A simple Bayesian linear excess relative risk model

Manuel Higueras1,*, Richard W. Harbron[2], Mark S. Pearce[2]
[1]Basque Center for Applied Mathematics, Bilbao, Spain
[2]Institute of Health and Society, Newcastle University; Newcastle upon Tyne, UK
* Corresponding author: mhigueras@bcamath.org

Abstract

A new Bayesian Poisson relative risk model is proposed for displaying the excess relative risk associated to a unique exposure as a probability distribution in a closed form. The background risk can be modelled by a unique two levels factor, *e.g.* gender or smoking status.

Keywords: radiation epidemiology; Poisson nonlinear regression; improper priors.

1 Introduction

The excess relative risk (ERR) represents the excess risk of disease (*e.g.*, leukaemia, brain tumour) per unit of exposure (*e.g.*, absorbed dose of ionising radiation). In a linear relative risk model with one exposure, the risk is modelled by $e^{\eta}(1 + \beta D)$, where the term β is the ERR, D represents the absorbed dose and e^{η} is the background risk. Poisson relative risk models are derived to calculate the ERR in follow-up studies, *i.e.*

$$C_i \sim \text{Pois}\left(PY_i e^{\eta_i}(1 + \beta D_i)\right), \tag{1}$$

where C_i, PY_i and D_i are the number of disease cases, the number of person-years of follow-up and D_i is the mean dose (weighted by the person-years) for stratum i respectively [1]. The linear predictor for the background risk has the form $\eta_i = \alpha_0 + \alpha_1 x_i$, where x_i represents a two levels factor, *e.g.* it is 1 if patient i is smoker, and 0 otherwise.

The model (1) is not the typical log-linear Poisson model [2], it mixes both log-linear and linear terms, so it is a Generalized Nonlinear Model. The model proposed here is simple, with only one exposure, and one 2-stage fixed effect factor in the background risk linear predictor.

The Bayesian analysis combines prior information, in form of probability distributions, with the likelihood function of an assumed model, providing posterior results as probability distributions too. The Bayes' theorem in its continuous version establishes

$$P(\Theta|X) = \frac{L(\Theta|X)P(\Theta)}{\int L(\Theta|X)P(\Theta)d\Theta}, \tag{2}$$

where Θ is the continuous parameter set, X is the observed data set, $L(\Theta|X)$ is the likelihood function, $P(\Theta)$ is the prior probability density function of Θ and $P(\Theta|X)$ is the posterior probability density of Θ given data X. See Christensen *et al.* 2011 [3], for further description.

2 The posterior ERR

Following expressions (1) and (2):

$$\begin{aligned}
\Theta = & \{\alpha_0, \alpha_1, \beta\}, \\
X = & \{C_i, PY_i, D_i\}_{i=1}^n, \\
L(\Theta|X) = & \prod_{i=1}^n \frac{(PY_i e^{\alpha_0 + \alpha_1 x_i}(1 + \beta D_i))^{C_i} \exp(-PY_i e^{\alpha_0 + \alpha_1 x_i}(1 + \beta D_i))}{C_i!}.
\end{aligned} \tag{3}$$

The value n represents the number of patients in the follow-up.

Assuming α_0, α_1 and β are independent, the prior probability density remains $P(\Theta) = P(\alpha_0)P(\alpha_1)P(\beta)$. It is also assumed that both priors are non informative, meaning that the probability is the same for all the values in the support of the parameters. This leads to the following improper uniform priors:

$$\begin{aligned}
\alpha_0 & \sim \mathcal{U}(-\infty, +\infty), \\
\alpha_1 & \sim \mathcal{U}(-\infty, +\infty), \\
\beta & \sim \mathcal{U}(0, +\infty).
\end{aligned} \tag{4}$$

The Bayesian framework affords the definition of improper prior distributions.

Applying the Bayes' theorem (2), the posterior of Θ is

$$
\begin{aligned}
\mathrm{P}(\Theta|X) \propto \quad & \mathrm{L}(\Theta|X) \propto \prod_{i=1}^{n}(PY_i \mathrm{e}^{\alpha_0+\alpha_1 x_i}(1+\beta D_i))^{C_i}\exp(-PY_i \mathrm{e}^{\alpha_0+\alpha_1 x_i}(1+\beta D_i)) \\
= \quad & \exp\left(T\alpha_0 - \mathrm{e}^{\alpha_0}\sum_{i=1}^{n}PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i)\right)\prod_{i=1}^{n}(PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i))^{C_i},
\end{aligned}
\tag{5}
$$

where $T=\sum_{i=1}^{n}C_i$ is the total number of diseases in the follow-up.

The goal here is to get the marginal posterior of the ERR, the posterior distribution of β. First it is calculated the joint marginal posterior of (α_1,β) which it is proportional to the integration of expression (5) over α_0, *i.e.*

$$
\begin{aligned}
\mathrm{P}(\alpha_1,\beta|X) \propto \quad & \int_{-\infty}^{+\infty}\mathrm{P}(\Theta|X)\mathrm{d}\alpha_0 = \left[\prod_{i=1}^{n}(PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i))^{C_i}\right]\int_{-\infty}^{+\infty}\exp\left(T\alpha - \mathrm{e}^{\alpha}\sum_{i=1}^{n}PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i)\right)\mathrm{d}\alpha_0 \\
= \quad & \left[\prod_{i=1}^{n}(PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i))^{C_i}\right]\frac{(T-1)!}{\left[\sum_{i=1}^{n}PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i)\right]^{T}} \propto \frac{\prod_{i=1}^{n}(\mathrm{e}^{\alpha_1 x_i}(1+\beta D_i))^{C_i}}{\left[\sum_{i=1}^{n}PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i)\right]^{T}}.
\end{aligned}
\tag{6}
$$

Then the marginal posterior of the ERR is proportional to the integration of expression (6) over α_0,

$$
\begin{aligned}
\mathrm{P}(\beta|X) \propto \quad & \int_{-\infty}^{+\infty}\mathrm{P}(\alpha_1,\beta|X)\mathrm{d}\alpha_1 \propto \int_{-\infty}^{+\infty}\prod_{i=1}^{n}(\mathrm{e}^{\alpha_1 x_i}(1+\beta D_i))^{C_i}\left[\sum_{i=1}^{n}PY_i \mathrm{e}^{\alpha_1 x_i}(1+\beta D_i)\right]^{-T}\mathrm{d}\alpha_0 \\
= \quad & \left[\prod_{i=1}^{n}(1+\beta D_i)^{C_i}\right]\int_{-\infty}^{+\infty}\mathrm{e}^{N\alpha_1}\left[\sum_{i|x_i=0}PY_i(1+\beta D_i)+\mathrm{e}^{\alpha_1}\sum_{i|x_i=1}PY_i(1+\beta D_i)\right]^{-T}\mathrm{d}\alpha_0 \\
= \quad & \left[\prod_{i=1}^{n}(1+\beta D_i)^{C_i}\right]\frac{(N-1)!}{\prod_{i=1}^{N-1}(T-i)}\frac{\left[\sum_{i|x_i=0}PY_i(1+\beta D_i)\right]^{N-T}}{\left[\sum_{i|x_i=1}PY_i(1+\beta D_i)\right]^{N}} \propto \frac{\left[\prod_{i=1}^{n}(1+\beta D_i)^{C_i}\right]\left[\sum_{i|x_i=0}PY_i(1+\beta D_i)\right]^{N-T}}{\left[\sum_{i|x_i=1}PY_i(1+\beta D_i)\right]^{N}},
\end{aligned}
\tag{7}
$$

where n_0 denotes the number of cases, PY_{0i} the person-years and D_{0i} the absorbed dose for each patient i such that $x_i=0$. Analogously for n_1, PY_{1i} and D_{1i} for those $x_i=1$. Consequently,

$$
\mathrm{P}(\beta|X)=\frac{\left[\prod_{i=1}^{n}(1+\beta D_i)^{C_i}\right]\left[\sum_{i|x_i=0}PY_i(1+\beta D_i)\right]^{N-T}}{K\left[\sum_{i|x_i=1}PY_i(1+\beta D_i)\right]^{N}},
\tag{8}
$$

where K is the normalising constant

$$
K=\int_{0}^{+\infty}\left[\prod_{i=1}^{n}(1+\beta D_i)^{C_i}\right]\left[\sum_{i|x_i=0}PY_i(1+\beta D_i)\right]^{N-T}\left[\sum_{i|x_i=1}PY_i(1+\beta D_i)\right]^{-N}\mathrm{d}\beta,
\tag{9}
$$

that is calculated by numerical integration (there is not analytical solution). The probability density (8) has not a recognizable form, but this is not rare when dealing with Bayesian analysis.

The integrals in Expressions (6) and (7) are calculated by recursive integration by parts.

3 Practical example

Pearce *et al.* 2012 [4] analysed the risk of leukaemia and brain tumour in young patients who were first examined with CT in National Health Service centres in England, Wales, or Scotland in a 23 years retrospective cohort study. Table 1 displays

Table 1: Cases of leukaemia, person-years, mean dose per group and relative risk for the different dose groups.

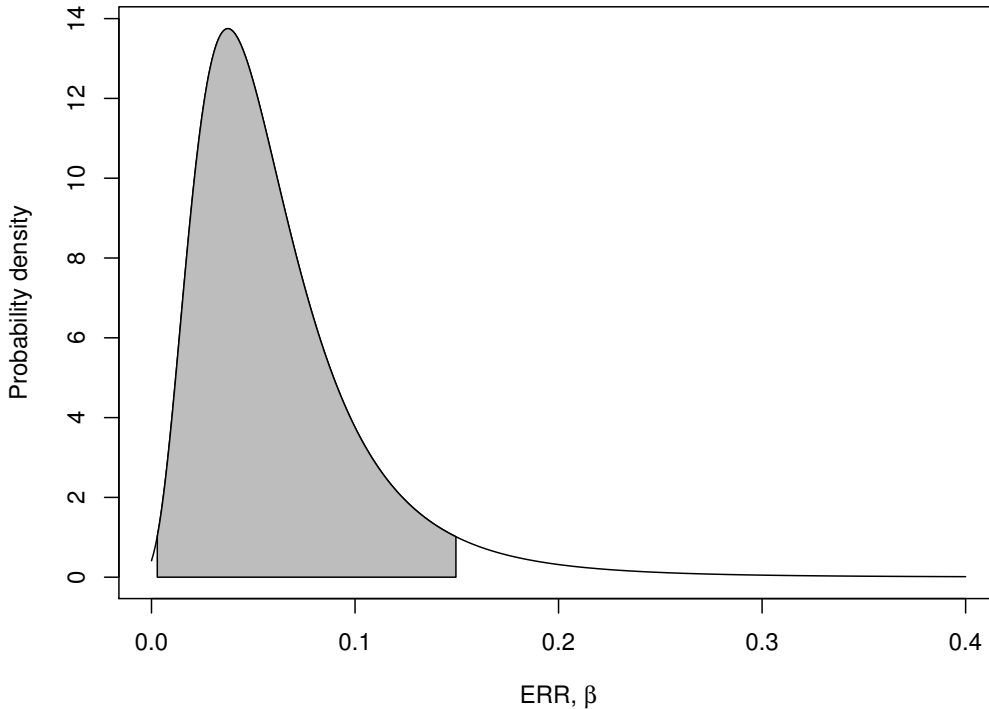| Dose group (mGy)	Cases	Person-years	Mean dose (mGy)	Relative risk
< 5	15	588,450	2.32	1.00
$5-9$	17	438,828	7.08	1.44
$10-14$	12	213,289	12.34	2.03
$15-19$	11	244,844	16.54	1.53
$20-29$	4	70,523	24.69	2.02
> 30	15	165,049	51.13	3.18

Figure 1: Posterior probability density of the ERR (solid line) and its 95% HPD (shaded grey).

the data of the leukaemia dose response in six different dose groups. There were 74 leukaemia diagnosis for 178,604 patients, and a total of 1,720,984 person-years in this study.

Following Equation (8), for the 2 years exclusion and lag-period person-year table in Pearce *et al.* 2012 [4], and taking the attained age greater than 20 as the categorical variable in the baseline rate, the ERR is inferred.

Figure 1 shows the posterior density function of the ERR following expression (8). The modal posterior ERR value is 0.038 and its 95% highest posterior density (HPD) is (0.003, 0.150).

One of the big advantages of the Bayesian framework is that it is possible to get the probability of the studied parameter for a range of values, for instance in this example there is a 53.15% chance for the ERR being greater than 0.050.

Following the classical procedure (frequentist), the maximum likelihood estimation of the ERR is $\hat{\beta} = 0.036$ and its 95% confidence interval is $(0.005, 0.120)$. Although both the Bayesian and the frequentist methods provide estimation and uncertainty results, when comparing them it is important to remark that they are not measuring the same thing, the frequentist one assumes that the parameter is a fixed value and the maximum likelihood estimator is a random variable whereas the Bayesian assumes the opposite.

4 Conclusion

The Bayesian model presented here for estimating the ERR in radiation epidemiology follow-up studies is simply and easy to implement. The Bayesian analysis provides an accurate framework for dealing with uncertainties, with the results being in the form of probability densities.

One of the main criticism of the Bayesian analysis is the use of prior distributions for the parameters, along with how the researcher defines them and how they can influence in the results. Here we propose a non-informative prior for both parameters, so the priors do not influence the final outcome. If desired, it is possible to define informative priors, for instance a normal for α and a gamma for β, but the model would lose its simplicity and closed form, although it would not be a very complex one.

References

[1] Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation. *Health Risks from exposure to low levels of ionizing radiation. BEIR VII Phase 2.* Washington DC: The National Academies Press, 2006.

[2] McCullagh P, Nelder J. *Generalized Linear Models*, 2nd edition. Champman & Hall/CRC Press, 1989.

[3] Christensen R, Johnson W, Brascum A, Hanson TE. *Bayesian ideas and data analysis.* Champman & Hall/CRC Press, 2011.

[4] Pearce MS, Salotti JA, Little MP, Mchugh K, Lee C, Kim KP, Howe NL, Ronckers CM, Rajaraman P, Craft AW, Parker L, Berrington De Gonzalez A. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet* **380**:499–505, 2012.