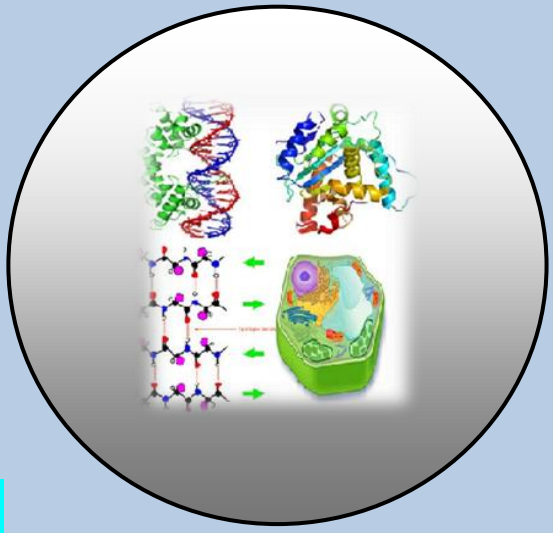
A Statistical Model of Prediction of Morbidity Pattern of Rare Diseases in a Particular Region

By
Syed Mohd Haider Zaidi

ISSN 2319-3077 Online/Electronic
ISSN 0970-4973 Print

Index Copernicus International Value
IC Value of Journal 46.52 Poland, Europe (2015)
Journal Impact Factor: 4.275
Global Impact factor of Journal: 0.876
Scientific Journals Impact Factor: 3.285
InfoBase Impact Factor: 3.66

J. Biol. Chem. Research
Volume 34 (1) 2017 Pages No. 345-349



Journal of Biological and Chemical Research

An International Peer Reviewed / Referred Journal of Life Sciences and Chemistry

Indexed Abstracted and Cited in various International and National Scientific Databases

Published by Society for Advancement of Sciences®

J. Biol. Chem. Research. Vol. 34, No. 1: 345-349, 2017
(An International Peer Reviewed / Refereed Journal of Life Sciences and Chemistry)
Ms 34/01/255/2017

ISSN 0970-4973 (Print)
ISSN 2319-3077 (Online/Electronic)

Dr. S.M.H. Zaidi
http:// www.sasjournals.com
http:// www.jbcr.co.in
jbiolchemres@gmail.com

RESEARCH PAPER

Received: 18/03/2017 Revised: 06/06/2017 Accepted: 07/06/2017

A Statistical Model of Prediction of Morbidity Pattern of Rare Diseases in a Particular Region

Syed Mohd Haider Zaidi

Department of Statistics, Shia PG College, Sitapur Road, Lucknow – 226020, U.P., India

ABSTRACT

Rare diseases represent a serious public health problem because of their seriousness and lack of sufficient number of researches. There are very few studies to examine the trends and patterns of morbidity status of rare diseases across the various populations. The Present Study used Poisson Distribution and Beta Priors for estimating the Morbidity pattern of Rare diseases.

Keywords: Statistical Model, Prediction Morbidity Pattern and Diseases.

INTRODUCTION

A rare disease is any disease that affects a small percentage of the population. The World Health Organization (W.H.O.) has suggested that a rare disease should be defined as one with frequency less than 6.5 – 10 per 10,000 people. Rare diseases range from cystic fibrosis and haemophilia to Angelman Syndrome, with an incidence of about 1 in 15000, to Opitz trigonocephaly syndrome, which is extremely rare with about one case per million people [Bulletin of the World Health Organization, 2012]. Rare diseases represent a serious public health problem because of their seriousness and lack of sufficient number of researches and financial hazard to treat or cure.

There are very few studies to examine the trends and patterns of morbidity status of rare diseases across the various populations. To find incidence, prevalence and pattern of morbidity of rare diseases is a difficult task because of big level screening requirement with a big sample size to detect it.

In this regard medical sciences provide small literature regarding discovering patterns of morbidity and seasonal variations among rare diseases or areawise variations.

A review of methods for the analysis of the geographical distribution of disease is presented. The topic is of increasing interest to statisticians, though much groundwork has

been done by epidemiologists and medical geographers. Methods for the detection and testing for apparent clusters of disease, including those near a possible environmental hazard, are reviewed. Estimating regional mortality rates, possibly to construct disease maps, and methods for investigating the association between disease rates and social, demographic and environmental factors are discussed. In addition, statistical analysis of the spatial spread of epidemics is mentioned [Roger et al., 1991].

One example is the study of Shahinaz M. Gadalla et.al. about the rare disease 'Thymomas', the rare tumors of the mediastinum. Only limited number of small studies have evaluated the outcomes in the patients of that disease. In this study there identified 668 patients with thymoma from the Swedish Cancer Registry, and compared them with 2,719 population-based matched controls. Then obtained information on autoimmunity from the nationwide inpatient/outpatient hospital discharge Registry. Survival analysis had been done using Kaplan-Meier curves, conditional regression and Cox proportional hazards models also used to evaluate the association between thymoma and autoimmune diseases, and standardized incidence ratios (SIRs) to evaluate the risk for second cancers following thymoma.It was found that for thymoma patients, younger age at diagnosis and being diagnosed in recent years were associated with a better survival and also found that Thymoma patients had twofold excess risk for second cancers compared with the general population and over time thymoma patients have worse survival than controls.

ESTIMATION OF MORBIDITY PATTERN OF RARE DISEASES

After some earlier trials to find morbidity pattern of rare diseases it was concluded by Cornfield in 1951, that when outcomes are sufficiently rare, the odds ratio from a case–control study will approximate the population risk ratio for the association of an exposure with a disease outcome. It was later realized that if controls are sampled as each case arises in time; the odds ratio will estimate the incidence-rate ratio even when outcomes are common [Peter Cummings, 2009].

Later on it was discovered that the extension of case-control methods to the study of common outcomes has led to the development of several design and analysis techniques which do not employ the rare-disease assumption. Unfortunately, the principles underlying valid application of these techniques are more subtle than those first considered by Cornfield in the rare-disease setting, and appear to be easily misunderstood [Greenland et al., 1986]. Another study of 1998 [Elham Rahme and Lawrence Joseph, 1998] used the adjusted maximum likelihood method for the estimation of rare diseases. Using this method the author made the estimator for the prevalence of rare diseases, the adjusted confidence interval and the formula for sample size calculation also had been developed.

In a more recent study [Middleton et al., 2004] of detecting patterns of morbidity and rehospitalisation following spinal cord injury a Longitudinal, descriptive design had been used. The aim of the study was to investigate the frequency, cause and duration of rehospitalisations in individuals with spinal cord injury (SCI) living in the community. Descriptive statistics and time to readmission using survival analysis, stratified by ASIA impairment grade, were calculated.

Here *Y* be a random variable denoting the number of persons suffering from the particular rare disease during a time t and assume the suffered persons to be independent with each other with probability of suffering in time δt as $\lambda\delta t$. Taking $P_k(t)$ and $P_k(t+\delta t)$ as chance of *k* sufferings in *t* and $t+\delta t$. We have two mutually exclusive possibilities:

1. There are *k* sufferings in time *t* and no one suffered in δt.

2. There are (*k* – 1) sufferings in time *t* and only one suffering in δt as δt is small.

Now

$$P_k(t+\delta t) = P_k(t)\cdot\{1-\lambda\delta t\} + P_{k-1}(t)\cdot\lambda\delta t$$
$$= P_k(t) - P_k(t)\cdot\lambda\delta t + P_{k-1}(t)\cdot\lambda\delta t$$

Or,

$$\lim_{\delta t\to 0}\frac{P_x(t+\delta t)-P_x(t)}{\delta t} = \lim_{\delta t\to 0}\lambda\left[P_{x-1}(t)-P_x(t)\right]$$

Or

$$\frac{d}{dt}P_k(t) = \lambda\left[P_{k-1}(t)-P_k(t)\right]$$

So we get a differential equation of first order and first degree.

Putting

$$P_k(t) = \frac{(\lambda t)^k}{k!}f(t) \text{ so that } P_0(t) = f(t) \text{ and } P_0(0) = f(0) = 1 \quad \text{……(1)}$$

$$\frac{d}{dt}\left[\frac{(\lambda t)^k}{k!}f(t)\right] = \lambda\left[\frac{(\lambda t)^{k-1}}{k-1!}f(t) - \frac{(\lambda t)^k}{k!}f(t)\right]$$

or, $$\frac{(\lambda t)^k}{k!}f'(t) + \frac{\lambda^k}{k!}f(t)\cdot kt^{k-1} = \frac{\lambda^k t^{k-1}}{k-1!}f(t) - \frac{\lambda t}{k}\frac{\lambda^k t^{k-1}}{k-1!}f(t)$$

Dividing the above equation by $\frac{\lambda^k t^{k-1}}{k-1!}f(t)$ we get

$$\frac{\lambda t}{k}\cdot\frac{f'(t)}{f(t)} + 1 = 1 - \frac{\lambda t}{k}$$

$$\Rightarrow \frac{f'(t)}{f(t)} = -\lambda$$

On integrating with applying initial conditions (1) we get

$$\log f(t) = -\lambda t \Rightarrow f(t) = e^{-\lambda t}$$

So that

$$P_k(t) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$$

This is the probability distribution of number of persons suffered during the time interval *t*. Obviously the above probability distribution is a well known distribution of Probability theory called the 'Poisson distribution'. Which is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event [Frank A. Haight, 1967] The Poisson distribution also known as the law of rare events. The law of rare events states that the total number of events will follow, approximately, the Poisson distribution if an event may occur in any of a large number of trials but the probability of occurrence in any given trial is small [Colin Cameron, 1998].

Estimation and Description of λ

λ is the prevalence of the particular rare disease. It can be estimated initially by some empirical data and then this estimator can be improved by Bayesian technique of conjugate priors as discussed below [Fink, 1995].

Let *n* people $x_1 x_2 \ldots x_n$ be selected such that,

$$x_i = \begin{cases} 1 & \text{if } i^{th} \text{ person suffered with disease} \\ 0 & \text{if } i^{th} \text{ person does not suffered with disease} \end{cases}$$

Then the likelihood function for λ with gamma prior is given by

$$p\left(x_1, x_2, \ldots x_n / \lambda\right) = \prod_{i=1}^{n} \frac{\lambda^{x_i} \exp\left(-\lambda\right)}{x_i !}$$

$$= \left(\prod_{i=1}^{n} x_i \right)^{-1} \lambda^{\sum_{i=1}^{n} x_i} \exp\left(-n\lambda\right)$$

So the conjugate prior density for λ is given by

$$p\left(\lambda / \overline{x}\right) = \eta \lambda^{n\overline{x}} \exp\left(-n\lambda\right)$$

Where η be the normalization coefficient.

In this way λ can be estimated.

CONCLUSION

So that it can be concluded that the Poisson probability distribution can be used to estimate the morbidity pattern of rare diseases like cystic fibrosis or haemophilia. Poisson probability distribution is very useful to estimate probabilities of rare events. Further the parameter of Poisson distribution can be estimated using Bayesian technique of conjugate priors.

ACKNOWLEDGEMENTS

Author is grateful to Principal Shia P.G. College, Lucknow, U.P., India for providing all facilities during course of present investigation.

REFERENCES

Coming together to combat rare diseases (2012). "Bulletin of the World Health Organization", Volume 90, Number 6, June 2012, pp. 406-407

Roger, J. Marshall (1991). "A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease", Vol. 154, No. 3, pp. 421-441

Shahinaz, M. Gadalla, Arun Rajan, Ruth Pfeiffer, Sigurdur Y. Kristinsson, Magnus Björkholm, Ola Landgren and Giuseppe Giaccone. "A population-based assessment of mortality and morbidity patterns among patients with thymoma", Volume 128, Issue 11, pp. 2688–2694.

Peter Cummings (2009). "Methods for estimating adjusted risk ratios", The Stata Journal, 9, Number 2, pp. 175–196

Greenland, S., Thomas, D.C. and Morgenstern, H. (1986). "The rare-disease assumption revisited. A critique of - estimators of relative risk for case-control studies", American Journal of Epidemiology, 124(6) pp. 869-883.

Elham Rahme and Lawrence Joseph (1998). "Estimating the Prevalence of a Rare Disease : Adjusted Maximum Likelihood", The Statistician, 47 (1) pp. 149-158.

Middleton, J.W., K. Lim, L. Taylor, R. Soden and S. Rutkowski (2004). "Patterns of morbidity and rehospitalisation following spinal cord injury", Spinal Cord, 42, 359–367.

Frank, A. Haight (1967). Handbook of the Poisson distribution. New York: John Wiley & Sons.

Colin Cameron, A. and Pravin K. Trivedi (1998). Regression Analysis of Count Data. Retrieved 2013-01-30 (p.5).

Fink, D. (1995). A Compendium of Conjugate Priors. In progress report: Extension and enhancement of methods for setting data quality objectives. (DOE contract 95‑831).

Corresponding author: Dr. Syed Mohd Haider Zaidi, Department of Statistics, Shia P.G. College, Sitapur Road, Lucknow – 226020, U.P., India
Email: syedmohdhaiderzaidi@yahoo.com