson regression model with both fixed and random effects is the working model. Its formula is:

$$\log \theta_i = \alpha_0 + \sum_{i=1}^3 X_i \alpha_i + \varepsilon_i \quad (1)$$

Where; ε_i the term for the disturbance is i.

3.3 Covariate Definition in the Regression Model

Three covariates (demographic factors) were used in our model to help explain the relative risk of the disease. Let the number of new instances of tuberculosis (TB) in area i, x_1 (percentage of adults without high school diplomas), x_2 (median household size), and x_3 (population density of the regions and municipalities) be the variables.

3.4 Bayesian Approach to Modeling

The foundation of a Bayesian analysis is the posterior probability distribution of model parameters. The conditional probability distribution of the unknown parameters, given the observed data and weighted by the prior knowledge, is known as the posterior probability distribution. To estimate each of the corresponding model parameters, Bayesian modeling relies on the capacity to generate posterior distributions. Most of these posterior distributions can be calculated easily. The posterior distribution of distributions with a conjugate prior often follows a conventional distributional pattern (Consul and Jain, 1973; Adams et al., 2020; Bliss and Owen, 1958; Poch and Mannering, 1996).

3.5 Bayesian Inference

The prior distribution, is the distribution of the parameter(s) before any data are observed. Identifying the prior distribution might not be simple.

$$p(\theta|\beta) \quad (2)$$

Before updating them with more recent observations in this situation, we can apply the Jeffreys prior to obtain the posterior distribution (Lord, 2006). The sampling distribution, often known as $p(X)$, is the distribution of the observed data in relation to its parameters

$$p(X|\theta) \quad (3)$$

Particularly when seen as a function of the parameter(s), this is also known as the likelihood and is occasionally written,

$$L(\theta|X) = p(X|\theta) \quad (4)$$

The distribution of the observed data minimized over the parameter(s) is known as the marginal likelihood, or evidence (Cameron et al., 1998),

$$p(X|\alpha) = \int_{\theta} p(X|\theta) p(\theta|\alpha) d\theta \quad (5)$$

The distribution of the parameter(s) after accounting for the observed data is known as the posterior distribution. Bayes' rule, the foundation of Bayesian inference, determines this.

$$p(\theta|X, \alpha) = \frac{p(X|\theta) p(\theta|\alpha)}{p(X|\alpha)} \propto p(X|\theta) p(\theta|\alpha) \quad (6)$$

But frequently the computation is more difficult, and the posterior distribution must be calculated using a more sophisticated technique. In order to estimate the distributions of the model parameters, these sophisticated approaches typically involve some kind of numerical simulation, typically by selecting a sample of parameter values from an approximation of the posterior distribution.

3.6 Fitting the Multilevel Bayesian Models

The suggested regression model would result in three (3) multilevel Bayesian models, which would be generated in the following order, for the study of the estimated risk factors to ascertain the impacts of these variables and the various mass densities of Poisson, Negative binomial, and Generalized Poisson (Winkelmann, 2003; Bliss and Fisher, 1953; Anscombe, 1950; Consul, 1989).

$$\log \lambda_i = \beta_0 + \sum_{i=1}^3 X_i \beta_i + \varepsilon_i \text{ with } O_i \sim \text{Poisson}(\lambda_i) \quad (7)$$

$$\log \lambda_i = \beta_0 + \sum_{i=1}^3 X_i \beta_i + \varepsilon_i \text{ with } O_i \sim \text{NB}(d, \lambda_i) \quad (8)$$

$$\log \lambda_i = \beta_0 + \sum_{i=1}^3 X_i \beta_i + \varepsilon_i \text{ with } O_i \sim \text{GP}(a, \lambda_i) \quad (9)$$

3.7 Comparison of Models

The probability of direction (pD) and deviance information criterion (DIC), which combines a measure of fit and a measure of model complexity based on the effective number of parameters, were used to compare different models. A more accurate model is indicated by lower pD and DIC values (Spiegelhalter et al., 2002).

4. RESULTS

Table 1 displays the outcomes of the Poisson, Negative Binomial, and Generalized Poisson models for addressing over-dispersion. For each model, the Likelihood of heading of over-scattering (pD), and the abnormality data measures (DIC) are given. From multilevel models, Negative Binomial and Generalized Poisson models produce the lowest DIC values, whereas Poisson models produce the highest DIC values. The pD and DIC therefore state that; With pD = 4.2 and DIC = 7123.56, the Poisson model clearly performed worse than the Negative Binomial (pD = 3.53 and DIC = 383.47) and Generalized Poisson (pD = 4.05 and DIC =

7123.56, respectively). This suggests that, based on both selection criteria, the Negative Binomial model performed slightly better than the Generalized Poisson model. It tends to be derived on the Poisson Regression models that the pervasiveness of Tuberculosis in South Africa's Eastern Cape Territory happened in overdispersion. Modeling with the Negative Binomial Regression (NBR) and Generalized Binomial Regression (GBR) solves the over-dispersion issue because both approaches can accommodate the dispersion parameter.

Estimates for the regression parameters in the Generalized Poisson count data regression model and the Negative binomial count data regression model are nearly identical, but the coefficient values for the Negative binomial model are preferred over those for the Generalized Poisson model. The Negative Binomial count model's results, which are presented in Table 2, indicate that, at levels of 1% and 5%, respectively, population density and the number of schools have a significant impact on the prevalence of tuberculosis in South Africa's Eastern Cape Province. However, the average size of the household was not statistically significant. The findings also indicated that the number of schools in a province has a negative but significant effect on tuberculosis prevalence in South Africa, while population density has a positive and significant effect.

Table 1: Model Selection Criteria for Over-Dispersion

Model Selection Criteria	Poisson	Negative Binomial	Generalized Poisson
(pD)	4.20	4.048	3.534
DIC	7123.56	383.47	384.40

DIC = deviance information criteria, pD = Number of Parameter

Table 2: Results of Poisson, Negative Binomial and Generalized Poisson Models in Handling Over-dispersion

Covariate	Poisson	Negative Binomial	Generalized Poisson
Intercept	5.99 (0.062)	6.96 (1.005)	6.41 (0.796)
Population Density	0.002 (0.000)**	0.60 (0.001)**	0.003 (0.000)**
Number of School	-496.34 (0.016)*	-5.80 (31.544)*	-8.44 (31.389)*
Average Household Size	0.44 (0.020)*	-0.02 (0.271)	0.15 (0.212)
Mean of Over-dispersion	-	3.314	14.14

Results in parenthesis are the standard errors, * indicates results that are significant at the 0.05 level and ** indicated result that are significant at the 0.01 level

5. DISCUSSION OF FINDINGS

In order to capture the over-dispersion of tuberculosis count data in conjunction with a few transmitting factors (demographic), the study compared the robustness of various Poisson regression models. According to our findings, in view of the took on model determination standards, the Negative Binomial model performed somewhat better compared to the Summed up Poisson model. This outcome is in accordance with previous study, which analyzed Summed up Poisson relapse and Negative Binomial relapse to pick the best count information model (Adams et al., 2021). When it comes to describing count data with excessive dispersion, their findings indicate that the Negative Binomial regression model is superior. In light of the review's discoveries, it was resolved that; the prevalence of tuberculosis is positively and significantly influenced by population density in South Africa. Others supported this finding, that a straightforward mathematical model of a known prevalence of hyper-susceptible individuals in the Ho Chi Minh City population is used in the study on the quantitative trend of tuberculosis dynamics (Vinh et al., 2018). The findings indicate that people who do not have HIV have a lower risk of contracting tuberculosis, with an estimated 18% (95 percent CI: 9–25%).

This investigation likewise discovered that the quantity of understudies in a school essentially affects the commonness of tuberculosis in South Africa, while the typical family fundamentally affects the pervasiveness of tuberculosis. A study compared the effects of various modeling assumptions and approaches on the epidemiologic projections of domestic LTBI control interventions in California, which confirmed our findings (Menzies et al., 2020). Under all models, it was found that the decline in tuberculosis incidence in California would remain stagnant.

6. CONCLUSION

This study primarily focused on the notable characteristic of over-dispersion, which is inherent to the Poisson distribution for count data. In this study, regression models for count data that allow for over-dispersion were presented. Utilizing zero expanded models or joining no swelled models with over-scattered conveyances, over-scattering can be displayed either by adding an extra boundary, as in the Negative Binomial and Generalized Poisson models, or by considering an extra extent of zero

perceptions. In this study, the Negative Binomial and Generalized Poisson distributions were used to take into account the additional variation in the data regarding tuberculosis counts. That's what this study showed, notwithstanding the way that the assessments of the relapse boundaries for the Negative binomial, the Generalized Poisson, and the Poisson were practically identical, the standard error of the Negative binomial and Generalized Poisson are more prominent than those of the Poisson. The Poisson distribution increased the influence of the regression parameters when there was over-dispersion.

Taking everything into account, both the Generalized Poisson model and the negative Binomial model are not especially hard to understand. Despite the fact that the probability density functions for both Negative Binomial and generalized Poisson models necessitate some mathematically complex formulas, the mean and variance for both models are abstractly simpler to comprehend. The Negative Binomial and Generalized Poisson models both have the same Poisson's mean. Because its variance is either the same as or greater than that of the Poisson, the Negative Binomial model can deal with over-dispersion. The Summed up Poisson can manage both over-scattering and under-scattering on the grounds that its change is something very similar, bigger, or more modest than that of the Poisson. It is advised that; the government of South Africa ought to intensify its efforts to combat the spread and prevalence of tuberculosis by identifying, isolating, and treating tuberculosis at an early stage, primarily within the population and schools in the Eastern Cape Province. The government should work to lower the country's high poverty rate because doing so will make it easier for tuberculosis patients to get medical diagnosis, treatment, and care. Lastly, there should be more programs to educate people about the importance of better nutrition and dieting. This can also significantly reduce the number of cases of tuberculosis in the Eastern Cape Province of South Africa.

ACKNOWLEDGMENTS

We are grateful to data and research department, administrators, staff, and members of Eastern Cape Department of Health for releasing the data of the province.

AUTHOR CONTRIBUTIONS

DAO: Conceptualization, Investigation, Methodology, Software, Validation,

Writing the original draft, writing the review, editing, data curation and investigation.

SOA: Conceptualization, Formal analysis Funding acquisition, Supervision, Writing the review, editing and visualization.

COMPLIANCE WITH ETHICAL STANDARDS

This study was carried out under the authorization and permission of the Ethical committee of the University of Fort Hare, Alice, Eastern Cape, South Africa and approval of the Eastern Cape Department of Health, with ethical clearance reference number QIN041SOBA01 and EC_2015RP24_398 respectively.

REFERENCES

- Adams, S.O. & Godwin, S. 2022. Comparative Study of the Error Trend and Seasonal Exponential Smoothing and ARIMA Model Using Covid-19 Death Rate in Nigeria, *International Journal of Natural Sciences Research* 2022; 10(1): 43-53. <http://doi.org/10.184488/63.v10i1.3031>
- Adams, S.O., Akano, R.O., Rauf, R.I. 2020. An Evaluation of the Chemical Composition of Soft drinks in Nigeria: A Principal Component Analysis Approach. *Advances in Modelling and Analysis A* 2020; 57(4):14-21. https://doi.org/10.18280/ama_a.571-403.
- Adams, S.O., Bamanga, M.O., Olanrewaju, S.O., Yahaya, HU and Akano, R.O. 2020. Modelling COVID-19 Cases in Nigeria Using Some Selected Count Data. *International Journal of Healthcare and Medical Sciences*. 6(4): 64-73. <https://doi.org/10.32861/ijhms.64.6.4.73>.
- Adams, S.O., Obaromi, A.D., Rauf, I.R. 2021. Modelling the Number of Household Visit to Health Care Centres in Some Nigeria Communities Using Count Data Regression Models. *Journal of Biostatistics and Epidemiology*. 7(1): 36-47. <https://doi.org/10.18502/jbe.v7i1.6293>
- Adebayo, A.M., Asuzu, M.C. 2015. Utilization of a community-based health facility in a low-income urban community in Ibadan, Nigeria. *African Journal of Pharmacy Health Care Family Medicine*. 7(1): 735-743. <http://dx.doi.org/10.4102/phcfm.v7i1.735>
- Anscombe, F. 1950. Sampling Theory of the Negative Binomial and Logarithmic Series Distribution, *Biometrika*, 37(3/4), 353-382. <https://doi.org/10.1093/biomet/37.3-4.358>
- Bliss, C.I., Fisher, R. A. 1953. Fitting the Negative binomial Distribution to Biological Data. *Biometrics* 1953; 9: 176-200.
- Bliss, C.I., Owen, A.R.G. 1958. Negative binomial Distributions with a Common K. *Biometrika* 1958; 45: 37-58.
- Cameron, A.C., Trivedi, P.K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press 1998.
- Consul, P. C. 1989. *Generalized Poisson Distributions: Properties and Applications*, New York: Marcel Dekker
- Consul, P.C. and Jain, G.C. 1973. A Generalization of Poisson Distribution. *Technometrics*, 15(4), 791 – 799.
- Davies, P. D. O., Gordon, S. B., Davies, G. (Eds.), 2014. *Clinical tuberculosis* (5th ed.). CRC Press, 2014, ISBN 978-1444154344
- De Souza W.M., Buss L.F., Da S. Candido D., et al., 2020. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil, *Nat. Hum. Behav.* 4 (8), 856-865, <https://doi.org/10.1038/s41562-020-0928-4>
- Drefahl S., Wallace M., Mussino E., et al., 2020. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden, *Nat. Commun.* 11 (1); 1-7, <https://doi.org/10.1038/s41467-020-18926-3>
- Haining, R.P. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge 1990.
- Haining, R.P. 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge 2003.
- Istituto Superiore di Sanit'a, Characteristics of COVID-19 Patients Dying in Italy, 2020.
- John H., Clarice G.B. Demetrio, 1998. Overdispersion: Models and Estimation. *Computation Statistics & Data Analysis* 1998; 27(2): 151-170. [https://doi.org/10.1016/S0167-9473\(98\)000073](https://doi.org/10.1016/S0167-9473(98)000073)
- John M. 1997. Heterogeneity, Excess Zeros, and the Structure of Count Data Models. *Journal of Applied Econometrics* 1997; 12(3): 337-350.
- Laxminarayan R., Wahl B., Dudala S.R., et al., 2020. Epidemiology and transmission dynamics of COVID-19 in two Indian states, *Science* 370 (6517), 691-697, <https://doi.org/10.1126/science.abd7672>
- Lee, Y., Nelder, J.A. 2000. Two Ways of Modelling Overdispersion in Non-Normal Data. *Journal of the Royal Society Series C. (Applied Statistics)*; 49(4): 591-598. <https://www.jstor.org/stable/2680790>
- Lewnard J.A., Liu V.X., Jackson M.L., et al., 2020. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study, *BMJ* 369 (9), <https://doi.org/10.1136/bmj.m1923>
- Lord, D. 2006. Modelling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining The Effects Of Low Sample Mean Values And Small Sample Size On The Estimation of the Fixed Dispersion Parameter. *Accident Analysis & Prevention*, 38(4): 751-766.
- Menzies N.A., Parriott A., Shrestha S., Dowdy, D.W., Cohen, T., Salomon, J.A., Marks, S.M., Hill, A.N., Winston, C.A., Asay, G.R., Barry, P., Readhead, A, Flood, J., Kahn, J.G., Shete, P.B. 2020. Comparative Modeling of Tuberculosis Epidemiology and Policy Outcomes in California. *Am J Respir Crit Care Med*. 2020; 1;201 (3):356-365. <https://doi.org/10.1164/rccm.201907-12890C>.
- Nikpouraghdam M., Jalali Farahani A., Alishiri G.H., et al., 2020. Epidemiological characteristics of coronavirus disease 2019 (COVID-19) patients in Iran: a single center study, *J. Clin. Virol.* 127, 104378, <https://doi.org/10.1016/j.jcv.2020.104378>
- Panthee B., Dhungana S., Panthee N., Gyawali S., Paudel A., Panthee S., 2020. Clinical and epidemiological features of COVID-19 deaths in Nepal, *New Microbes New Infect.* 38, 100797, <https://doi.org/10.1016/j.nmni.2020.100797>
- Poch, M., Mannering, F. L. 1996. Negative binomial Analysis of Intersection Accident Frequency. *Journal of Transportation Engineering*; 122(2):105-113.
- Ragonnet, R., Flegg, J. A., Brilleman, S. L., Tiemersma, E. W., Melsew, Y. A., McBryde, E. S., & Trauer, J. M. 2021. Revisiting the natural history of pulmonary tuberculosis: A bayesian estimation of natural recovery and mortality rates. *Clinical Infectious Diseases* 2021; 73(1), e88ee96. <https://doi.org/10.1093/cid/ciaa602>
- Richard, B. MacDonald, J. 2008. Overdispersion and Poisson Regression. *Journal of Quantitative Criminology* 2008; 24(3): 269-234. <https://doi.org/10.1007/s10940-008-9048-4>
- Rodriguez, G. 2013. Models for Count Data with Over-Dispersion. Addendum to the WWS 509 Notes on Extra-Poisson Variation and the Negative Binomial Model 2013.
- Spiegelhalter D, Best N, Carlin BP. 2002. Bayesian measures of model complexity and fit. *Journal of Royal Statistics Society Series B Methodology* 2002; 64:539-583.
- Stokes E.K., Zambrano L.D., Anderson K.N., et al., 2020. Coronavirus disease 2019 case surveillance — United States, January 22–May 30, *MMWR Morb. Mortal. Wkly. Rep.* 69 (24), 759-765, <https://doi.org/10.15585/mmwr.mm6924e2>.
- Tiemersma, E. W., Van Der Werf, M. J., Borgdorff, M. W., Williams, B. G., & Nagelkerke, N. J. D. 2011. Natural history of tuberculosis: Duration and fatality of untreated pulmonary tuberculosis in HIV-negative patients: A systematic review. *PLoS One*, 6(4), Article e17601. <https://doi.org/10.1371/journal.pone.0017601>
- Vinh, D.N., Ha, D.T.M., Hanh, N.T. et al. 2018. Modeling tuberculosis dynamics with the presence of hyper-susceptible individuals for Ho Chi Minh City from 1996 to 2015. *BMC Infect Dis* 18, 494 (2018). <https://doi.org/10.1186/s12879-018-3383-3>

Williamson E.J., Walker A.J., Bhaskaran K., et al., 2020. Factors associated with COVID- 19-related death using Open SAFELY, *Nature* 584 (7821), 430–436, <https://doi.org/10.1038/s41586-020-2521-4>

Winkelmann, R. 2003. *Econometric Analysis of Count Data*. 4th Edition. Berlin: Springer-Verlag 2003.

World Health Organization (WHO). 2020. *Global tuberculosis report 2020*. World Health Organization, ISBN 978-92-4-001314-8.

Xekalaki E. 2014. On the Distribution Theory of Over-dispersion. *Journal of Statistical Distributions and Application* 2014; 1(19): 1-12. <https://doi.org/10.1186/s40488-014-0019-z>

