RIF - Statistical Manual

***Disease mapping***

Disease maps aim at representing the geographical distribution of the incidence of disease. In the frame of the RIF software, only maps produced with counts data are considered. Counts of disease cases are reported for a list of regions, denoted here areas, and delimited by geographical boundaries. The easiest way to map the geographical variations of the disease would be to directly map the counts. However, these counts depend strongly of the structure of the population at risk within each area, and cannot be directly compared. Two variables are thus considered to exclude the effect of populations, the directly standardized rates and the relative risk. These two variables demand the definition of a ‘comparison’ population associated to each area. The comparison population may be the total population of all study areas, or subsets of study areas.

***Directly standardised rates***

Rates of disease are calculated by dividing the number of observed cases by the underlying populations at risk. However, because this variable does not take into account the age and sex structure of the population, rates between different areas cannot be compared.

The directly standardised rates are calculated by applying to the comparison populations the rates of the disease in each age and sex strata of the corresponding study area . Thus

where is the population–years at risk in age/sex strata in the comparison population. The rate is multiplied by to obtain a rate per person years. Directly standardised rates of several areas can be directly compared because they are calculated for the same standard population (the comparison population).

95% confidence intervals (95% CI) indicate the degree of uncertainty in the estimation of . Assuming that the number of cases in strata and area follow a Poisson distribution, the directly standardised rates are weighted sums of Poisson variables:

* If , a Gaussian approximation of is done. Then,
  + the lower 95% CI
  + the upper 95% CI

where is an estimate of the standard deviation of the log directly standardised rate.

* If , 95% CI are obtained with the formula by Dobson et al. (1991) for weighted sums of Poisson parameters.
  + Lower 95% CI
  + Upper 95% CI

Where

and is the lower 95% CI of the mean of the Poisson variable ,

and is the upper 95% CI of the mean of the Poisson variable .

*Variable names:*

|  |  |
| --- | --- |
| Column name | Description |
| RATE\_UNADJ | Directly standardised rate adjusted for age and sex only |
| RATEL95\_UNADJ | Lower bound of the 95% CI of the directly standardised rate adjusted for age and sex only |
| RATEU95\_UNADJ | Upper bound of the 95% CI of the directly standardised rate adjusted for age and sex only |
| RATE\_ADJ | Directly standardised rate adjusted for age, sex and provided adjustment variables |
| RATEL95\_ADJ | Lower bound of the 95% CI of the directly standardised adjusted for age, sex and provided adjustment variables |
| RATEU95\_ADJ | Upper bound of the 95% CI of the directly standardised adjusted for age, sex and provided adjustment variables |

***Relative risk ratio***

To compare areas, one can also consider the relative excess or deficit of observed counts in comparison with the expected value if rates in all strata of area , where identical to rates in the comparison population. This value names the relative risk, and it is roughly estimated by the ratio between the observed counts and expected counts . This estimate names the standardised mortality rate () or the standardised incidence rate (). The expected counts are calculated by applying to each strata of the study population, the rates of the comparison population,

where is the study population in strata of area .

Values of the relative risk larger than one indicate an excess of risk relatively to the underlying ‘comparison population’, whereas values smaller than 1 indicate a deficit of risk. Since each observation, is divided to the expected counts given the structure of the population, this variable has no unit, and comparisons between areas can be done.

Again, the 95% credible intervals indicate the uncertainty associated to the estimate. Here we notice that relative risk is the parameter of a Poisson distribution,

* If , confidence intervals are find by using the Chi-squared method.

For instance, if is the upper bound, then, by definition, for any , .

We know that , with . Consequently,

* + upper 95% CI
  + lower 95% CI .
* If , a Gaussian approximation of the log relative risk is done, is assumed to follow a Gaussian distribution with mean , and variance . Then,
  + lower 95% CI
  + upper 95% CI

*Variable names:*

|  |  |
| --- | --- |
| Column name | Description |
| EXP\_UNADJ | Expected count adjusted for age and sex only |
| RR\_UNADJ | SMR estimate of the relative risk adjusted for age and sex only |
| RRL95\_UNADJ | Lower bound of the 95% CI of the SMR estimate of the relative risk, adjusted for age and sex only |
| RRU95\_UNADJ | Upper bound of the 95% CI of the SMR estimate of the relative risk adjusted for age and sex only |
| EXP\_ADJ | Expected count adjusted for age, sex and adjustment variables |
| RR\_ADJ | SMR estimate of the relative risk adjusted for age sex and adjustment variables |
| RRL95\_ADJ | Lower bound of the 95% CI of the SMR estimate of the relative risk, adjusted for age, sex and adjustment variables. |
| RRU95\_ADJ | Upper bound of the 95% CI of the SMR estimate of the relative risk, adjusted for age, sex and adjustment variables. |

***Empirical Bayes***

The maps of the standardised mortality or incidence ratio may lead to misinterpretations, since the extreme values are more often the consequence of small counts than a true extreme relative risk. Consequently, a non-significantly positive standardized mortality risk may be higher than a significant one for which the population at risk is higher. To reduce the influence of the small counts, Clayton and Kaldor (1987) proposed empirical Bayes estimates of the relative risk. They are based on a Poisson-Gamma hierarchical model. By assuming that all the relative risks are sampled in the same gamma distribution, their estimates are smoothed toward a global value. Moreover, the smaller is the count, the stronger is the shrinkage effect. In detail, relative risks are assumed to come from a single Gamma distribution of scale and shape ,

.

Approximation of the posterior mean of relative risk are given by the empirical Bayes estimates,

where and   are the maximum likelihood estimates of and .

In practice, these estimates of the relative risk are obtained through the following iterative procedure:

1. Start with initial values for the relative risk . For instance .
2. Obtain estimators   and using equations and
3. Obtain new estimated values for the relative risks
4. Repeat steps 2 and 3 until estimated values for and   do not change significantly.

*Variable names:*

|  |  |
| --- | --- |
| Column name | Description |
| SMRR\_UNADJ | Empirical Bayes estimate of the relative risk, adjusted for age and sex only |
| SMRR\_ADJ | Empirical Bayes estimate of the relative risk, adjusted for age, sex and adjustment variables. |

***Bayesian smoothing***

If we assume spatial dependence of the risk, meaning that the risks in close areas are similar, we can estimate the risk in one area with information borrowed to close areas. Since these estimates are based on more information than SMR estimates, they are more robust. The BYM model (Besag, York, and Mollié 1991) has been developed in this aim. First, one must consider a neighbouring graph which specifies the neighboured relationships between areas. Here, we denotes the set neighbours for area . The log relative risk is then split into two terms, a spatial term which accounts for the spatial variations within the risk, and a noise which accounts for independent local variations,

The spatial term is modelled with an intrinsic conditional auto-regressive model, meaning that for each area , is given conditionally to its neighbours. Specifically, it is assumed to follow a normal distribution with mean, the neighbour mean and variance, a variance parameter, , divided by the number of neighbours, ,

The larger is the number of neighbours, the smaller is the variance. The Gaussian noises, denoted , are supposed to be independent and identically distributed with variance .

One can consider the map of the uniquely spatially structured term given by (). They display uniquely the part of the risk which has a smooth spatial distribution. The independent term is then seen as residual. But one can also consider the estimate of the entire relative risk which leads to robust estimate thanks to the spatial CAR term, but also allows individual variability.

Other models are also proposed in the RIF, the ‘CAR’ model, in which the log relative risk is only modelled by the conditional autoregressive term. In this model, the relative risk is assumed to be spatially smooth without any local independent variations. In the CAR and the BYM model, estimates are smoothed towards a local mean.

Finally the ‘HET’ model is also proposed. In this model, the log relative risk is only composed of the independent term . As for empirical Bayes estimates, with this model the log relative risk estimates are smoothed toward a global mean. These two models are thus similar, they only differ by their prior (Gaussian vs gamma), and their inference method (fully Bayesian vs empirical Bayes).

*Prior specification*

Independent inverse gamma priors are assigned to the model parameters and .

*Variable names:*

|  |  |
| --- | --- |
| Column name | Description |
| BYM\_ssRR\_UNADJ | Relative risk estimate obtained with the spatially structured term of the BYM model uniquely, adjusted for sex and age only. |
| BYM\_ssRRL95\_UNADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the spatially structured term of the BYM model uniquely, adjusted for sex and age only. |
| BYM\_ssRRU95\_UNADJ | Upper bound of the 95% CI of the relative risk estimated obtained with the spatially structured term of the BYM model uniquely, adjusted for sex and age only. |
| BYM\_RR\_UNADJ | Relative risk estimate obtained with the BYM model, adjusted for sex and age only. |
| BYM\_RRL95\_UNADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the BYM model, adjusted for sex and age only. |
| BYM\_RRU95\_UNADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the BYM model, adjusted for sex and age only. |
| HET\_RR\_UNADJ | Relative risk estimate obtained with the HET model, adjusted for sex and age only. |
| HET\_RRL95\_UNADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the HET model, adjusted for sex and age only. |
| HET\_RRU95\_UNADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the HET model, adjusted for sex and age only. |
| CAR\_RR\_UNADJ | Relative risk estimate obtained with the CAR model, adjusted for sex and age only. |
| CAR\_RRL95\_UNADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the CAR model, adjusted for sex and age only. |
| CAR\_RRU95\_UNADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the CAR model, adjusted for sex and age only |
| BYM\_ssRR\_ADJ | Relative risk estimate obtained with the spatially structured term of the BYM model uniquely, adjusted for sex, age and adjustment variables. |
| BYM\_ssRRL95\_ADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the spatially structured term of the BYM model uniquely, adjusted for sex, age and adjustment variables. |
| BYM\_ssRRU95\_ADJ | Upper bound of the 95% CI of the relative risk estimated obtained with the spatially structured term of the BYM model uniquely, adjusted for sex, age and adjustment variables. |
| BYM\_RR\_ADJ | Relative risk estimate obtained with the BYM model, adjusted for sex, age and adjustment variables. |
| BYM\_RRL95\_ADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the BYM model, adjusted for sex, age and adjustment variables. |
| BYM\_RRU95\_ADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the BYM model, adjusted for sex, age and adjustment variables. |
| HET\_RR\_ADJ | Relative risk estimate obtained with the HET model, adjusted for sex, age and adjustment variables. |
| HET\_RRL95\_ADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the HET model, adjusted for sex, age and adjustment variables. |
| HET\_RRU95\_ADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the HET model, adjusted for sex, age and adjustment variables. |
| CAR\_RR\_ADJ | Relative risk estimate obtained with the CAR model, adjusted for sex, age and adjustment variables. |
| CAR\_RRL95\_ADJ | Lower bound of the 95% CI of the relative risk estimate obtained with the CAR model, adjusted for sex, age and adjustment variables. |
| CAR\_RRU95\_ADJ | Upper bound of the 95% CI of the relative risk estimate obtained with the CAR model, adjusted for sex, age and adjustment variables. |