HIERARCHICAL BAYESIAN MODEL WITH HEALTH COUNT DATA:
MALIGNANT MELANOMA MORTALITY IN THE EUROPEAN COMMUNITY

Giulia Lumia, Marco D'Antoni

ABSTRACT

A hierarchical Bayesian approach is adopted in order to model the variability in the mortality rates from malignant melanoma among the nine nations within the European Community (1980). Previous beliefs are updated by the mean of the available data, allowing an assessment of the mortality trend: Western Germany, Denmark, UK, Luxembourg and Netherland seem to need some intervention.
The geographical pattern of the disease within Europe shows some marked discrepancies that may be explained by the uvb radiations dose.

INTRODUCTION

This report is aimed to show the results of an analysis carried out in order to:

- investigate the rate of Malignant Melanoma Mortality in the European Community;
- understand the mortality rate at a European geographical level;
- assess nations' need for a dedicated policy to reduce the mortality risk.

Data were collected between 1971 and 1980 with respect to the EC-9 nations.
They refer not only to the observed number of deaths due to malignant melanoma, but also to the expected one, i.e. how many deaths were expected to happen in that specific time interval on the base of a priori knowledge. Moreover, a UVB dose value was also collected for each county, since that exposure to ultraviolet radiation from sunlight is well-known to be an important risk factor of malignant melanoma.

The dataset is made up of 6 variables:

- **Nation** (Belgium, Western Germany, Denmark, France, UK, Italy, Ireland, Luxembourg, Netherlands);
- **Region** ID;
- **County** ID;
- **Observed deaths** due to malignant melanoma;
- **Expected deaths** due to malignant melanoma;
- **UVB dose** reaching the earth's surface in each county.

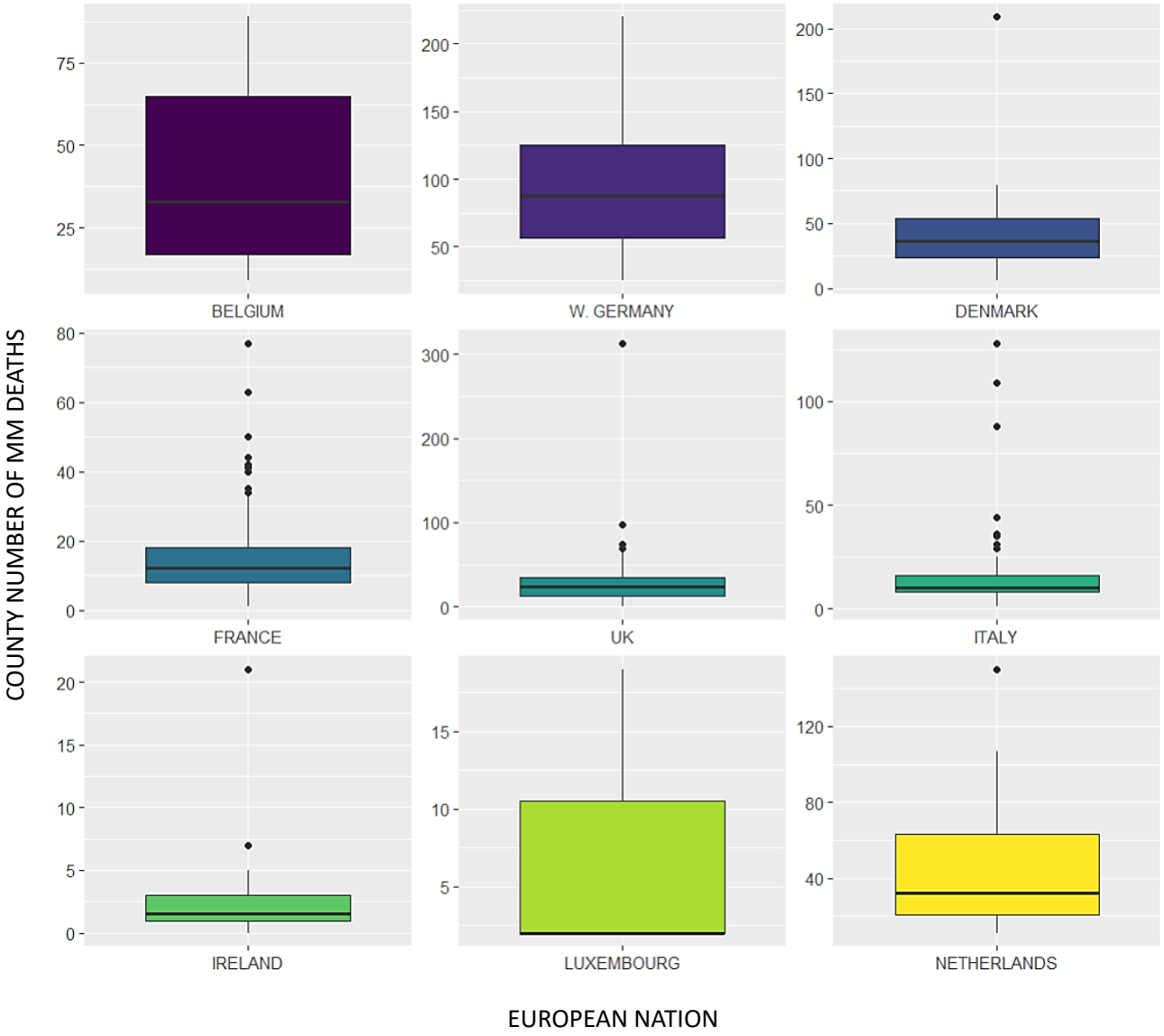
The analysis develops in three main stages:

1. Exploratory analysis;
2. Hierarchical Bayesian model adaptation;
3. Further analysis developments.

1. EXPLORATORY ANALYSIS

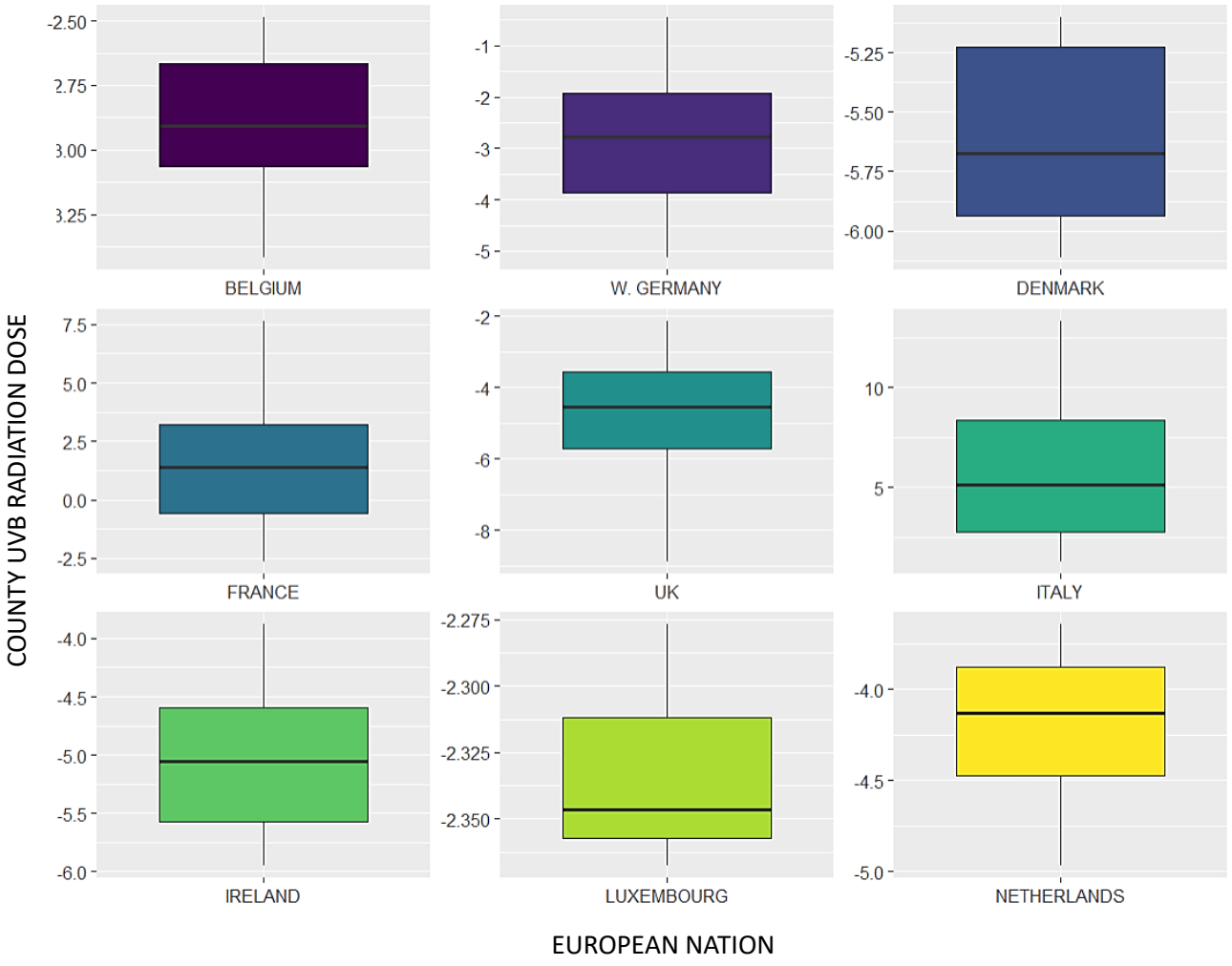
Three levels of geographical aggregation are available (national, regional and provincial).
According to the main aims of the analysis only the national level is considered later on.
However, before carrying out such an aggregation it is logical to explore provincial data conditional to each nation.
PLOT 1.1 shows the distribution of the number of deaths due to MM (malignant melanoma) in the counties of each of the 9 nations considered, during the period 1971-1980.
It is possible to notice that not only the median number of deaths but also the variation range change when the nation changes too: e.g. in Italy the median number of deaths in each county (province) was 10, while the 75% of all the Italian counties didn't register more than 16 deaths.
On the other hand, nations like western Germany registered an higher mortality (the median german value is about 80 deaths per county).
However, one should take into account that these values are absolute, due to the lack of information about the population density of each county.

PLOT 1.1: DISTRIBUTION OF DEATHS IN THE COUNTIES OF EACH NATION.



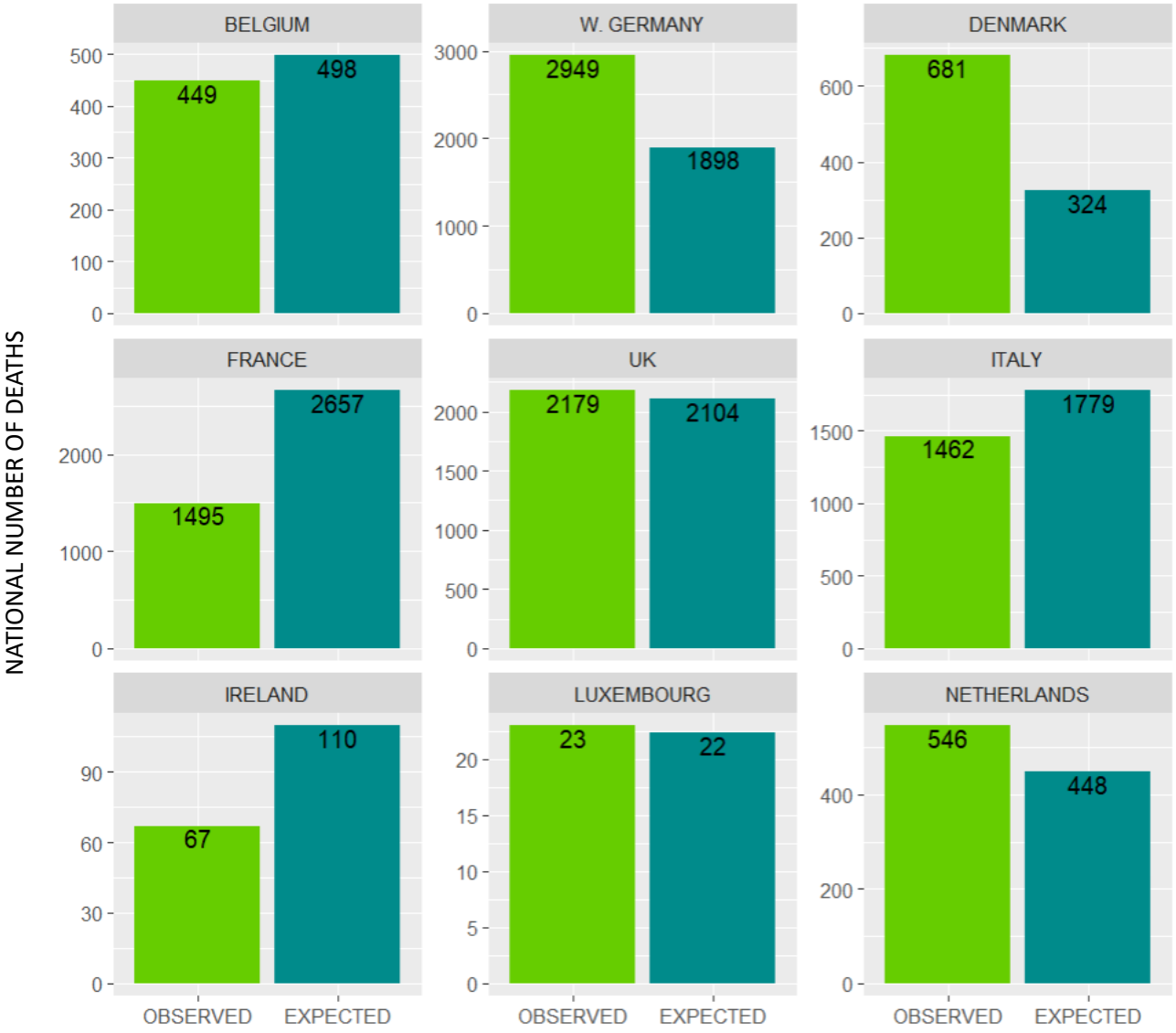
The differences highlighted before may depend on a wide range of reasons. The data collected allow to consider the effect of the UVB radiations dose whose national distribution is shown in PLOT 1.2. The median provincial radiations dose seems to change with the latitude of the country considered. In fact, southern countries like Italy and France receive the highest dose of UVB radiation (around 5 and 1.25 respectively) whereas northern countries like Denmark and Ireland the lowest. Also in this case national counties register some variability about the UVB dose.
This may be related to the longitudinal expansion of the country they belong to, but more precise further analysis may be needed in order to assess this relationship.

PLOT 1.2: DISTRIBUTION OF UVB DOSE IN THE COUNTIES OF EACH NATION.



PLOT 1.3 highlight the difference between the observed number of deaths due to malignant melanoma and the expected number on the base of a priori knowledge i.e. values computed before observing deaths in 1971-1989, both at a national aggregated level.
While the magnitude of each singular value is poorly informative alone (it is absolute), the comparison between reality and expectations gives rise to interesting considerations.
It is possible to notice that the situation varies from country to country:
in nations like Belgium, France, Ireland and Italy the mortality was less severe then expected;
on the contrary nations such as Denmark and Western Germany registered an higher fatality (the expected number of deaths is less than the real one).

PLOT 1.3: REAL vs EXPECTED NUMBER OF DEATHS IN EACH NATION.



2. HIERARCHICAL BAYESIAN MODEL

Hierarchical models are commonly adopted when the i.i.d. assumption doesn't hold.
For many applications, in fact, some observations share characteristics that distinguish them from other observations, therefore multiple distinct groups are observed.
The malignant melanoma dataset presents a natural spatial hierarchy since the geographical levels considered are nested among them.

EUROPEAN COMMUNITY → NATIONS → REGIONS → COUNTIES

According to the particular type of variable to model i.e. the number of deaths, the choice fell on a **Gamma-Poisson hierarchical Bayesian model**: the model assumes that the number of deaths due to MM in each j-th country can be modelled by a Poisson distribution.
All the j-th parameters come from a gamma distribution whose parameters (called hyperparameters) have their own prior distributions (called hyperprior distributions).

The full model specification consists of:

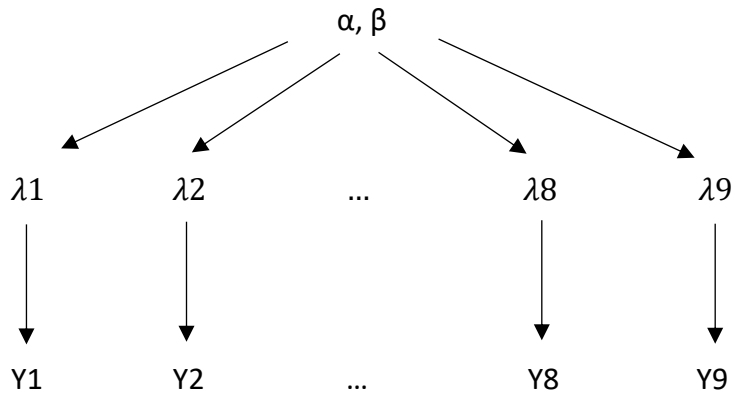
$Yj|\lambda j \sim Poisson(\lambda j) \quad j = 1, \dots 9$ ← SAMPLING DISTRIBUTION FOR THE NUMBER OF DEATHS IN THE j^{th} NATION

$\lambda j|(\alpha, \beta) \sim Gamma(\alpha, \beta)$ ← PRIOR DISTRIBUTION FOR THE PARAMETERS λ_j

$\alpha \sim Gamma(a, b)$ ← HYPERPRIOR FOR THE HYPERPARAMETER α

$\beta \sim Gamma(c, d)$ ← HYPERPRIOR FOR THE HYPERPARAMETER β

The hierarchical model specified can be represented by the mean of a DAG, highlighting a crucial assumption: data depends on the hyperparameters only through the parameters λ_j.



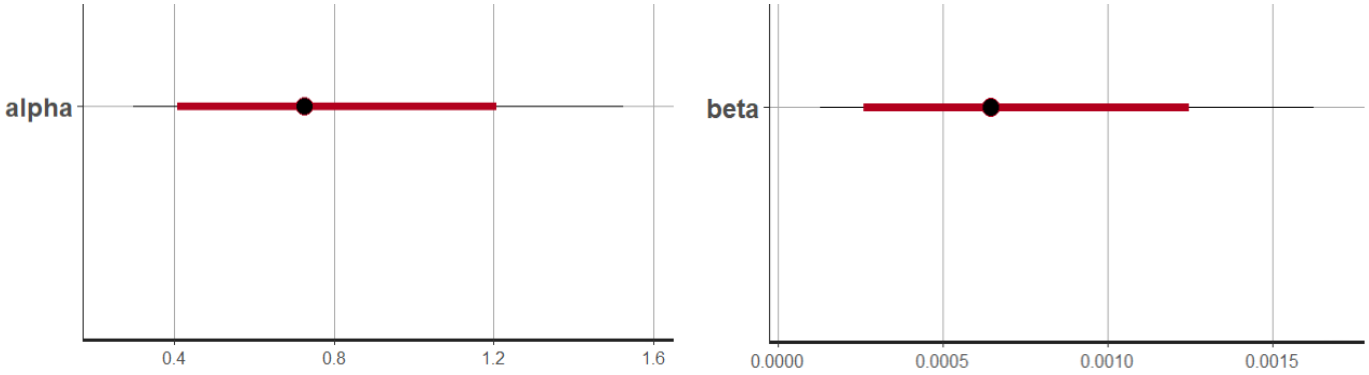
The values of the hyperparameters were chosen on the base of prior knowledge, that is represented in this case by the expected number of deaths at a national level.
They were set in order to have an expectation almost equal to the mean of the prior parameters and a variance almost equal to their variance.

$|E[\lambda|(\alpha, \beta)] = \frac{\sum_{j=1}^{9} \lambda j}{9}$ ← EXPECTATION OF THE GAMMA DISTRIBUTION GENERATING λj

The posterior distribution of the parameters given the observed data and the prior information cannot be derived analytically, thus a sampling algorithm is needed:
the chosen algorithm, namely NUTS, is based on a MCMC (Markov Chain Monte Carlo) in which the dependence is break up through the thinning method.
The resulting simulated posterior distribution can be used to obtain some a posteriori quantities.

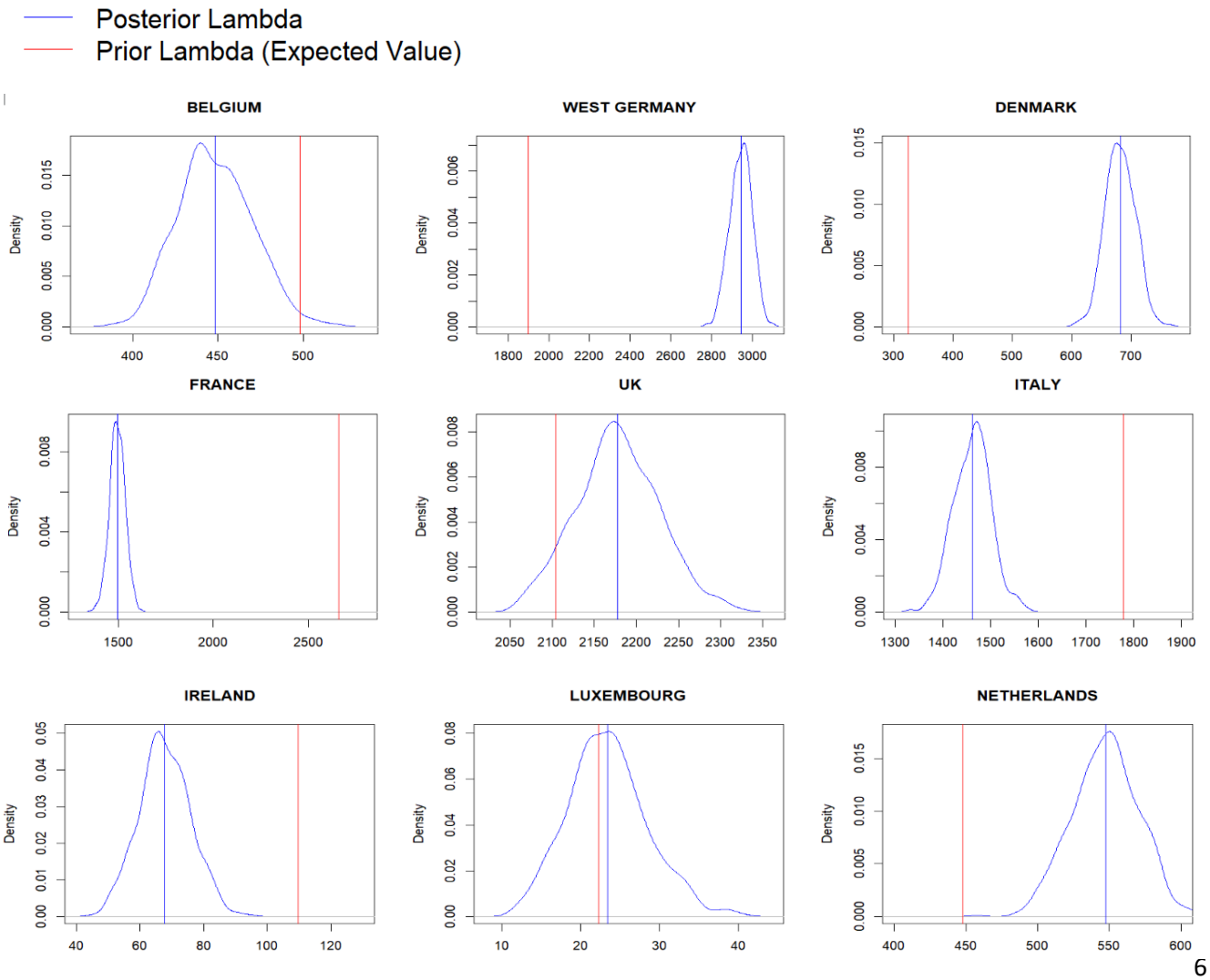
PLOT 2.1 shows the posterior hyperparameters and their corresponding Bayesian confidence intervals. Taking into account the observations Yj it is possible to say that the expected number of deaths due to MM for the generic European country is about 1128.

PLOT 2.1: BAYESIAN CONFIDENCE INTERVALS FOR THE HYPERPARAMETERS



After having sampled the posterior distribution of the parameters it is also possible to do a comparison between the prior expectation for each country, i.e. the expected number of deaths due to MM without taking into account the collected data, with the posterior expectation.
In this way one can notice the effect of including the evidence into the prior knowledge.

PLOT 2.2: COMPARISON BETWEEN PRIOR AND POSTERIOR EXPECTED NUMBER OF DEATHS BY COUNTRY



PLOT 2.2 can be interpreted as follows:
Western Germany, Denmark, UK, Luxembourg and Netherlands registered an increase in the value of the expected number of deaths when the collected data are combined with the prior ones. These country seems to need some intervention policies in order to reduce the fatality of the Malignant melanoma.
On the contrary, the remaining countries shown an totally opposite situation.

3. FURTHER ANALYSIS DEVELOPMENTS

The variance in melanoma rates between areas can then be modelled as functions of explanatory variables, including UVB dose, at different levels of a geographical hierarchy.
This analysis only focuses on mortality according to the country, but it is reasonable to include also the doses of UVB arriving at the surface in different countries.
All things considered, further developments of the current analysis include:
- taking population density into account;
- taking UVB radiations dose into account;
- taking the latitude expansion of each country into account.