Modelling Short-Term Effects of Meteorological Variables on Mortality[1]

Modelli statistici per lo studio degli effetti a breve termine delle variabili metereologiche sulla mortalità

Annibale Biggeri, Michela Baccini

Dipartimento di Statistica, Università di Firenze
Unità di Biostatistica CSPO, Firenze

Riassunto: Diversi studi epidemiologici mettono in evidenza che l'esposizione a condizioni meteorologiche estreme ha effetti dannosi sulla salute umana. In questo lavoro viene valutato l'effetto della temperatura apparente (un indice ottenuto come combinazione di temperatura e umidità) sulla mortalità durante la stagione calda in 15 città europee partecipanti al progetto PHEWE. In particolare l'interesse è sull'analisi della relazione esposizione-risposta, sulla definizione di un indice sintetico di effetto delle alte temperature apparenti e sullo studio dell'effetto ritardato dell'esposizione ed eventuali fenomeni di anticipazione del decesso. L'analisi città-specifica è basata sull'approccio GEE. Un modello di meta-analisi ad effetti casuali Bayesiano è stato utilizzato per combinare i risultati città-specifici.

Keywords: Generalized Estimating Equations, Hierarchical Bayesian model, Meta-analysis, Transfer Function.

1. Introduction

Several epidemiological studies indicate that exposure to extreme meteorological conditions is associated with an increase in mortality, but further investigation is necessary in order to address relevant issues, such as the pattern of exposure-response relation and the lagged effect of heat (see for example Kunst *et al.*, 1993; Greenberg *et al.*, 1983; Saez *et al.*, 1995; Khaw, 1995; Pan *et al.*, 1995). In this paper we evaluated the health effects of apparent temperature (a combination of the dry bulb and dew point temperatures) during warm season in 15 European cities participating to the PHEWE project. We focused on study of the city-specific exposure-response curves describing relationship between apparent temperature and mortality and on investigation of delayed effect of exposure, in order to detect possible mortality displacement phenomena. Results from different cities were compared and combined using random effects meta-analysis models.

In section 2, a brief description of data is provided. In section 3, the modelling approach used for the city-specific analysis is described and in section 4 the meta-analysis

---

[1] The PHEWE project is supported by the European Commission, DG Research within the 5th Framework programme, contract N°QLK4-CT-2001-00152.

methods adopted to compare the city-specific results are reported. In section 5 we show the main results of the study and in section 6 we discuss them.

2. Data

The analysis was performed on 15 European cities: Athens, Barcelona, Budapest, Dublin, Helsinki, Ljubljana, London, Milan, Paris, Praha, Rome, Stockholm, Turin, Valencia, Zurich. All cities provided age specific (0-64, 65-74, 75+) daily counts of deaths for all causes and for cardiovascular and respiratory causes and 3-hourly meteorological data (namely temperature and humidity) retrieved from the nearest airport weather station. Data on several confounders including other meteorological variables and air pollution variables were also provided.

Analyses were carried out for warm season, defined as the period ranging from April to September. Apparent temperature (a combination of the dry bulb and dew point temperatures, ref) were computed for each city. 3 hour maximum apparent temperature was considered as daily exposure indicator, except for Barcelona for which 3 hour measurements were not available and mean apparent temperature was used .

Data covered the period 1990-2001 and time series length ranged from 6 to 11 years. A large variability in exposure and in mortality was observed in the cities involved (Tab.1).

Table 1. *Number of inhabitants, study period, average daily number of deaths for all natural causes and mean of daily maximum apparent temperature during warm season (April-September) by city.*

City	Population	Study Period	Average Daily Number of Deaths	Mean of Daily Maximum Apparent Temperature
Athens	3 188 305	1992-1996	67.5	27.6
Barcelona	1 512 971	1993-2000	35.9	23.3
Budapest	1 797 222	1992-2001	71	21.9
Dublin	481 854	1990-2000	11.4	14.7
Helsinki	955 143	1990-2000	17	14.3
Ljubljana	263 290	1992-1999	6.3	20.1
London	6 796 900	1992-2000	149	18.0
Milan	1 304 942	1990-1999	26.3	25.4
Paris	6 161 393	1991-1998	115.7	19.5
Praha	1 183 900	1992-2000	34.9	17.8
Rome	2 812 573	1992-2000	52.8	26.1
Stockholm	1 173 183	1990-2000	27.8	15.4
Turin	901 010	1991-1999	19.1	23.4
Valencia	739 004	1995-2000	14.6	29.5
Zurich	990 000	1990-1996	11.6	19.0

3. Statistical analysis: city-specific analysis

The city-specific analysis was carried out in two steps. At the first step, an exploratory analysis was performed in order to obtain information to be used in building the base model and to highlight common characteristics among cities. At the second step, a Generalized Estimating Equations approach was used to obtain city–specific results.

The exploratory analysis was based on Transfer Function approach (Chiogna and Gaetan, 2003). TF models are able to deal with many potential covariates, lagged effects, seasonal patterns, autoregressive and moving average terms. We combined this approach with a model selection strategy based on Genetic Algorithm (GA), in order to choose the “best” model out of a large set of competing models. We performed TF analyses for each city, for each year and season, for each outcome. Our principal interest was in the autoregressive order selected by GA, which could range from 0 (independence of errors) to 3.

At the second stage, a Generalized Estimating Equations approach was used (Diggle *et al.*, 1994). We modelled the marginal relation between daily mortality count and apparent temperature, adjusted for confounders, assuming independence among summers and treating serial dependence of the daily number of deaths within each summer as a nuisance. On the basis of the TF exploratory analysis results (not reported), an AR(1) structure of correlation within summer was specified.

The inference on the exposure-response curve was performed using lag-average indicators of exposure: the mean of current day and previous 3 days maximum apparent temperature (lag 0-3).

A flexible parametric approach was adopted to model the relationship, introducing in the model a cubic regression spline with a pre-defined number of knots placed at the percentiles of the exposure variable distribution.

The following model was assumed:

$$y_i \sim Poisson(\lambda_i) \qquad \log(\lambda_i) = \alpha + s(AppT_i) + \sum_k \beta_k x_{ki} \ ,$$

where y_i indicates the number of deaths on day $i, (x_{1i}, x_{2i}, ..., x_{ki})$ is a vector of covariates and $s(AppT)$ a cubic regression spline of exposure variable with 1 knots every 8 units of apparent temperature.

We adjusted for the effect of seasonality and time-related confounders including in the model indicators for month, day of week and holidays and linear and quadratic terms for long term time trend. We controlled for the air pollution and meteorological confounding by linear terms for daily NO_2 concentration (Black Smoke for Dublin) (lag 0-1), mean sea level pressure (lag 0-3) and wind speed (lag 0).

Separate analyses were performed for all ages mortality and for each age group. Mortality for all natural causes and for respiratory, cardiovascular and cerebrovascular causes was considered.

Due to “V” shapes with different minimum points were observed for exposure-response curves, in a further analysis we described the relationship between apparent temperature and mortality during summer by two linear terms constrained to join in a point chosen on the basis of the city specific exposure-response curve. Then we used the estimated slope above the breakpoint as an indicator of high temperature exposure effect.

In the basic analysis, city-specific thresholds were obtained by model, treating the apparent temperature corresponding to the minimum of exposure-response curve as a unknown parameter (Muggeo, 2003). Analyses performed specifying a common breakpoint for all the cities or different breakpoints defined by visual inspection of dose-response curves, showed a strong sensitivity of results to the definition of the threshold. In particular, using common breakpoint bring to underestimation of the slope when the actual minimum of exposure-response curve was placed over the chosen threshold (results not reported).

Influence of extreme values of apparent temperature on the slope estimate was checked introducing in the model an indicator of very high apparent temperature days, defined as the days with maximum apparent temperature (lag 0-3) higher than the 90th percentile.

The delayed effect of exposure was studied by restricted distributed lag models in a GEE framework. Modeling the apparent temperature effect by a "V" shape function, we studied the lagged effect of the high exposure linear term up to 40 days, using a 5 degree polynomial constrain:

$$\log(\lambda_i) = \alpha + \sum_k \beta_k x_{ki} + \lambda_0 y_i + \lambda_1 y_{i-1} + \ldots + \lambda_L y_{i-40} \qquad \lambda_\ell = \sum_{h=0}^{5} \eta_h \ell^h ,$$

where $\lambda_0, \lambda_1, \ldots, \lambda_{40}$ are unknown coefficients describing the exposure effect from lag 0 to lag 40 and $\eta_0, \eta_1, \ldots, \eta_5$ are the coefficients that define the constrain. Introducing a constrain reduces the problems related to collinearity between lagged covariates and produces more stable estimates of the λ_ℓ coefficients. The estimation of the model was performed using standard software for GEE estimation having recourse to the Almon's parametrization (Almon, 1965). As a sensitive analysis, alternative constrain definitions varying degrees of polynomial or using spline based approach (Green and Silvermann, 1994) were considered (not reported).

All the statistical analyses were performed using STATA (*gee* function for GEE estimation) and R software version 1.8 (*smiso* and *genstep* libraries for TF analysis and *mgcv* library for spline basis construction). Ad hoc codes were written for distributed lag analysis.

4. Combining city-specific results

Part of the city-specific results were combined using meta-analytic methods. In order to reduce heterogeneity among cities, results were combined separately by three groups of cities, a priori defined on the basis of meteorological and geographical criteria by Kassomenos (see the PHEWE project protocol). The groups were the following:

- Group 1 (Mediterranean cities): Athens, Rome, Barcelona, Valencia, Torino, Milano, Ljubljana;
- Group 2 (Continental cities): Prague, Budapest, Zurich, Cracow, Paris, Helsinki, Stockholm;
- Group 3: (North Atlantic cities): London, Dublin.

For each group, a Bayesian random effects meta-analysis model was specified (Gelman *et al.*, 1995; Sutton and Abrams, 2001). We assumed that the city-specific estimates of the slope were independent realizations from Gaussian populations with mean λ_c and known variance $\hat{\sigma}_c^2$.

$$\hat{\lambda}_c \sim N\left(\lambda_c, \hat{\sigma}_c^2\right)$$

and that each city-specific effect λ_c was drawn from a Normal population with mean λ and variance τ^2:

$$\lambda_c \sim N\left(\beta, \tau^2\right)$$

λ represents the average effect of high temperature adjusted for the inter-city variation τ^2. The Bayesian formulation needs to specify prior distributions on hyper-parameters λ and τ^2. We placed upon these parameters vague proper priors:

$$\lambda \sim N\left(0{,}10^5\right)$$

$$\tau^2 \sim IG(0.001{,}0.001).$$

Posterior distributions of parameters have been obtained with WinBUGS (Spiegelhalter *et al.,* 1996). 100,000 iterations (excluding the first 4,000) were retained. Achieved convergence was assessed using the Gelman and Rubin approach (Gelman and Rubin, 1992), based on 3 parallel chains.

5. Results

Tables and Figures in this paper are relative to the effect of apparent temperature on all age mortality for all natural causes.

Figure 1 reports the estimated curves describing the relationship between maximum apparent temperature (lag 0-3) and mortality in a *log* scale. For most of the cities enrolled, the estimated curve showed a clear “V” shape, with city-dependent minimum.

On the basis of these results it appeared appropriate specifying a simpler model with two linear terms constrained to joint in a breakpoint. Table 2 reports the Maximum Likelihood estimates of breakpoints by city. They ranges from 21-22 (Barcelona and Praha) to 33 (Rome and Athens) units of maximum apparent temperature. Greater uncertainty was around threshold estimates in Valencia and Zurich. For Dublin and Ljubljana not significant breakpoints were found.

Linear effects above threshold were very consistent among cities (descriptive thresholds of 21 units of apparent temperature based on visual inspection of exposure-response curves were adopted for Dublin and Ljubljana) and in most of cases the slope estimates appeared to be robust when a dummy variable for extreme apparent temperature days was included in the model (Tab. 2).

Figure 2 reports the scatter plot of estimated slope in respect to the estimated breakpoint for the cities enrolled in the study. Rome, Athens and Milan showed the highest values of slope and threshold.

Distributed lag analysis up to 40 days showed an excess risk concentrated in the first days and limited evidence of harvesting at days 5-20 (Fig. 3). Different specification of constrain did not bring to different results (not reported).

Figure 1. *Estimated relationship between maximum apparent temperature (lag 0-3) and mortality for all natural causes with related 95% point confidence intervals, by city.*

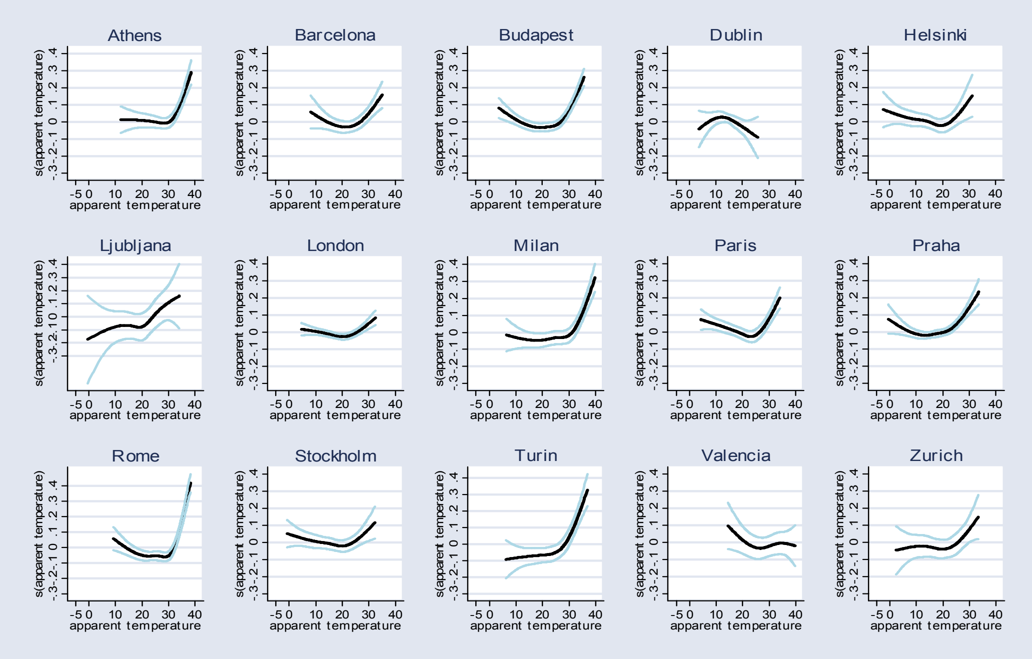


Table 2. *City-specific estimates of the threshold and related standard error and city-specific estimates of the slope above the estimated threshold (95% CI) under the basic model and adding in the model an indicator of days with apparent temperature higher than the 90th percentile.*

			Basic Model			**Model with indicator of high temperature days**		
City	**Threshold**	**se**	**Slope above Threshold**	**95% CI**		**Slope above Threshold**	**95% CI**	
Athens	32.7	0.3	0.054	0.042	0.066	0.053	0.034	0.072
Barcelona	21.3	0.9	0.014	0.009	0.019	0.009	0.003	0.015
Budapest	26.5	0.4	0.028	0.024	0.033	0.031	0.023	0.039
Dublin	-	-	*-0.011*	*-0.048*	*0.026*	*0.002*	*-0.038*	*0.042*
Helsinki	23.8	0.9	0.039	0.018	0.060	0.037	0.013	0.061
Ljubljana	-	-	*0.020*	*0.008*	*0.033*	*0.026*	*0.008*	*0.045*
London	24.9	0.6	0.020	0.013	0.026	0.019	0.011	0.028
Milan	31.7	0.6	0.041	0.032	0.050	0.052	0.035	0.069
Paris	24.9	0.5	0.023	0.018	0.029	0.021	0.011	0.031
Praha	22.4	0.8	0.020	0.014	0.025	0.021	0.010	0.031
Rome	33.0	0.2	0.093	0.081	0.104	0.109	0.087	0.130
Stockholm	23.5	1.2	0.019	0.008	0.030	0.017	0.002	0.031
Turin	28.1	1.0	0.037	0.028	0.046	0.037	0.021	0.053
Valencia	28.0	3.1	0.005	-0.004	0.014	0.010	0.000	0.020
Zurich	22.9	2.2	0.015	0.005	0.025	0.023	0.005	0.041

Figure 2. *Scatter plot of estimated slope in respect to the ML estimate of breakpoint by city.*

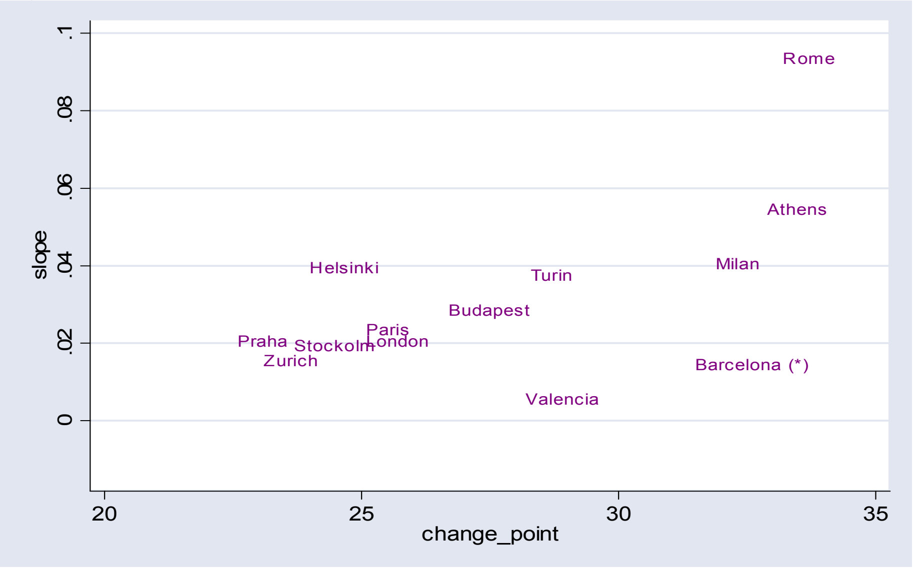
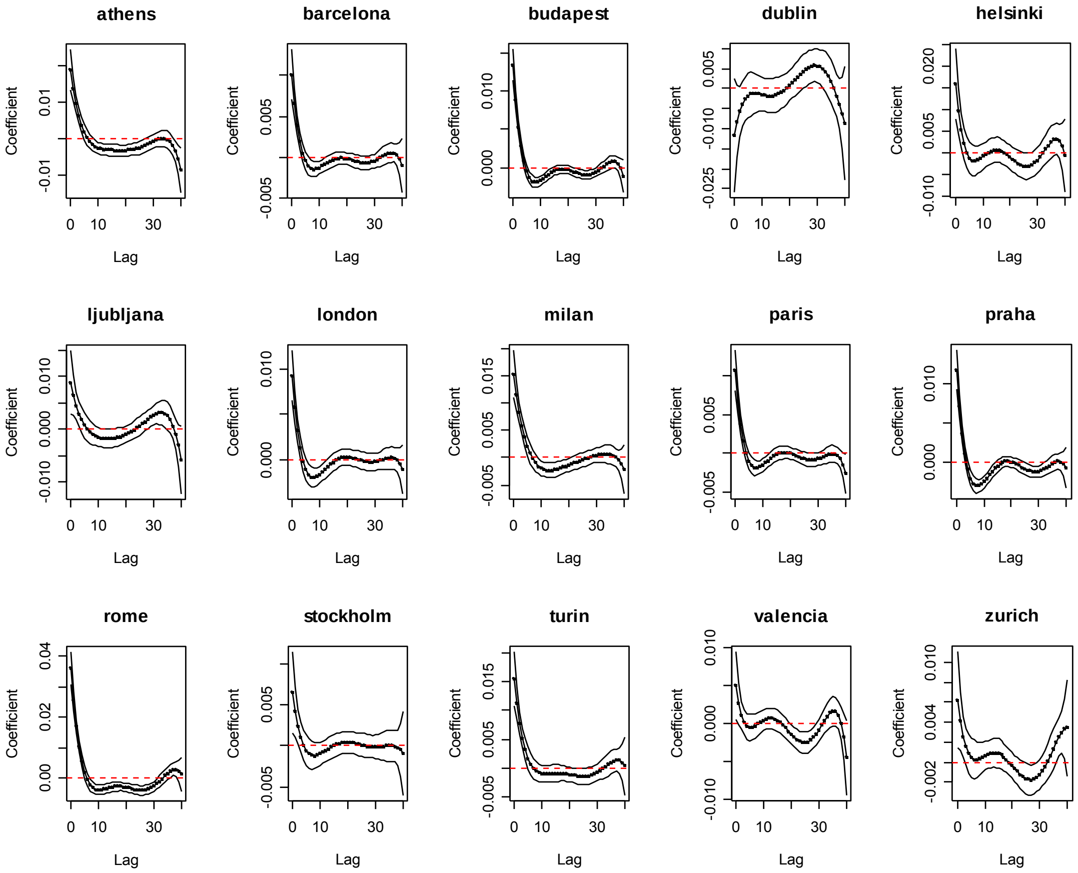


Figure 3. *Estimated lagged effect of apparent temperature above the threshold on total mortality, under a 5 degree polynomial constrain for the lag 0-40 coefficients.*



Results from the Bayesian meta-analysis models are reported in Figure 4. As expected, we observed the strongest effect of high apparent temperature in the Mediterranean cities, the overall estimate of the percent increase in mortality associated to 1 unit increase of apparent temperature being 3.8 (95% Credibility Interval: 1.1, 6.7). For the second and the third group, the overall estimate of percent increase was 2.3 (1.8, 2.8) and 0.2 (-8.2, 8.9), respectively. In interpreting the overall effect for the third group, it is necessary to take into account that the estimate was based only on two cities for which the sign of the slope was opposite.

In Figure 4 the posterior distribution of the heterogeneity variance for the first and the second group are reported. The amount of heterogeneity was larger among the Mediterranean cities than among the Continental cities. As a direct consequence, we observed a large variability of the overall estimate in the first group of cities.

Figure 4. *City-specific estimates (95% CI) and overall estimate (95% CrI) of the effect of high maximum apparent temperature on mortality for all natural causes (lag 0-3). Effect are expressed in terms of percent variation associated to 1 unite increase of apparent temperature.*

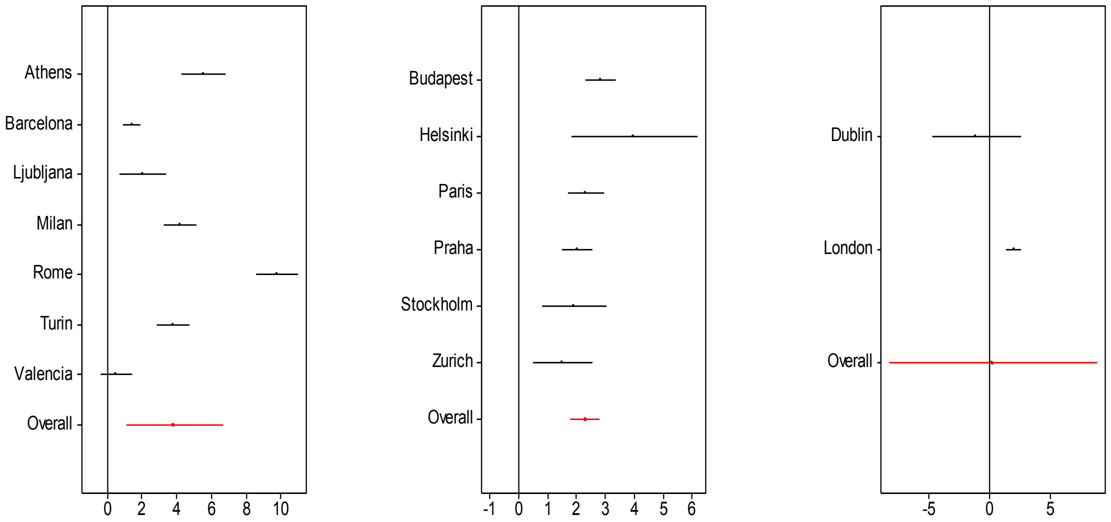
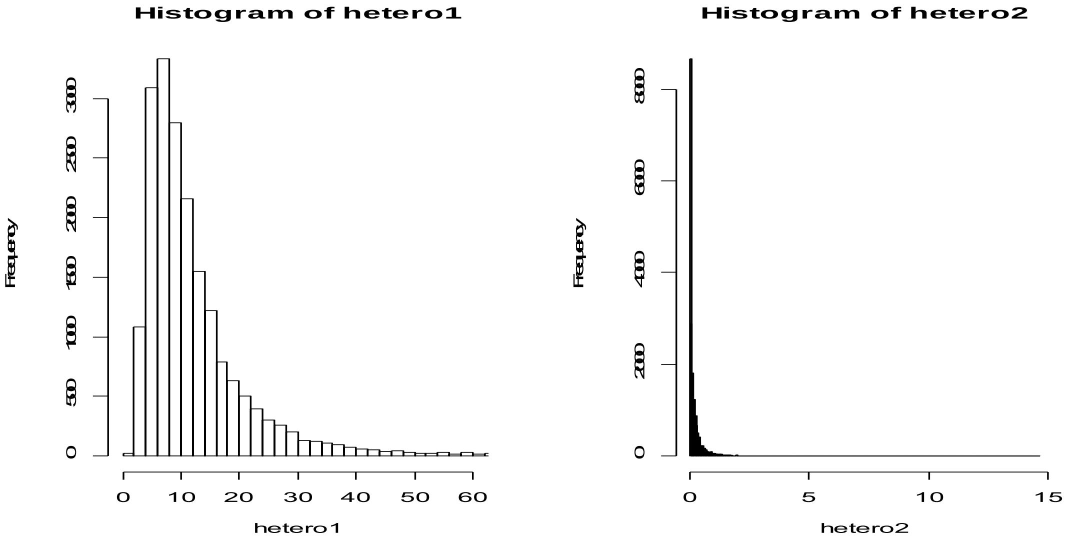


Figure 5. *Histograms of the posterior distributions of the heterogeneity variance among cities belonging in the first and in the second group.*



6. Discussion

We modelled the effect of apparent temperature on mortality. This exposure indicator, which combine temperature and dew point values, is not usually investigated in the literature.

In the basic analysis, maximum daily values of apparent temperature were considered. A sensitivity analysis using minimum apparent temperature produced shifted exposure-response curves with similar shape (not reported).

We proposed to summarize the city-specific exposure-response relationship by two linear terms constrained to join in a point. The results appeared very sensitive to the choice of breakpoint and using city-specific thresholds appeared to be more appropriate than using a common breakpoint.

The distributed lag analysis showed that the effect of adverse meteorological conditions may be immediate or may occur with some lag.

Bayesian meta-regression should be addressed in order to investigate the relevant heterogeneity of the high apparent temperature effects, in particular among Mediterranean cities.

References

Almon S. (1965) The distributed lag between capital appropriations and expenditures. *Econometrica*, 147:178-196.

Chiogna M., Gaetan C. (2003). Transfer function modelling strategy for combining evidence on air pollution and daily mortality. *In Proceedings of SCO2003,* Treviso 4-6 settembre 2003, 140-145.

Diggle P. J., Liang K. Y., Zeger S. L. (1994) *Analysis of longitudinal data*, Oxford: Clarendon Press.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (1995) *Bayesian Data Analysis,* Chapman and Hall.

Gelman A., Rubin D.R. (1992) Inference from iterative simulation using multiple sequence (with discussion). *Statistical Science*, 7: 457-511.

Green P.J., Silverman B.W. (1994) *Nonparametric Regression and Generalized Linear Models, A roughness penalty approach*, London, Chapman and Hall.

Greenberg J.H., Bromberg J., Reed C.M., Gustafson T.L., Beauchamp R.A. (1983) The epidemiology of heat related deaths. Texas-1950, 1970-1979, and 1980. *American Journal of Public Health*, 73: 805-807.

Khaw K.T. (1995) Temperature and cardiovascular mortality. *The Lancet,* 345: 337-8.

Kunst A.E., Looman C.W.N., Mackenbach J.P. (1993) Outdoor Air Temperature and Mortality in the Netherlands: A Time-Series Analysis. *American Journal of Epidemiology*, 137: 331-41.

Muggeo V.M.R. (2003) Estimating regression models with unknown break-points. *Statistic in Medicine*, 22:3055-3071.

Pan W.H., Li L.A., Tsai M.J. (1995) Temperature extremes and mortality from coronary heart disease and cerebral infarction in elderly Chinese. *The Lancet*, 345: 353-5.

Saez M., Sunyer J., Castellsague J., Murillo C., Anto J.M. (1995) Relationship between Weather Temperature and Mortality: A Time Series Analysis Approach in Barcelona. *International Journal of Epidemiology*, 24: 576-82.

Spiegelhalter D.J., Thomas A., Best N.G., Gilks W.R. (1996) *BUGS: Bayesian Inference Using Gibbs Sampling*, Medical Research Council Biostatistics Unit, Cambridge.

Sutton A.J., Abrams K.R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research,* 10:277-303.