2020 555 Statistical Methods for Spatial Epidemiology: Disease Mapping

Jon Wakefield
Departments of Statistics and Biostatistics, University of Washington

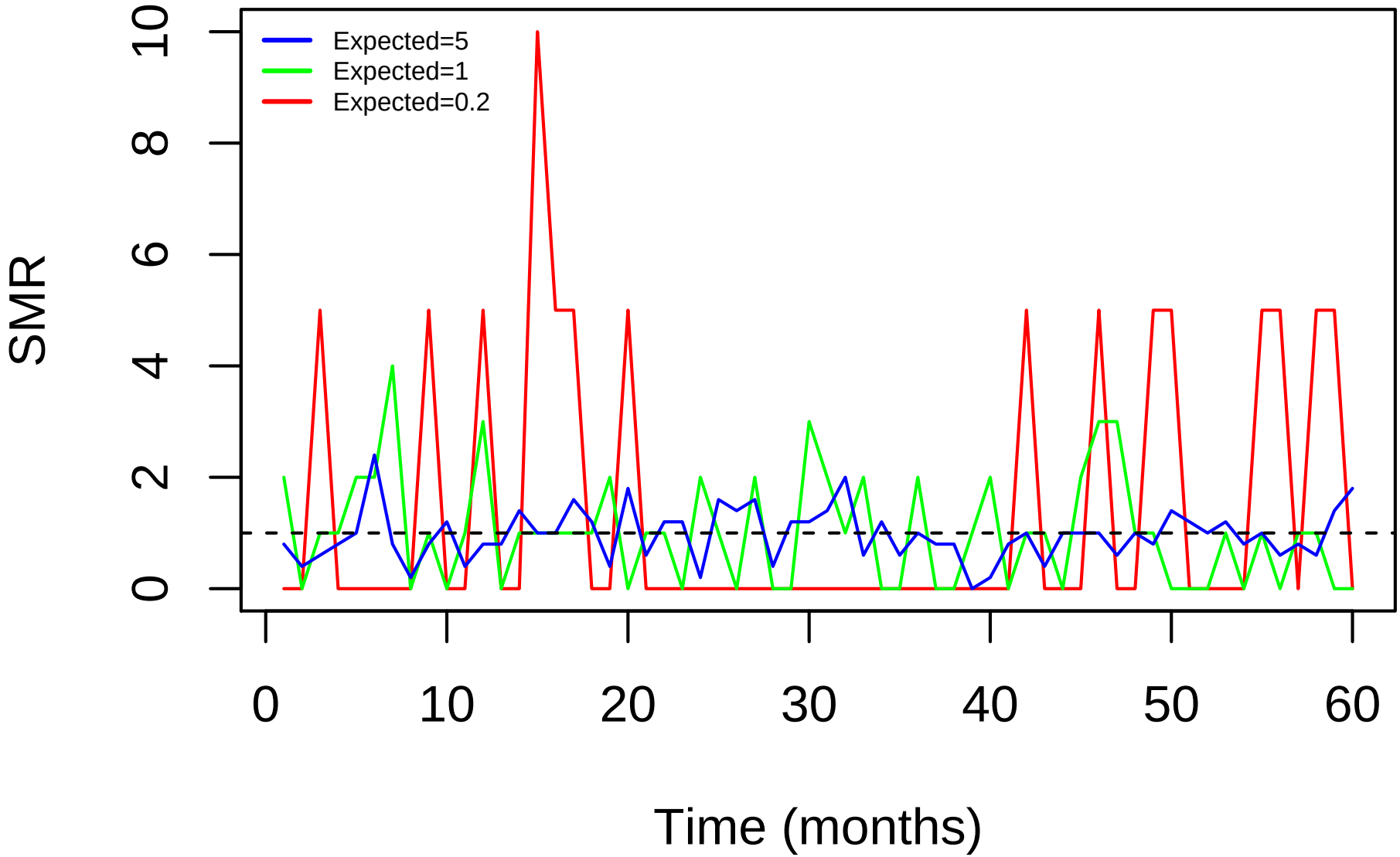
2020-01-31

Surveillance

Surveillance

We demonstrate the effect of different expected numbers in the null situation.

```
set.seed(2113)
time <- seq(1, 60, 1)
theta <- 1
Expect1 <- 5
Expect2 <- 1
Expect3 <- 0.2
plot(rpois(length(time), theta * Expect3)/Expect3 ~
    time, col = "red", type = "l", ylab = "SMR", xlab = "Time (m
lines(rpois(length(time), theta * Expect2)/Expect2 ~
    time, col = "green", add = T)
lines(rpois(length(time), theta * Expect1)/Expect1 ~
    time, col = "blue", add = T)
abline(h = 1, col = "black", lty = 2)
legend("topleft", legend = c("Expected=5", "Expected=1",
    "Expected=0.2"), col = c("blue", "green", "red"),
    lwd = 1.5, bty = "n", cex = 0.5)
```


Expected=5
Expected=1
Expected=0.2
SMR
0
2
4
6
8
10
0
10
20
30
40
50
60
Time (months)

SMR Estimates

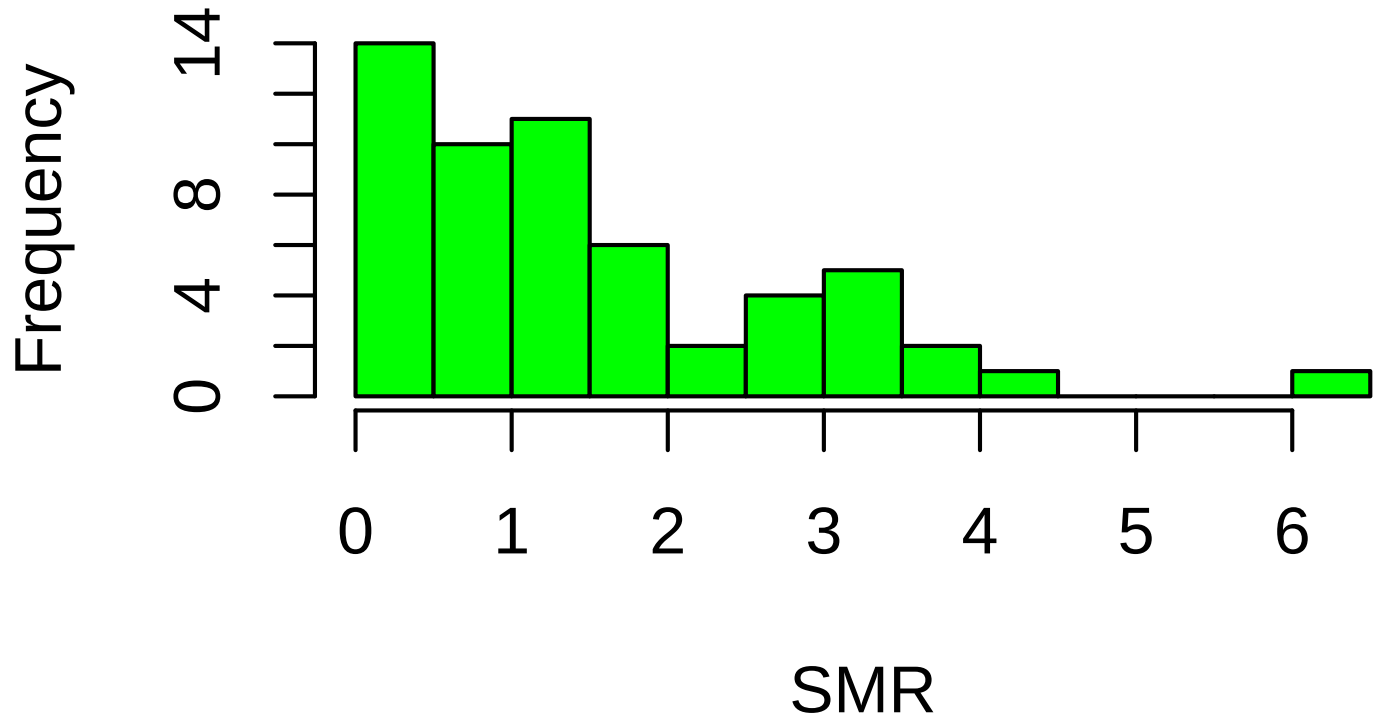
Scottish lip cancer data

We will first fit a number of models to the famous Scottish lip cancer data.

We have counts of disease, expected numbers and an area-based covariate (proportion in agriculture, fishing and farming) in each of 56 areas.

```
library(SpatialEpi)
data(scotland)
Y <- scotland$data$cases
X <- scotland$data$AFF
E <- scotland$data$expected
# Relative risk estimates
smr <- Y/E
summary(E)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.100   4.050   6.300   9.575  10.125  88.700
summary(smr)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.496   1.111   1.522   2.241   6.429
```

Distribution of SMRs



Scottish lip cancer data

The SMRs have a large spread with an increasing trend in the south-north direction.

```
scotland.map <- scotland$spatial.polygon
par(mar = c(1, 1, 1, 1))
mapvariable(smr, scotland.map)  # Function in SpatialEpi
```

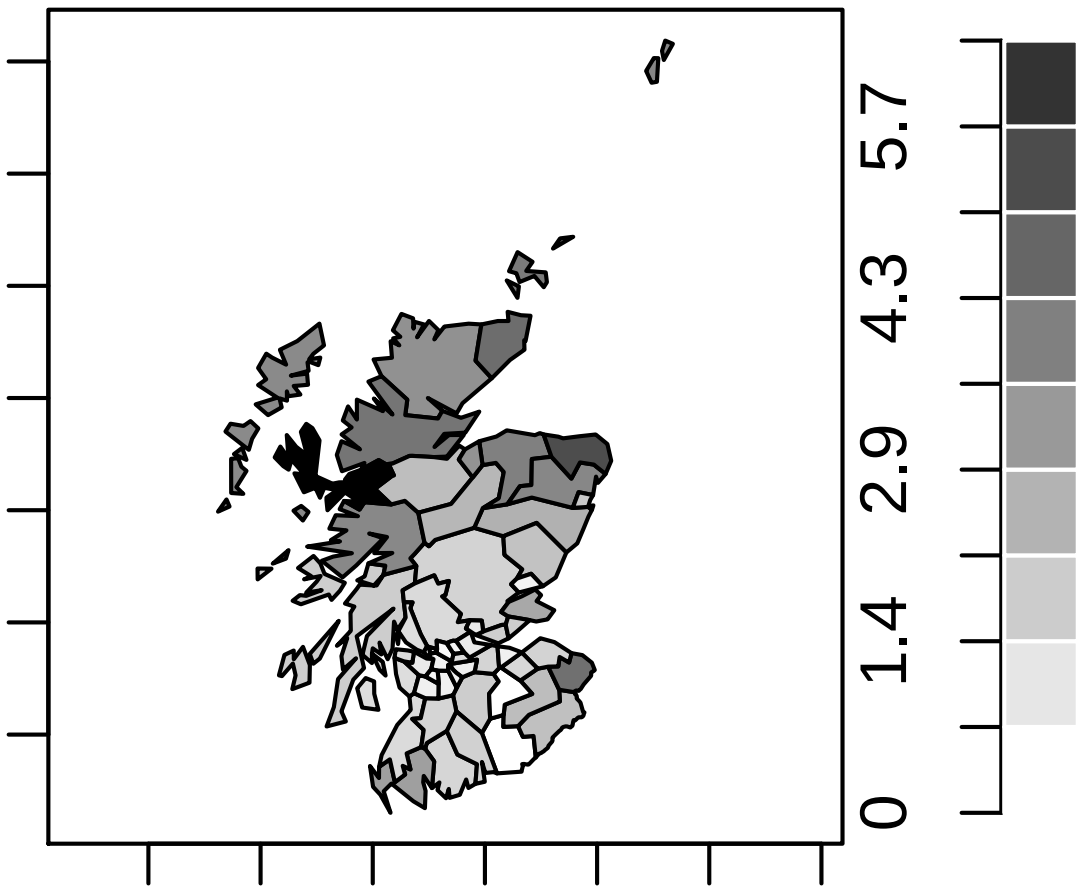


Figure 1: SMRs for Scottish lip cancer data

Scottish lip cancer data

The variance of the estimate in area i is

$$\text{var}(\text{SMR}_i) = \frac{\text{SMR}_i}{E_i},$$

which will be large if E_i is small.

For the Scottish data the expected numbers are highly variable, with range 1.1–88.7.

This variability suggests that there is a good chance that the extreme SMRs are based on small expected numbers (many of the large, sparsely-populated rural areas in the north have high SMRs).

Expected numbers for Scottish lip cancer data

```
par(mar = c(1, 1, 1, 1))
mapvariable(E, scotland.map)
```

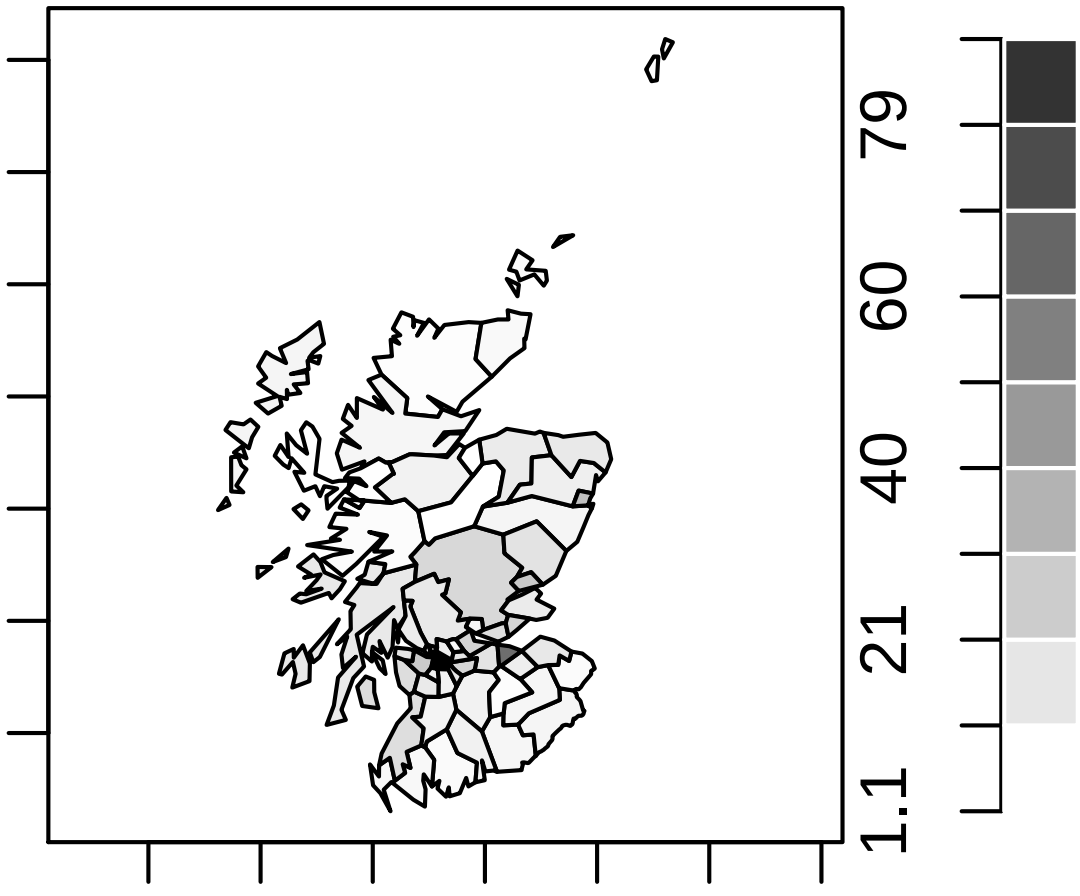
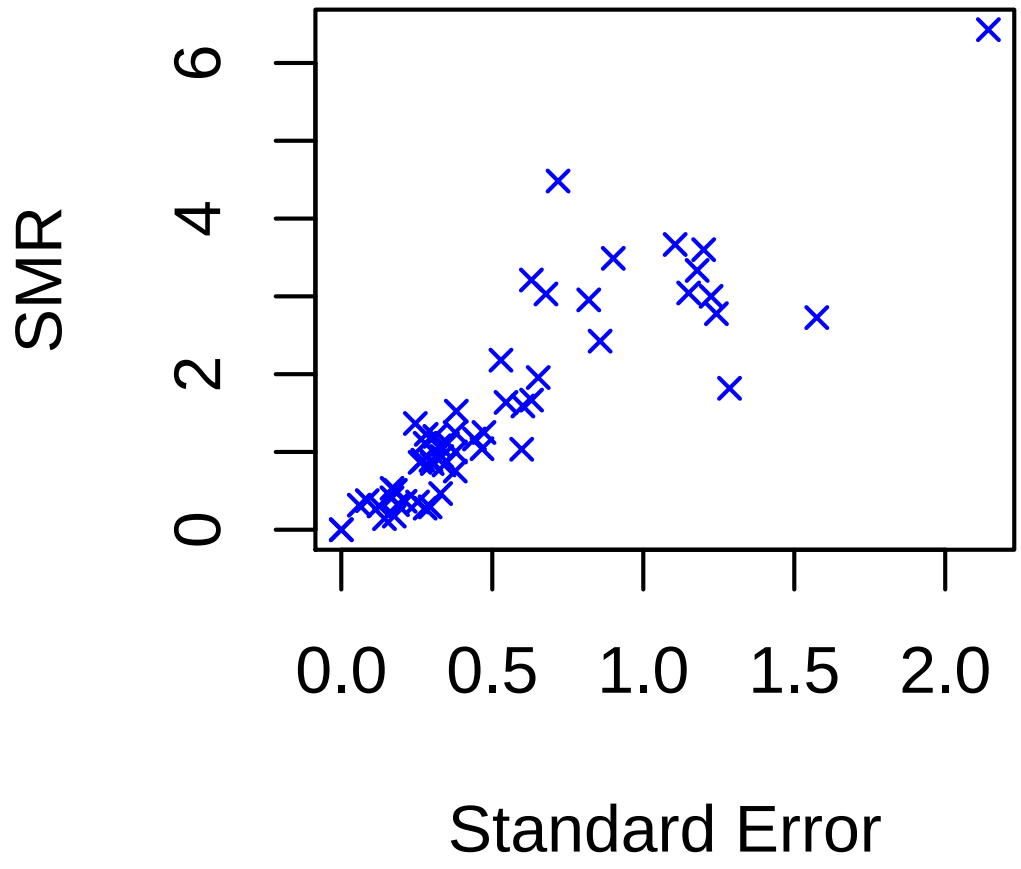


Figure 2: Expected numbers for Scottish lip cancer data

SMR for Scottish lip cancer data

The highest SMRs tend to have the largest standard errors.

```
se <- sqrt(smr/E)
plot(smr ~ se, xlab = "Standard Error", ylab = "SMR",
     col = "blue", cex = 0.7, pch = 4)
```



SMR interval estimates

Let $\theta_i = \text{SMR}_i$.

We obtain an interval estimate for $\alpha_i = \log \theta_i$ (since the normality of the estimator is likely to be better on this scale) and then transform.

Via the delta method

$$\widehat{\text{var}}(\widehat{\alpha}_i) = \widehat{\text{var}}(\widehat{\theta}_i)|J|^2$$

where $J = \frac{d\alpha_i}{d\theta_i} = \exp(-\alpha_i)$ and $\widehat{\text{var}}(\widehat{\theta}_i) = \widehat{\theta}_i / E_i$.

We obtain:

$$\widehat{\text{var}}(\widehat{\alpha}_i) = [E_i \exp(\widehat{\alpha}_i)]^{-1},$$

to give a 95% confidence interval for θ_i of

$$\exp\left(\widehat{\alpha}_i \pm 1.96 \times \sqrt{\widehat{\text{var}}(\widehat{\alpha}_i)}\right).$$

SMR estimates when $Y_i = 0$

When $Y_i = 0$, we obtain an SMR of 0, and (more worryingly) a standard error of zero.

In this case, we carry out an adjustment and set $Y_i^\star = Y_i + 0.5$ and $E_i^\star = E_i + 0.5$ to give the estimator

$$\theta_i^\star = \text{SMR}_i^\star = Y_i^\star / E_i^\star,$$

with $\hat{\text{var}}(\widehat{\theta}_i^\star) = \widehat{\theta}_i^\star / E_i^\star$.

Also let $\alpha_i^\star = \log \theta_i^\star$.

We obtain:

$$\widehat{\text{var}}(\widehat{\alpha}_i^\star) = (E_i^\star \exp(\widehat{\alpha}_i^\star))^{-1},$$

to give a 95% confidence interval of

$$\exp\left(\widehat{\alpha}_i^\star \pm 1.96 \times \sqrt{\widehat{\text{var}}(\widehat{\alpha}_i^\star)}\right).$$

SMR interval estimates

The addition of 0.5 is somewhat ad hoc but corresponds to a Ga(0.5,0.5) prior on the relative risk. This prior has 0.025, 0.5, 0.975 quantiles of 0.00098, 0.45, 5.0.

The addition of a non-integer also highlights that some adjustment has been made!

This prior is contributing information equivalent to observing an expected number of 0.5 and 'half a case'.

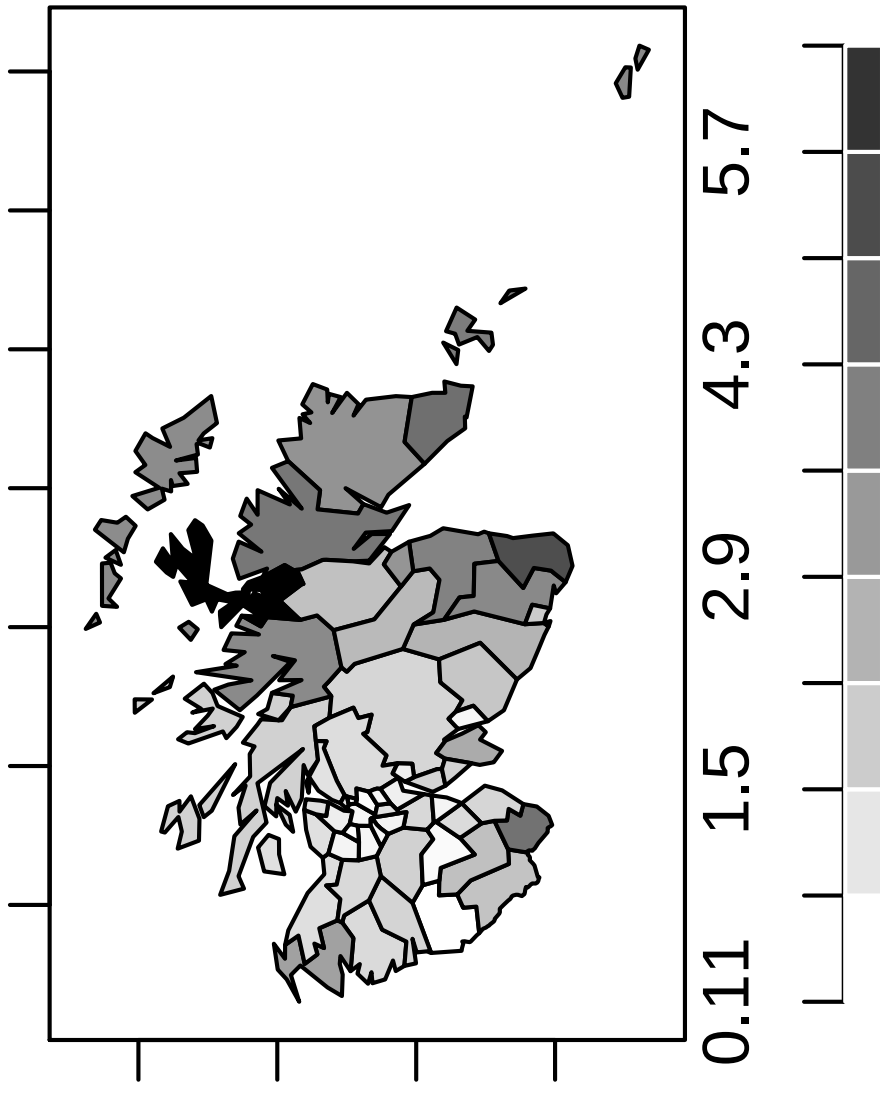
SMR estimates with adjustment

Create estimates adjusted for zeroes.

```
Ystar <- ifelse(Y == 0, 0.5, Y)
Estar <- ifelse(Y == 0, E + 0.5, E)
SMRstar <- Ystar/Estar
alphastar <- log(SMRstar)
varalphastar <- 1/(SMRstar * Estar)
SMRlower <- exp(alphastar - 1.96 * sqrt(varalphastar))
SMRupper <- exp(alphastar + 1.96 * sqrt(varalphastar))
SMRwidth <- SMRupper - SMRlower
```

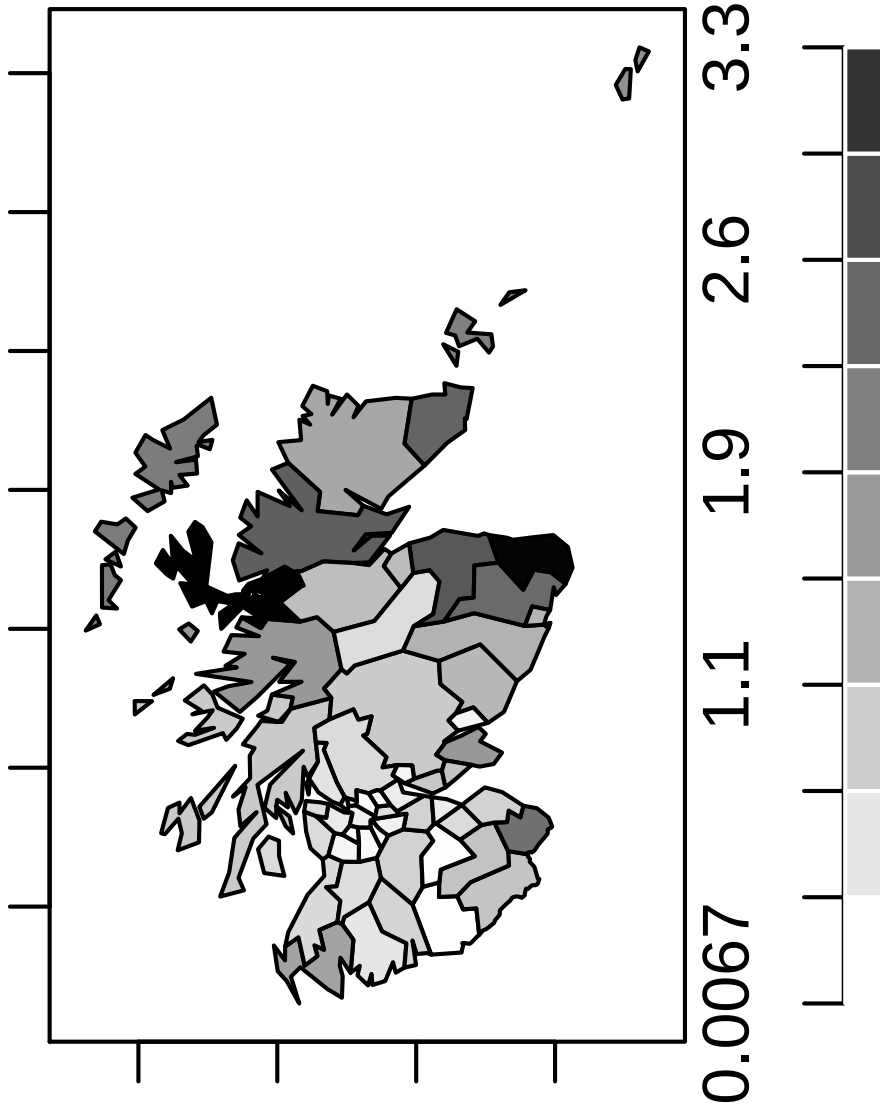
Point estimates with adjustment

```
par(mar = c(1, 1, 1, 1))
mapvariable(SMRstar, scotland.map)
```



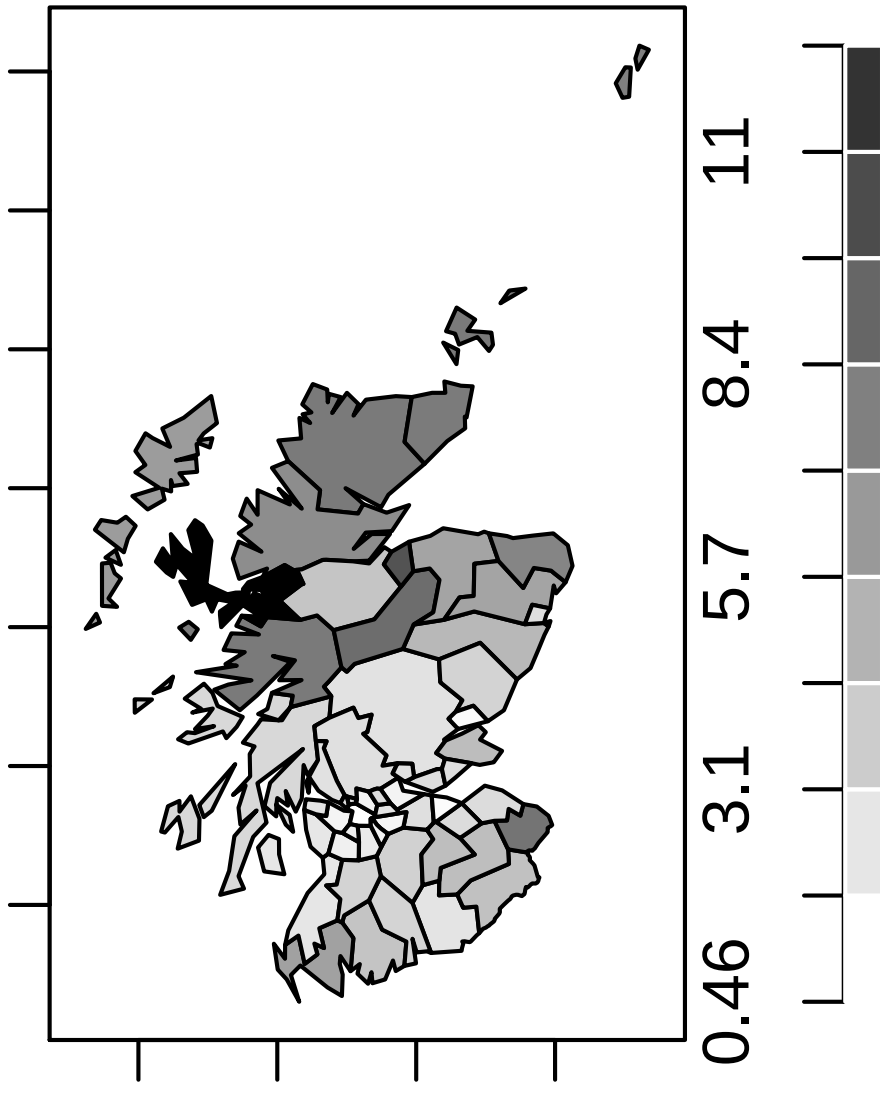
Lower confidence intervals with adjustment

```
par(mar = c(1, 1, 1, 1))
mapvariable(SMRlower, scotland.map)
```



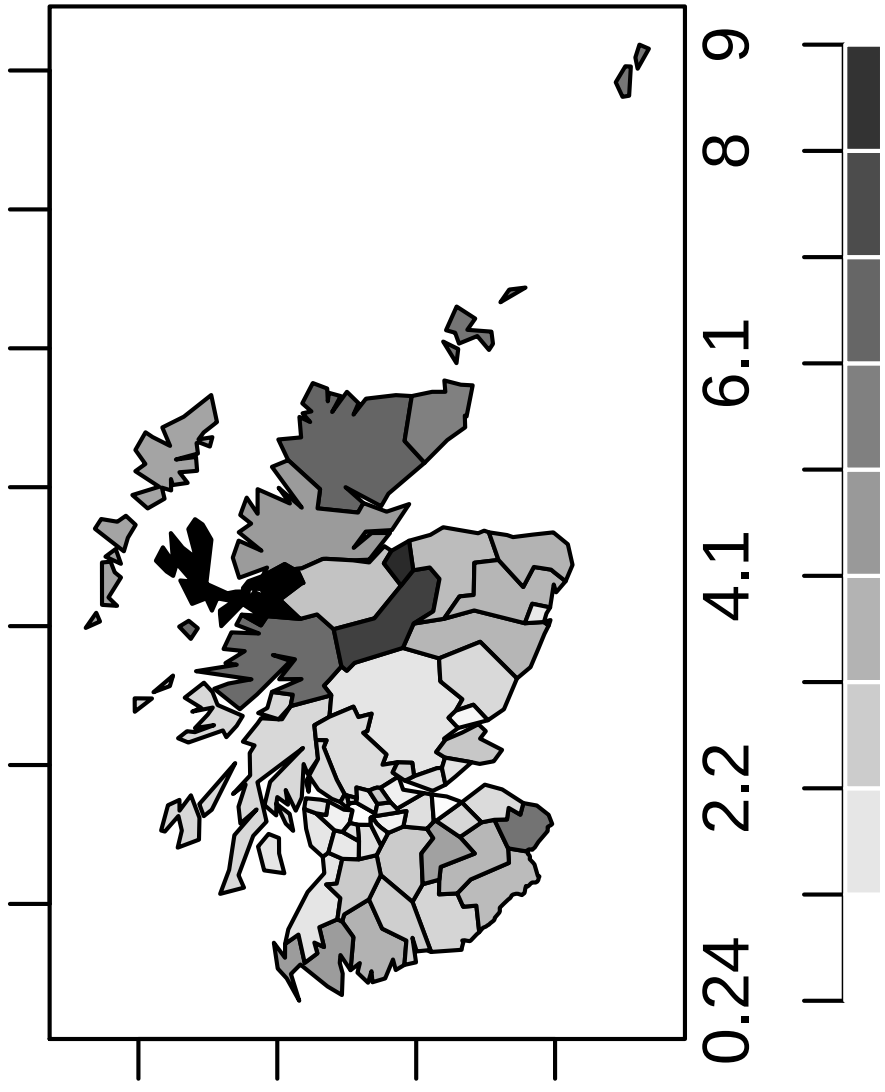
Upper confidence intervals with adjustment

```
par(mar = c(1, 1, 1, 1))
mapvariable(SMRupper, scotland.map)
```



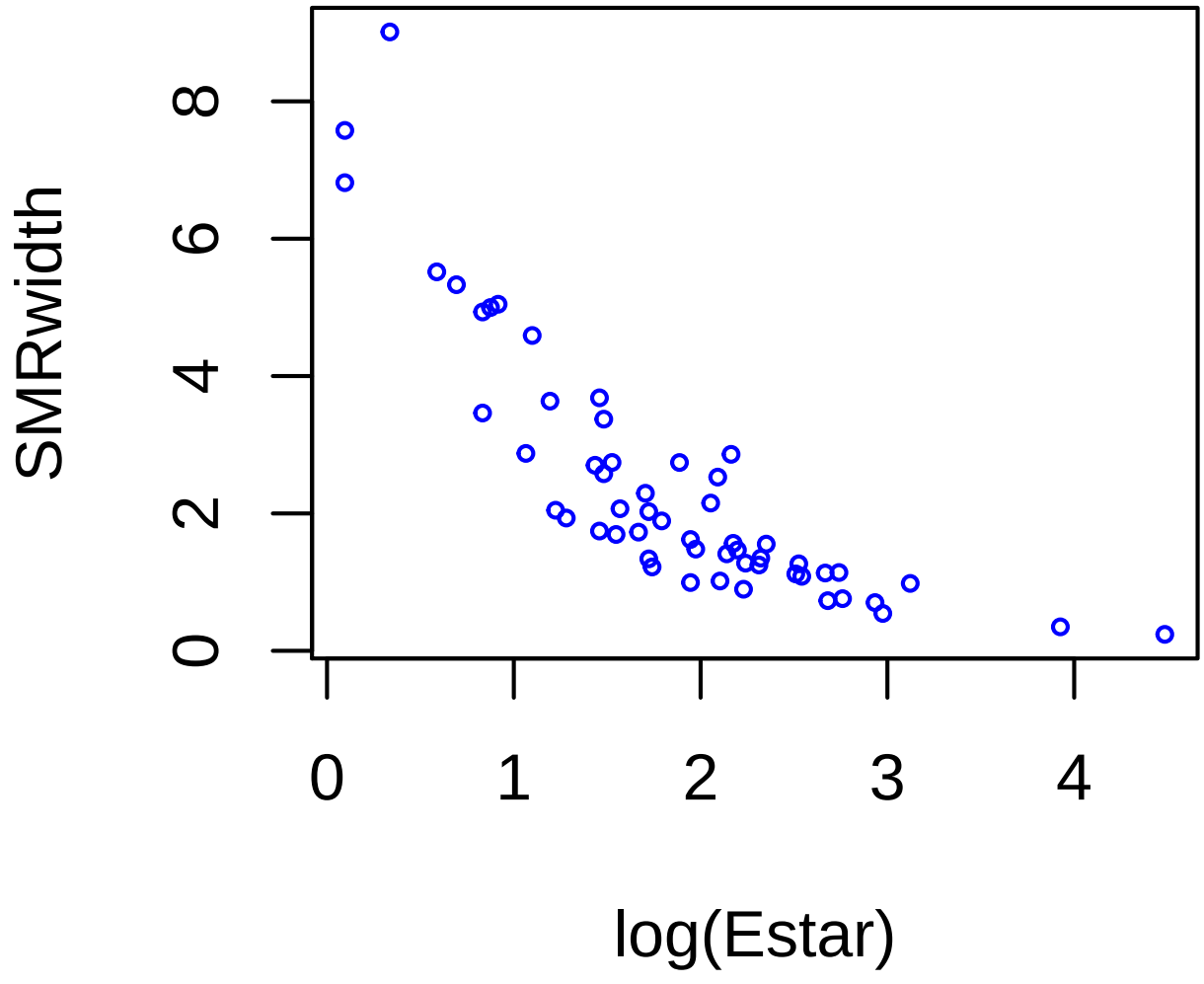
Interval widths with adjustment: note the variability!

```
par(mar = c(1, 1, 1, 1))
mapvariable(SMRwidth, scotland.map)
```



The width decreases with increasing expected numbers!

```
plot(SMRwidth ~ log(Estar), col = "blue", cex = 0.5)
```



Gamma Smoothing Model

Gamma smoothing model (empirical Bayes)

In the gamma model it is assumed that

$$Y_i|\beta_0, \delta_i \sim_{iid} \text{Poisson}(E_i \exp(\beta_0)\delta_i)$$

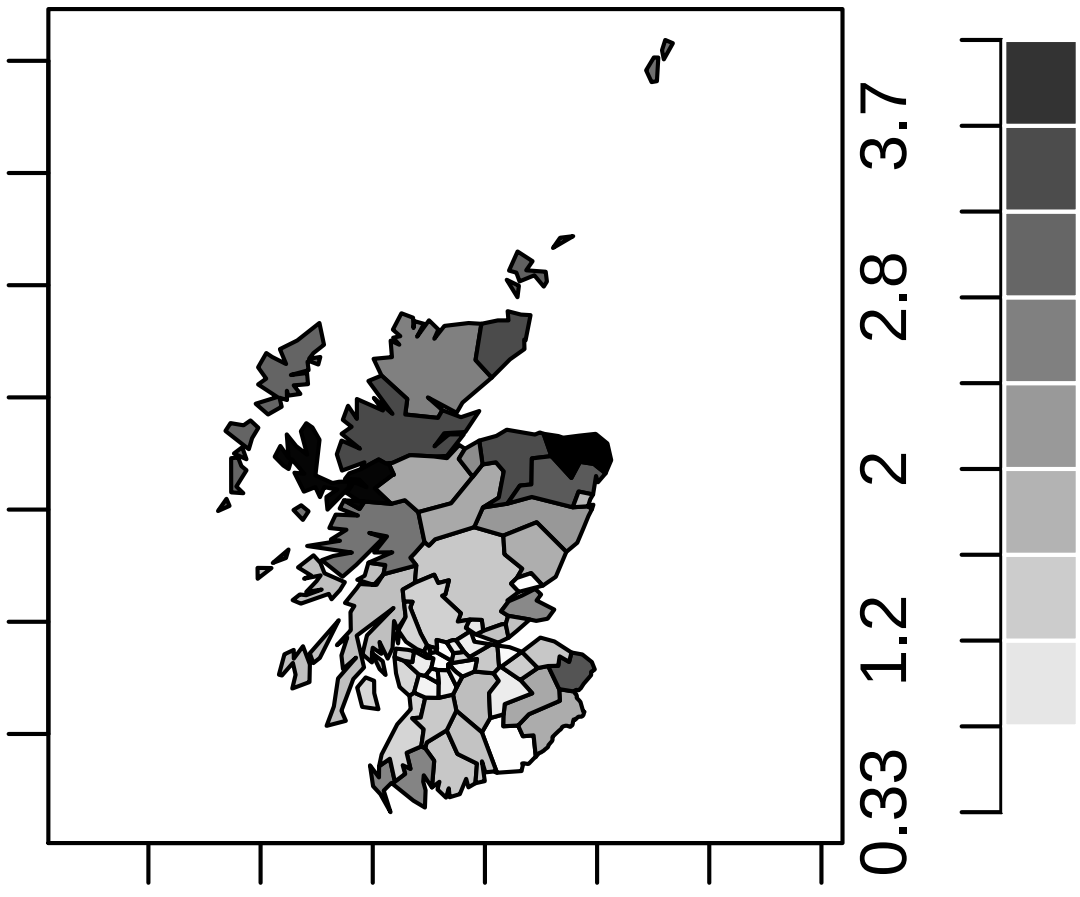
with the relative risks

$$\delta_i \sim_{iid} \text{Ga}(\alpha, \alpha).$$

```
# Gamma random effects model
ebresults <- eBayes(Y, E)
ebresults$alpha  #  the estimate of alpha in our model
## [1] 1.87949
exp(ebresults$beta)  # the estimate of exp(beta0) in our model
## (Intercept)
##     1.42206
# the RRs are greater in the study region than in
# the reference region
mean(Y/E)
## [1] 1.522366
summary(ebresults$RR)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3319  0.6079  1.1534  1.4221  2.0286  4.0791
```

Relative risk estimates (posterior means)

```
par(mar = c(1, 1, 1, 1))
mapvariable(ebresults$RR, scotland.map)
```



Assessment of gamma assumption

Below we assess the gamma assumption.

The ordered observed residual relative risks $\frac{Y_i/E_i}{e^{\beta_0}}$ are plotted against those expected from the gamma distribution. Let f_i be the ordered expected relative risk if n samples were taken from a gamma distribution.

If the gamma distribution is appropriate the points should lie approximately on a straight line.

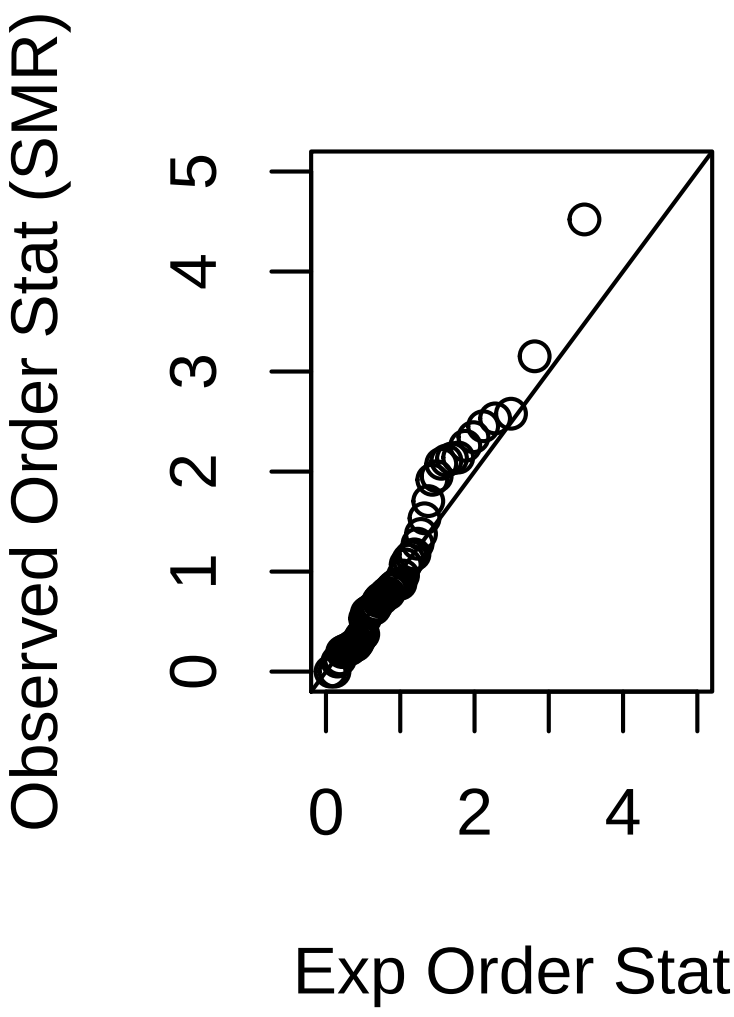
We can also plot the estimated relative risks $\widehat{\text{RR}}_i$ under the model against f_i.

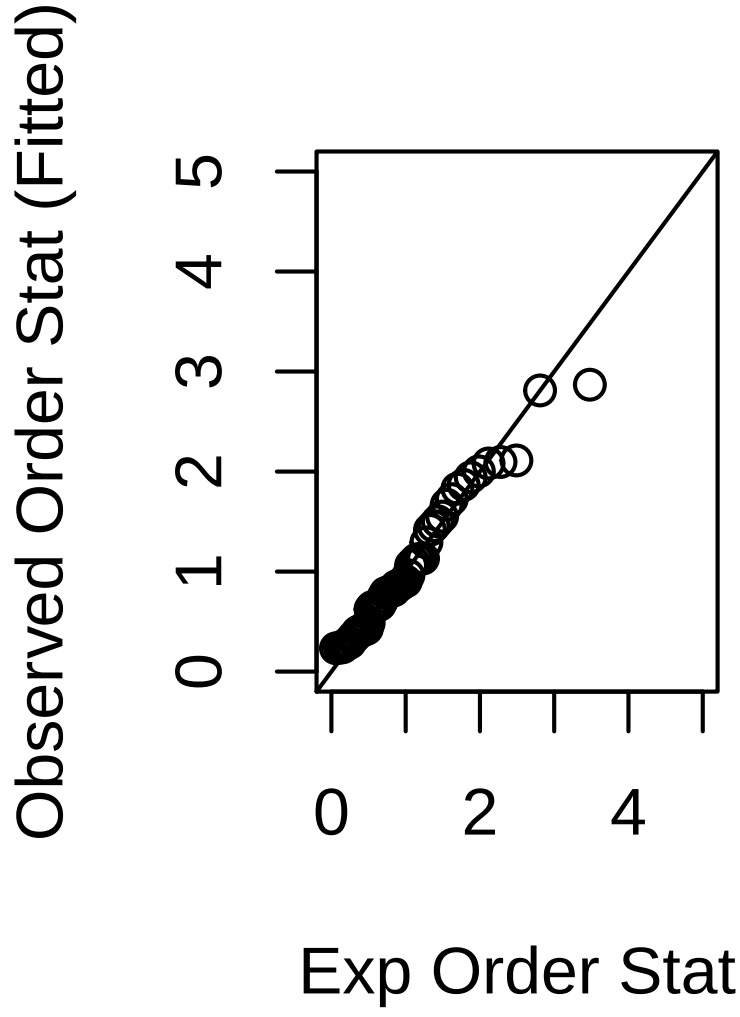
Note that the estimates have been constructed under the assumption of gamma random effects, which can make the estimates look more like a gamma sample than they really are.

Assessment of gamma assumption

```
# Now let's examine the gamma assumption
egamma <- qgamma(seq(0.5, length(Y), 1)/length(Y),
    ebresults$alpha, ebresults$alpha)
par(mfrow = c(1, 2))
# First plot is the SMR estimates
plot(egamma, exp(-ebresults$beta) * sort(smr), xlim = c(0,
    5), ylim = c(0, 5), xlab = "Exp Order Stat", ylab = "Obs Ord
abline(0, 1)
# Second plot is the estimates from the gamma model
plot(egamma, exp(-ebresults$beta) * sort(ebresults$RR),
    xlim = c(0, 5), ylim = c(0, 5), xlab = "Exp Order Stat",
    ylab = "Obs Order Stat (Gamma)")
abline(0, 1)
```

Assessment of gamma assumption: looks reasonable





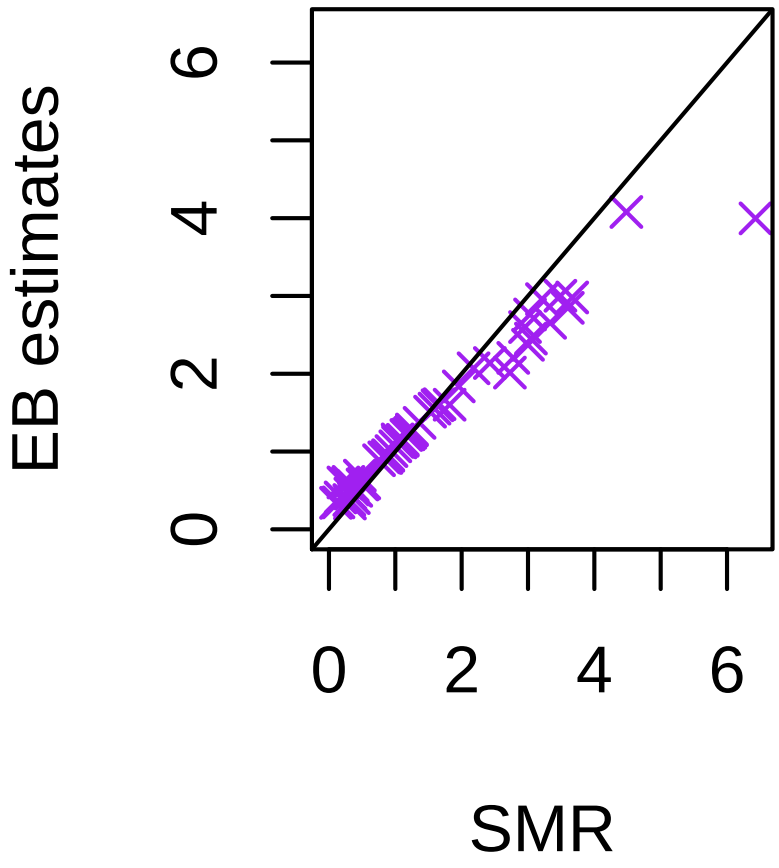
Gamma smoothing model

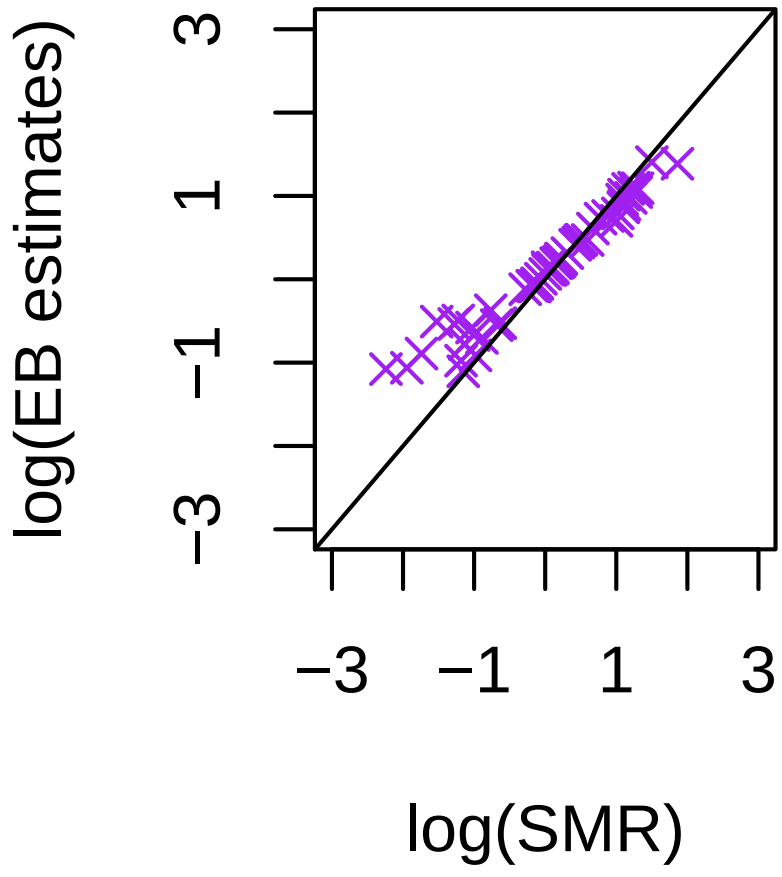
Below we give code to compare the SMRs and empirical Bayes gamma smoothed estimates.

We plot on the original and on the log scale

```
# Comparison of point estimates
par(mfrow = c(1, 2))
plot(SMRstar, ebresults$RR, xlab = "SMR", ylab = "EB estimates",
    xlim = c(0, max(SMRstar)), ylim = c(0, max(SMRstar)),
    pch = 4, col = "purple")
abline(0, 1)
plot(log(SMRstar), log(ebresults$RR), xlim = c(-3,
    3), ylim = c(-3, 3), xlab = "log(SMR)", ylab = "log(EB estim
    pch = 4, col = "purple")
abline(0, 1)
```

Gamma smoothed estimates are shrunk





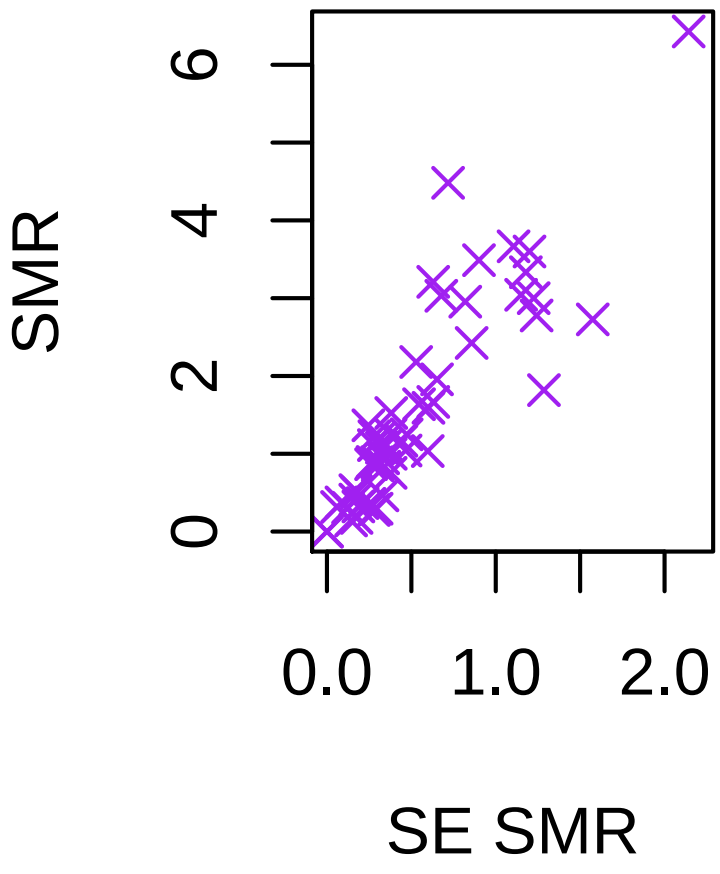
Gamma smoothing model

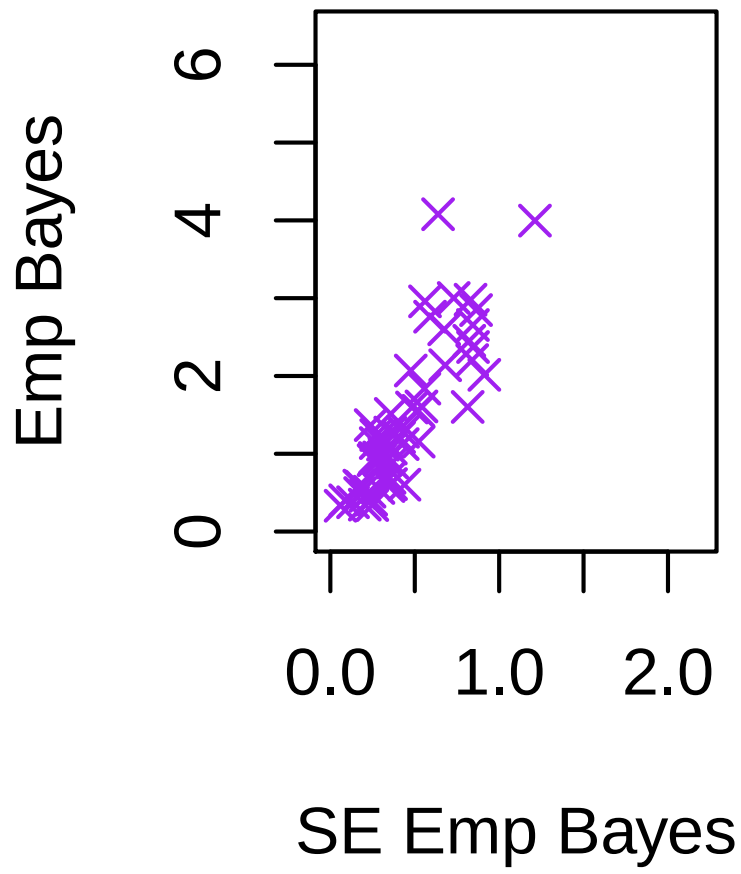
We repeat the earlier plot of SMRs versus standard errors of SMRs, and also empirical smoothed Bayes estimates versus their standard error.

Again the shrinkage of the Bayes estimates is apparent, and the standard errors are smoothed also (note no zeroes on the right).

```
seEBests <- sqrt((ebresults$alpha + Y) * exp(2 * ebresults$beta)
    E * exp(ebresults$beta))^2)
par(mfrow = c(1, 2))
plot(se, smr, ylim = c(0, max(smr)), xlim = c(0, 2.2),
    xlab = "SE SMR", ylab = "SMR", pch = 4, col = "purple")
plot(seEBests, ebresults$RR, ylim = c(0, max(smr)),
    xlim = c(0, 2.2), xlab = "SE Emp Bayes", ylab = "Emp Bayes",
    pch = 4, col = "purple")
```

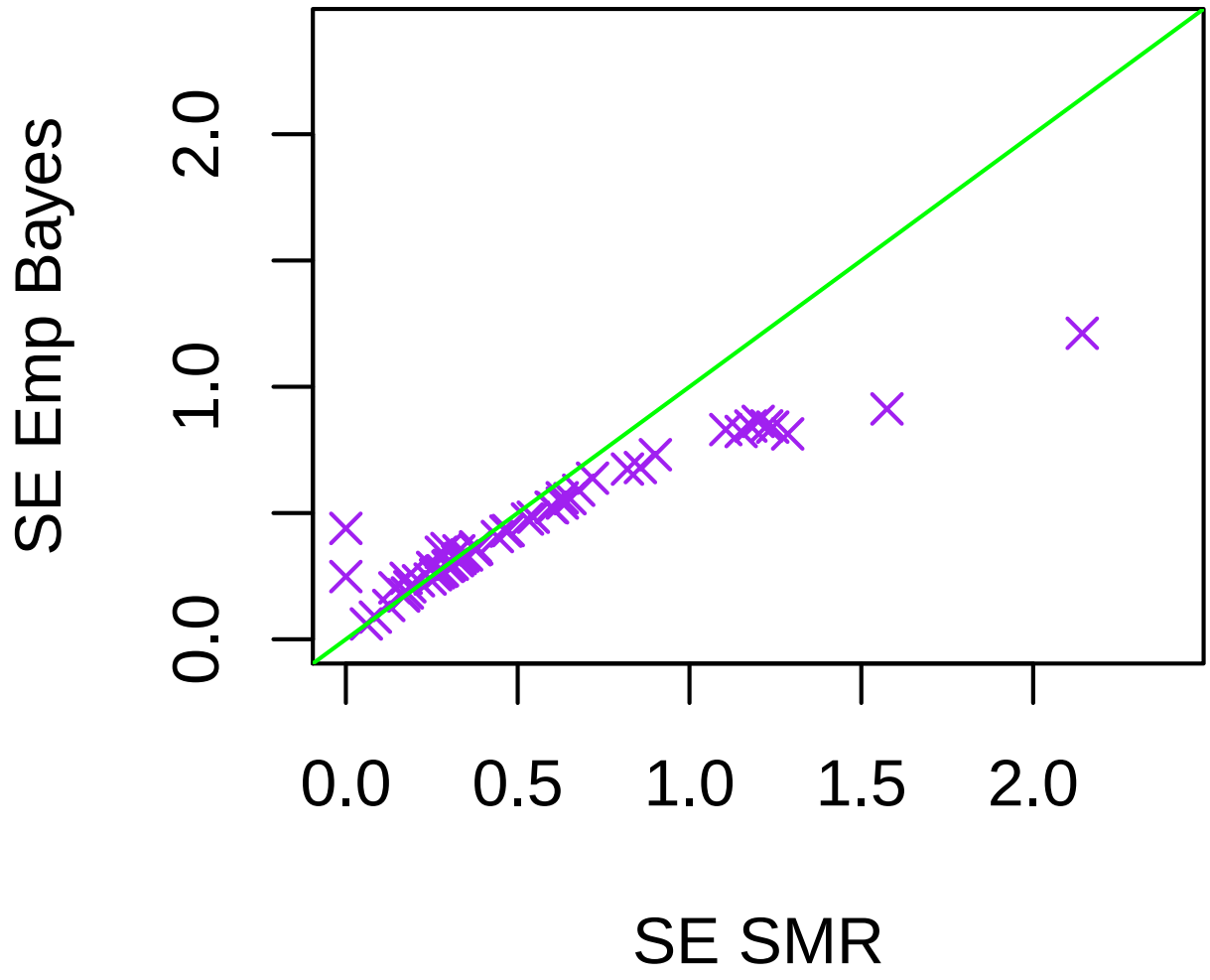
Gamma smoothing model





Gamma smoothing model: comparison of standard errors

```
plot(seEBests ~ se, ylab = "SE Emp Bayes", xlab = "SE SMR",
    xlim = c(0, 2.4), ylim = c(0, 2.4), pch = 4, col = "purple")
abline(0, 1, col = "green")
```



Gamma smoothing model

Given that the posterior distributions of the relative risks have a known distribution (which is gamma, with known β_0 and α), we can examine the complete distribution.

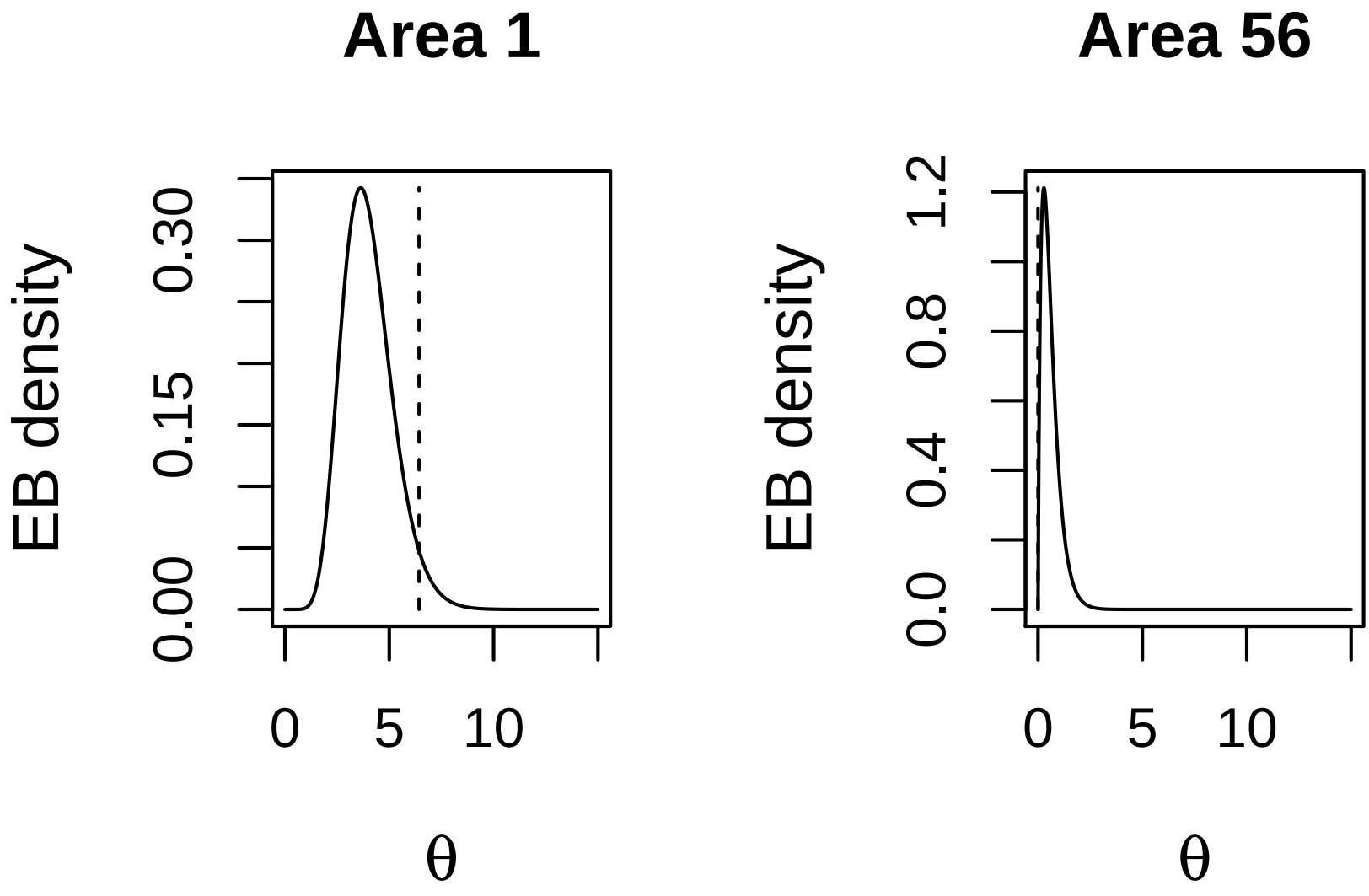
Previously we looked at the posterior means, below we look at the distributions for areas 1 and 56.

The function `EBpostdens` is in the `SpatialEpi` package.

```
# Densities for areas 1 and 56
par(mfrow = c(1, 2))
EBpostdens(Y[1], E[1], ebresults$alpha, ebresults$beta,
    lower = 0, upper = 15, main = "Area 1")
EBpostdens(Y[56], E[56], ebresults$alpha, ebresults$beta,
    lower = 0, upper = 15, main = "Area 56")
```

The SMRs are indicated as dashed line - note the zero on the right.

Posterior distributions of relative risks



Posterior probabilities of exceeeding a threshold

We calculate the posterior probability that the relative risk exceeds a value of 3 (this value was chosen for illustration).

We then map these probabilities.

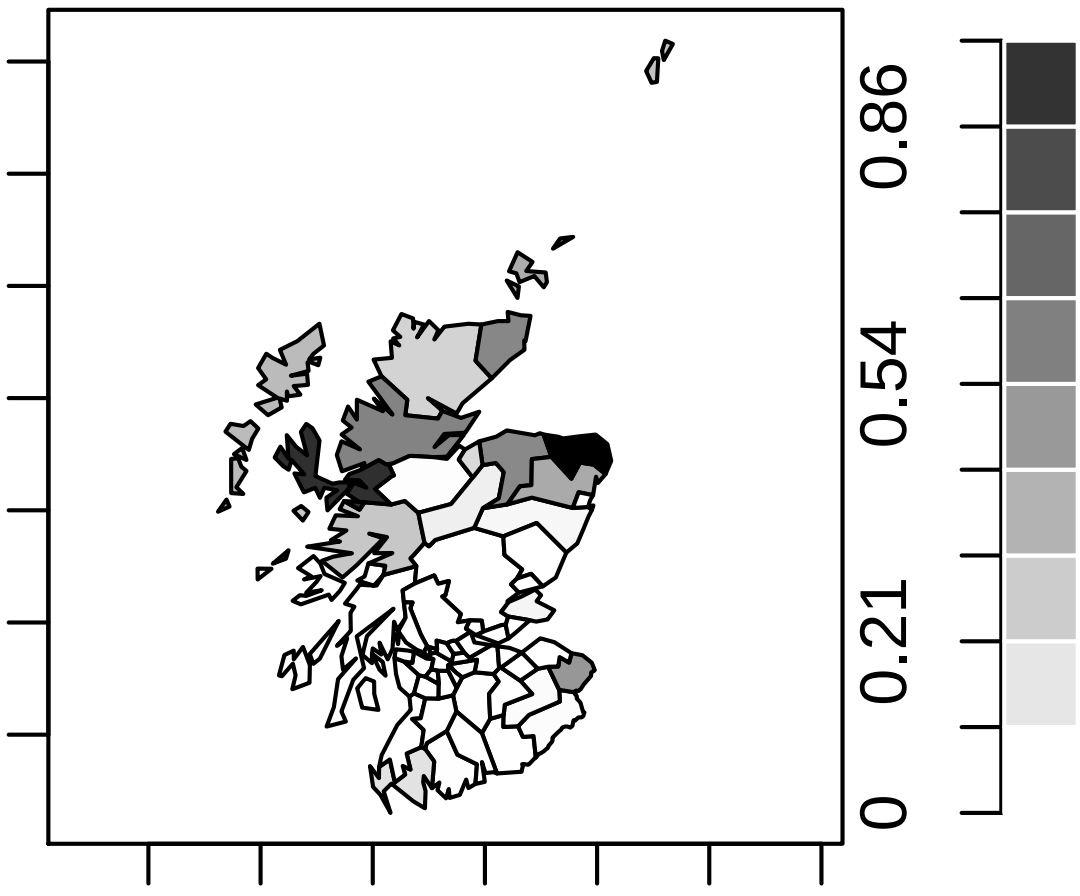
The function `EBpostthresh` is in the `SpatialEpi` package.

In general, we are attempting to understand the variability in posterior distributions across a map, and a single map can only represent one aspect. Hence, presenting multiple maps is a good idea, for example, maps of locations and spreads.

```
par(mfrow = c(1, 1), mar = c(1, 1, 1, 1))
thresh3 <- EBpostthresh(Y, E, alpha = ebresults$alpha,
    beta = ebresults$beta, rrthresh = 3)
mapvariable(thresh3, scotland.map)
```

Posterior probabilities of exceeeding a threshold

Two areas in particular have high values.



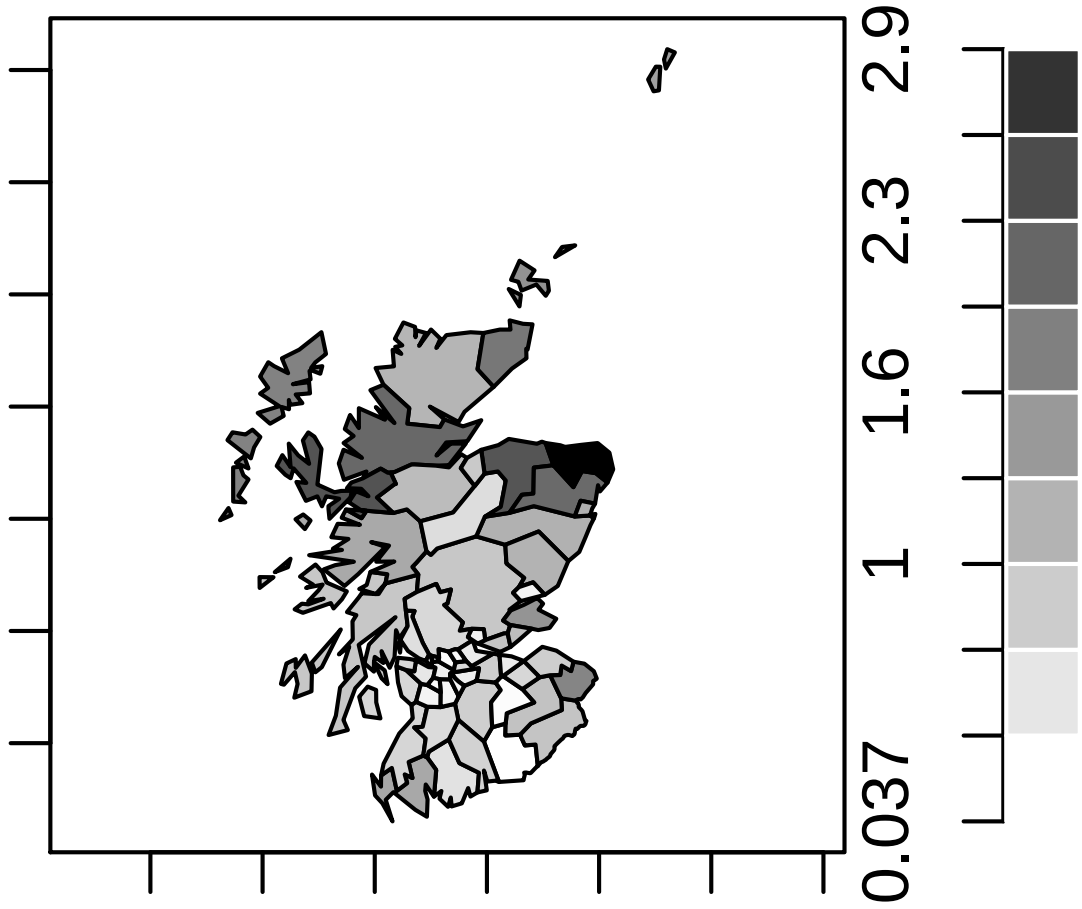
Uncertainty estimates of EB estimates

We calculate the 2.5% and 97.5% points of the gamma posteriors of the relative risks.

```
apost <- ebresults$alpha + Y
bpost <- (ebresults$alpha + E * exp(ebresults$beta))/exp(ebresul
EBlower <- qgamma(0.025, apost, bpost)
EBupper <- qgamma(0.975, apost, bpost)
EBwidth <- EBupper - EBlower
```

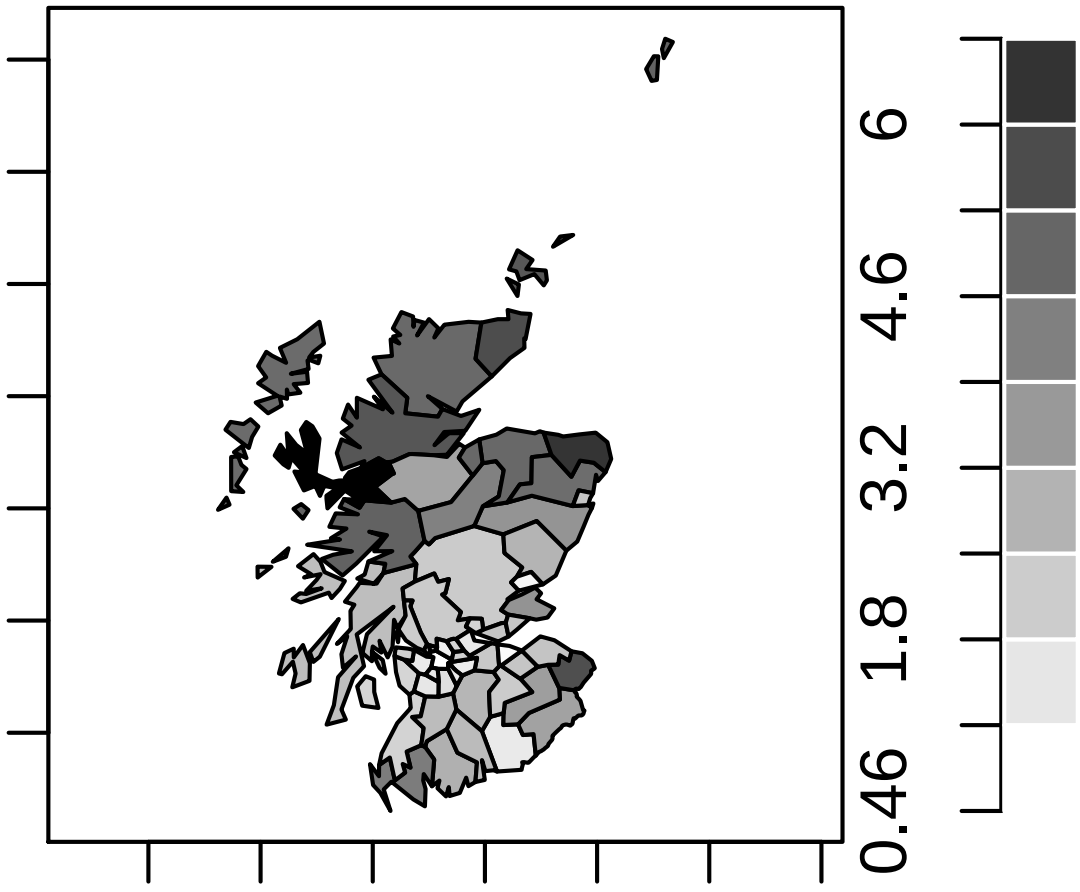
Lower 2.5% point of EB estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(EBlower, scotland.map)
```



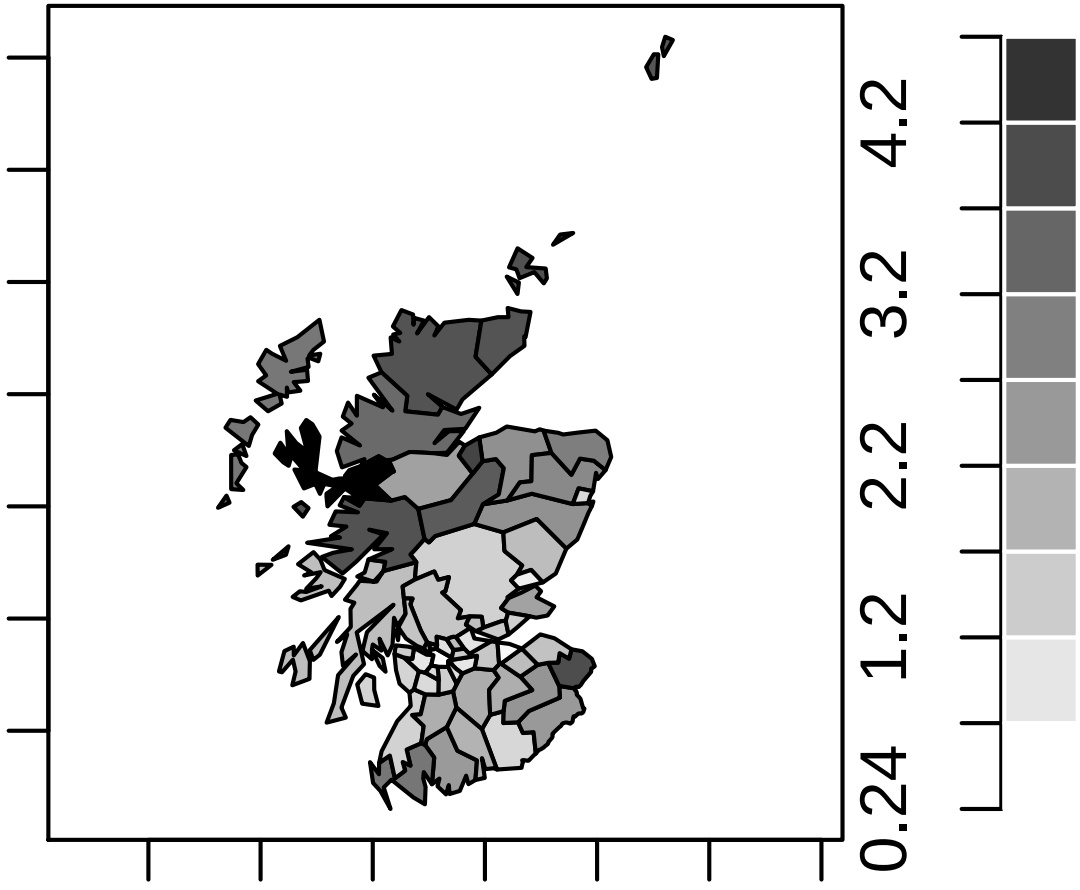
Lower 97.5% point of EB estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(EBupper, scotland.map)
```



Width of EB estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(EBwidth, scotland.map)
```

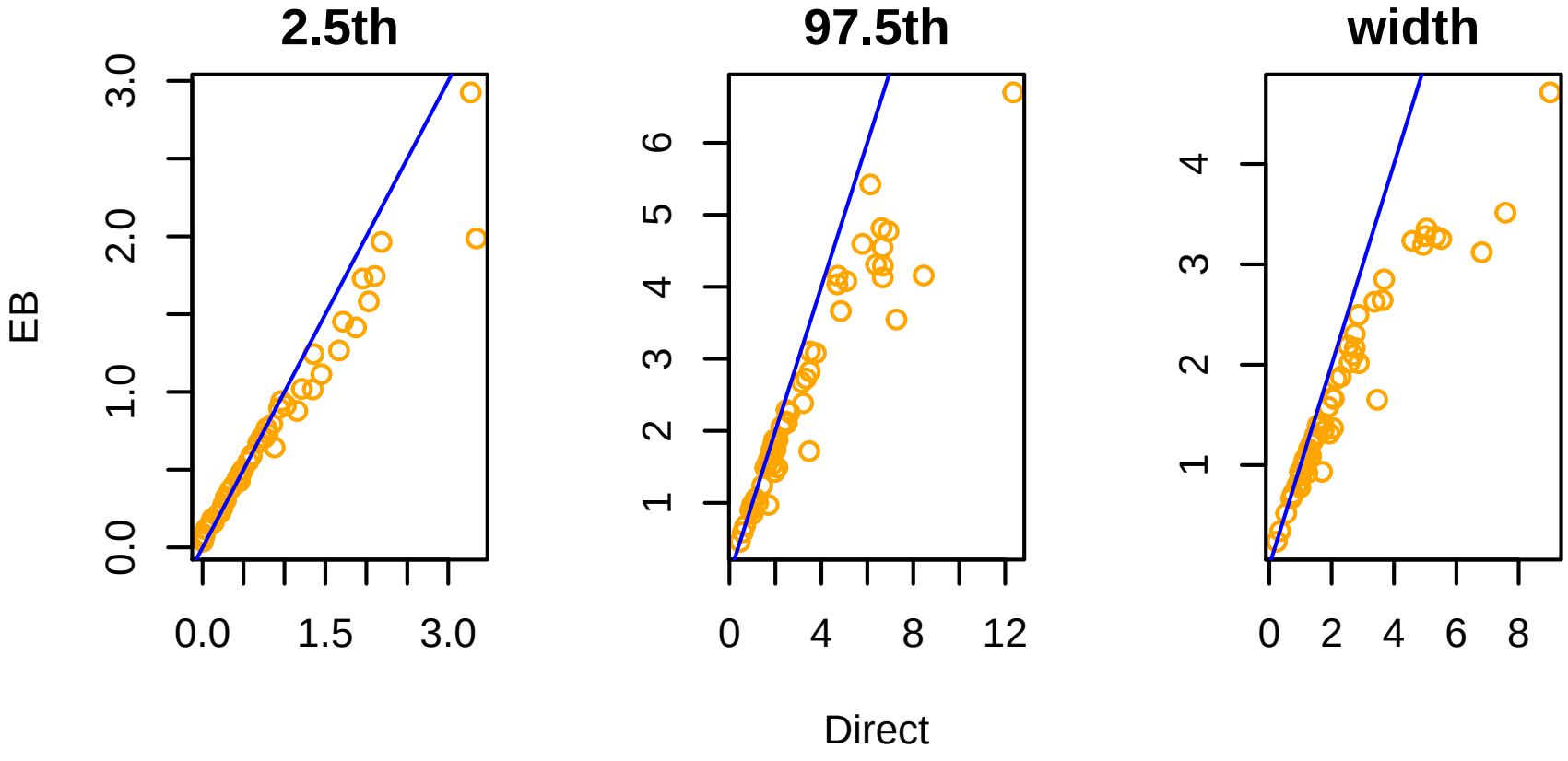


Comparison of interval widths: EB is in general narrower

```
par(mfrow = c(1, 3), mar = c(5, 4, 2, 1))
plot(EBlower ~ SMRlower, col = "orange", ylab = "EB",
    xlab = "", main = "2.5th")
abline(0, 1, col = "blue")
plot(EBupper ~ SMRupper, col = "orange", ylab = "",
    xlab = "Direct", main = "97.5th")
abline(0, 1, col = "blue")
plot(EBwidth ~ SMRwidth, col = "orange", ylab = "",
    xlab = "", main = "width")
abline(0, 1, col = "blue")
```

Comparison of interval widths

We see the effect of the smoothing in terms of narrowing the interval widths as compared to the analysis based on the SMRs (which only uses the data from each area); using the data from all areas increases the precision of the estimates.



Gamma smoothing model with one covariate

We fit the model with the covariate AFF:

$$Y_i|\theta_i \sim_{iid} \text{Poisson}(E_i\theta_i)$$

with

$$\theta_i = \exp(\beta_0 + \beta_1 x_i)\delta_i,$$

and $\delta_i \sim_{iid} \text{Ga}(\alpha, \alpha)$.

```
# Now with AFF
ebresultsX <- eBayes(Y, E, X)
ebresultsX$alpha
## [1] 2.98428
ebresults$alpha
## [1] 1.87949
# note the reduction in excess-Poisson variation
# compared to the no covariate model
```

Gamma smoothing model with covariate

```
ebresultsX$beta
## (Intercept)        Xmat
##  -0.3527686   7.1481551
exp(0.1 * ebresultsX$beta[2])
##    Xmat
## 2.04381
```

We estimate that a 10% increase in the percentage in AFF is associated with an 2.0438096 increase in risk of male lip cancer, which is large.

There may be a small part of exposure to sunlight, but more likely AFF is acting as a surrogate for smoking, alcohol, diet,...

A new function to allow construction of interval estimates

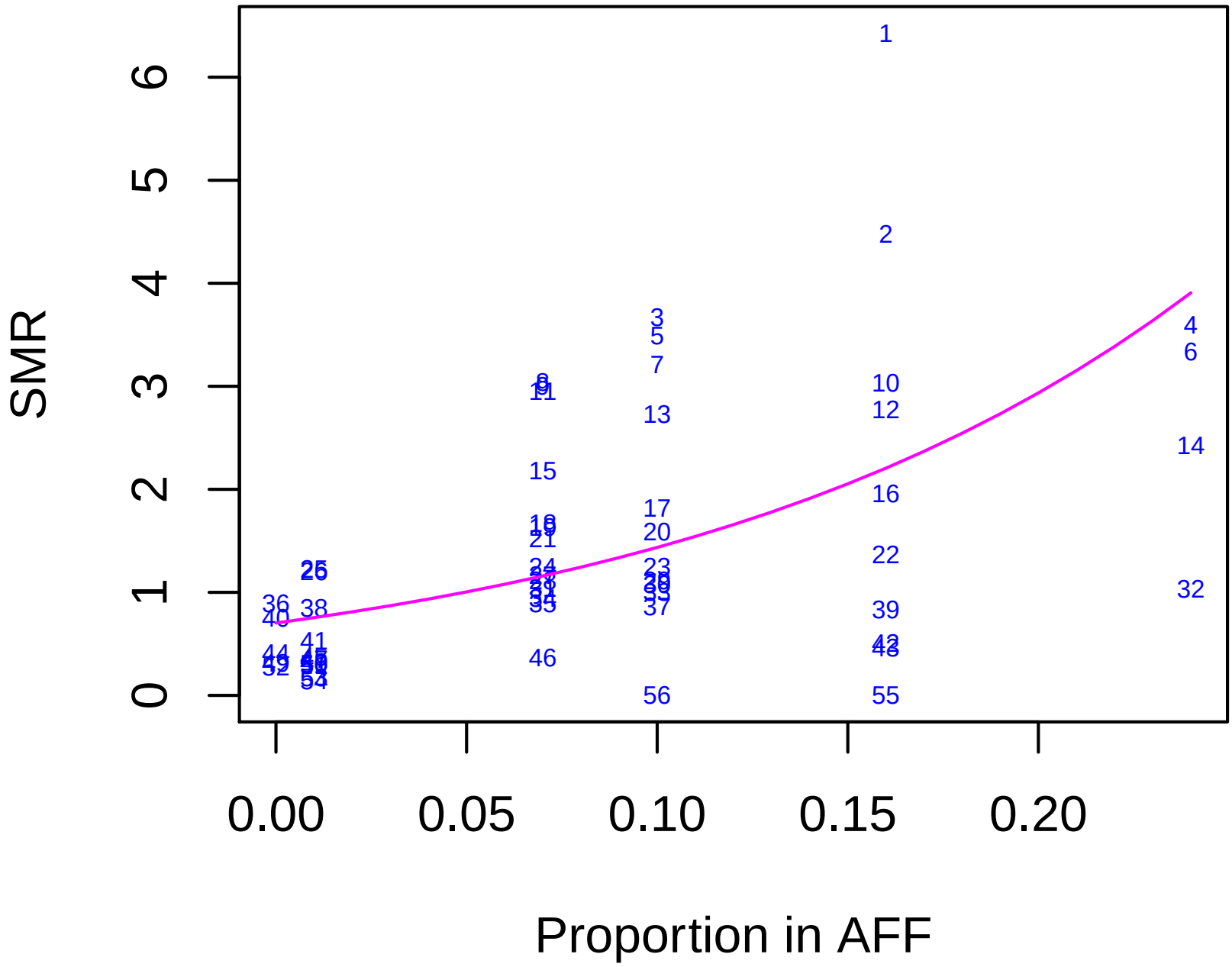
```
library(MASS)
eBayes2 <- function(Y, E, Xmat = NULL) {
    if (is.null(Xmat)) {
        mod <- glm.nb(Y ~ 1 + offset(log(E)), link = log)
    } else {
        mod <- glm.nb(Y ~ Xmat + offset(log(E)), link = log)
    }
    alpha <- mod$theta
    muhat <- mod$fitted/E
    wgt <- E * muhat/(alpha + E * muhat)
    SMR <- Y/E
    RR <- as.numeric(wgt * SMR + (1 - wgt) * muhat)
    RRmed <- qgamma(0.5, alpha + Y, (alpha + E * muhat)/muhat)
    list(RR = RR, RRmed = RRmed, beta = mod$coeff,
        mod = mod, alpha = alpha, SMR = SMR)
}
gammafit <- eBayes2(Y, E, X)
confint(gammafit$mod, level = 0.95)
##                  2.5 %      97.5 %
## (Intercept) -0.6632437 -0.03391441
## Xmat         4.2782203 10.08359144
```

Gamma smoothing model with covariate

We fit the model with the covariate AFF below and plot the fitted line versus the data.

```
# Now with AFF
ebresultsX <- eBayes(Y, E, X)
par(mfrow = c(1, 1))
plot(X, smr, type = "n")
text(X, smr)
xval <- seq(0, max(X), 0.01)
lines(xval, exp(ebresultsX$beta[1] + ebresultsX$beta[2] *
    xval))
```

Gamma smoothing model with covariate



Gamma smoothing model

Recall the variance of the Ga(α, α) is $1/\alpha$.

In the above, the variance of the random effects is

- $1/1.88 = 0.53$ without AFF and
- $1/2.98 = 0.34$ with AFF,

showing the variability "explained" by the covariate.

Comparison of models

We compare the SMRs with the gamma estimates with and without covariates.

```
# Now let's compare estimates from the three
# models:
x0 <- rep(0, length(Y))
x1 <- rep(1, length(Y))
x2 <- rep(2, length(Y))
plot(x0, smr, xlim = c(0, 2), type = "n", ylab = "Relative risks
    xlab = "", axes = F)
axis(2)
axis(1, at = c(0, 1, 2))
text(x0, smr, cex = 0.5)
text(x1, ebresults$RR, cex = 0.5)
text(x2, ebresultsX$RR, cex = 0.5)
for (i in 1:length(Y)) {
    lines(c(0, 1, 2), c(smr[i], ebresults$RR[i], ebresultsX$RR[i
        lty = 2, col = "grey")
}
abline(1, 0, col = "red")
```

Comparison of models

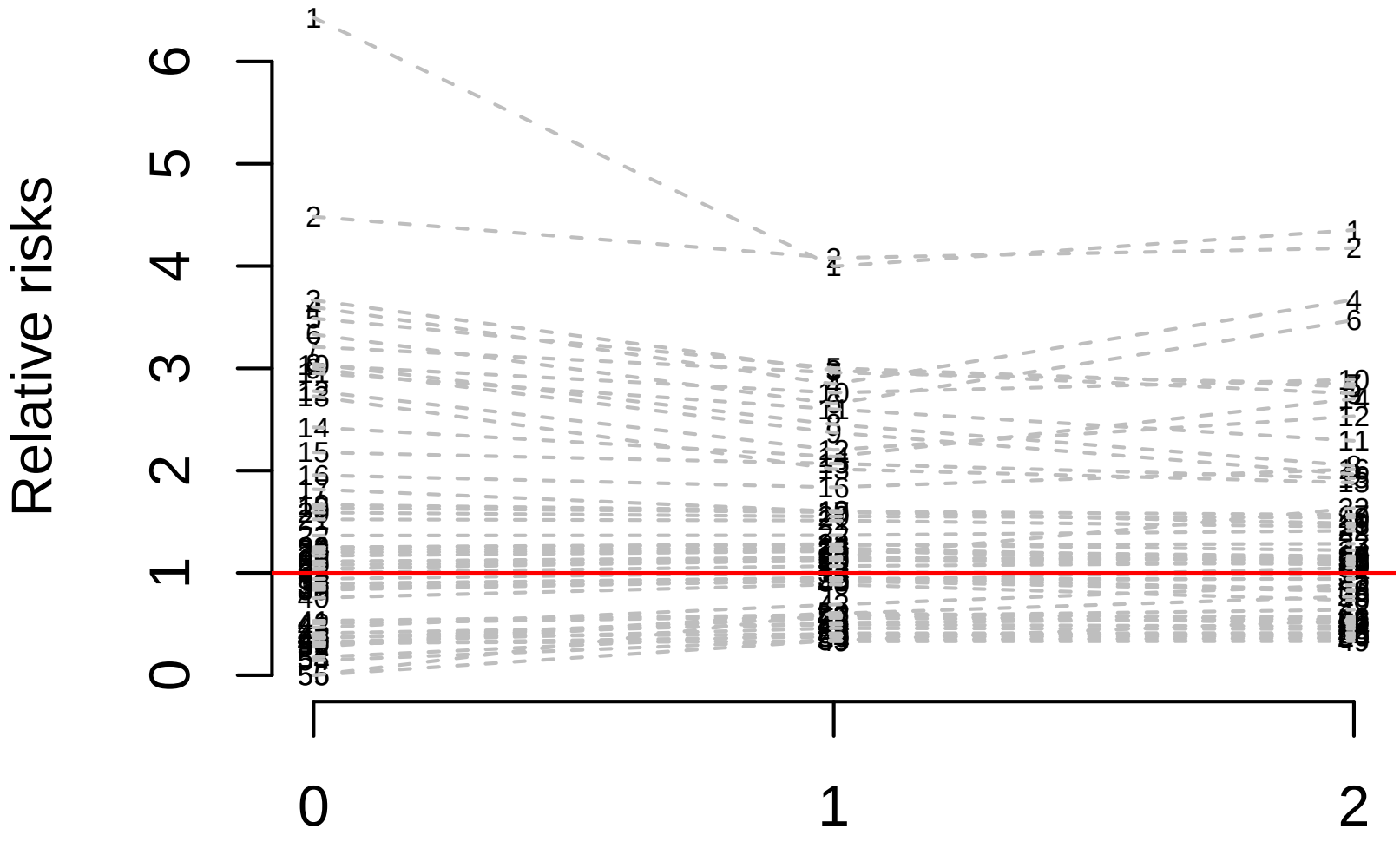


Figure 3: SMRs at 0, EB estimates without covariate at 1 and with EB estimates with covariate AFF at 2.

Comparison of models

We note that the random effects estimates (from the two gamma models) are relatively similar, except in areas 4 and 6 whose estimates are increased, based on the loglinear model.

The fit of this model was shown earlier and we might question the adequacy of this model for large AFF.

Lognormal Non-Spatial Smoothing Model

Integrated Nested Laplace Approximation (INLA)

- Relatively recently an approach has emerged that combines Laplace approximations and numerical integration in a very efficient manner, see Rue et al (2009) for more detail.
- The method is designed for "latent Gaussian models", which for our purposes, means modeling with latent normal random effects, that have independent, spatial, or space-time structure.
- Suppose the outcomes y_i, $i = 1, \ldots, n$ have density $p(y_i|x_i, \alpha^\star)$ where $x|\theta \sim \mathsf{N}(0, Q(\theta)^{-1})$ and $\alpha^\star$ are dispersion parameters (e.g.~the measurement error variance for a normal sampling model). So the x represent the random effects, along with regression coefficients.
- We also have priors for $\alpha^\star$ and θ — these priors may be non-normal.
- Let $\alpha = [\theta, \alpha^\star]$ denote the non-Gaussian parameters.

Integrated Nested Laplace Approximation (INLA)

- Then,

$$\begin{aligned}\pi(x, \alpha \mid y) &\propto \pi(\alpha)\pi(x \mid \alpha) \prod_i p(y_i \mid x_i, \alpha) \\ &\propto \pi(\alpha)|Q(\theta)|^{p/2} \exp\left\{-\frac{1}{2}x^T Q(\theta)x + \sum_i \log p(y_i|x_i, \alpha)\right\}\end{aligned}$$

The random effects are always added on the linear predictor scale:

- linear for a normal sampling model
- log-linear for a Poisson sampling model
- logistic for a sampling binomial model

The INLA Algorithm

- Suppose we wish to obtain the posterior marginals $\pi(x_i \mid y)$ and $\pi(\alpha \mid y)$. We have

$$\pi(x_i \mid y) = \int \pi(x_i \mid \alpha, y) \times \pi(\alpha \mid y)\, d\alpha$$

which is evaluated via the approximation

$$\begin{aligned} \widetilde{\pi}(x_i \mid y) &= \int \widetilde{\pi}(x_i \mid \alpha, y) \times \widetilde{\pi}(\alpha \mid y)\, d\alpha && (1) \\ &= \sum_k \widetilde{\pi}(x_i \mid \alpha_k, y) \times \widetilde{\pi}(\alpha_k \mid y) \times \Delta_k && (2) \end{aligned}$$

where

- $\widetilde{\pi}(x_i \mid \alpha, y)$ is approximated by Laplace (or other analytical approximations).
- For $\widetilde{\pi}(\alpha_k \mid y)$ the mode is located and then the Hessian is approximated, from which a grid of points is found that cover the density.

Pros and Cons of INLA

Advantages:

- Quite widely applicable: Generalized Linear Mixed Models (GLMMs) including temporal and spatial error terms.
- Very fast.
- An R package is available (though non-standard so can't load in the usual way).

Disadvantages:

- Restricted to models with Gaussian random effects.
- Experience required to assess when the approximation is failing, though lots of empirical evidence being gathered.

Lognormal model

We now consider an alternative lognormal model for the relative risks, but still independent.

A Poisson-lognormal non-spatial random effect model is given by:

$$\begin{aligned} Y_i|\beta_0, \epsilon_i &\sim_{ind} \text{Poisson}(E_i e^{\beta_0} e^{\epsilon_i}), \\ \epsilon_i|\sigma_\epsilon^2 &\sim_{iid} \text{N}(0, \sigma_\epsilon^2) \end{aligned}$$

where ϵ_i are area-specific random effects that capture the residual or unexplained (log) relative risk of disease in area i, $i = 1, ..., n$.

In the previous notation $x = [\epsilon_1, \ldots, \epsilon_n, \beta_0]$, $\theta = \sigma_\epsilon^{-2}$; no $\alpha^\star$ since no dispersion parameter in a Poisson.

Note that in INLA the random effect variances are reported in terms of the precisions (the reciprocal of the variance).

Lognormal model

This model gives rise to the posterior distribution;

$$p(\beta_0, \tau_\epsilon, \epsilon_1, \ldots, \epsilon_n | y) = \frac{\prod_{i=1}^{n} \Pr(Y_i | \beta_0, \epsilon_i) p(\epsilon_i | \tau_\epsilon) p(\beta_0) p(\tau_\epsilon)}{\Pr(y)}.$$

The full posterior is an $(n+2)$-dimensional distribution and INLA by default produces summaries of the univariate posterior distributions for α and τ_ϵ.

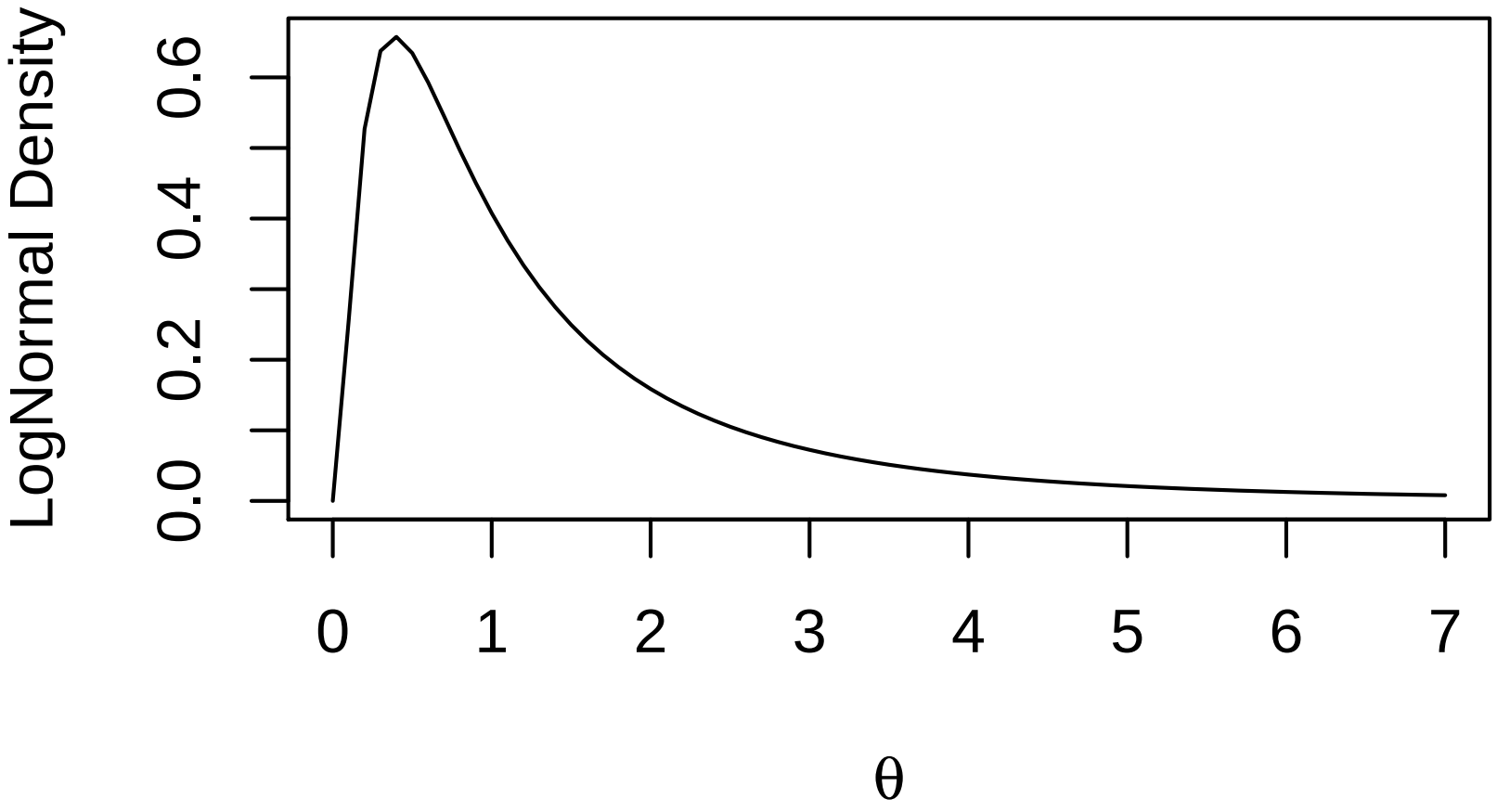
The posteriors on the random effects $p(\epsilon_i | y)$ can be extracted, as we will show in subsequent slides.

Prior choice for lognormal model

We specify the 5% and 95% points of the relative risk associated with β as 1 and 5.

```
lnprior <- LogNormalPriorCh(1, 5, 0.5, 0.95)
lnprior
## $mu
## [1] 0
##
## $sigma
## [1] 0.9784688
plot(seq(0, 7, 0.1), dlnorm(seq(0, 7, 0.1),
    meanlog = lnprior$mu, sdlog = lnprior$sigma),
    type = "l", xlab = expression(theta),
    ylab = "LogNormal Density")
```

Prior choice for Lognormal model



Prior choice for lognormal model

The priors $\sigma_\epsilon^{-2} \sim \text{Ga}(1, 0.0260)$ or $\sigma_\epsilon^{-2} \sim \text{Ga}(0.5, 0.0005)$ will often be suitable in a mapping context.

The 2.5%, 50% (median) and 97.5% quantiles for σ_ϵ are calculated below.

Recall that

$$\text{s.d.(Relative Risk)} \approx \sigma_\epsilon.$$

```
1/sqrt(qgamma(c(0.975, 0.5, 0.025), 0.5, 5e-04))
## [1] 0.01410848 0.04688400 1.00908784
1/sqrt(qgamma(c(0.975, 0.5, 0.025), 1, 0.026))
## [1] 0.08395362 0.19367517 1.01338302
```

Scottish lip cancer data

We present an alternative version of the dataset, with more information on the polygon files.

```
# First time: install.packages('INLA',
# repos=c(getOption('repos'),
# INLA='https://inla.r-inla-download.org/R/stable'),
# dep=TRUE)
library(INLA)
library(rgdal)
data(Scotland)
# Next read shapefiles, The shapefiles can be read
# from here:
# http://faculty.washington.edu/jonno/SISMIDmaterial/
# they are also available on the Canvas website
scot_shape <- readOGR(dsn = "./examples", layer = "scot")
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/austin/Dropbox/555-Materials/555-2020/555-202
## with 56 features
## It has 2 fields
## Integer64 fields read as strings:  ID
proj4string(scot_shape)  # No coordinate system so we assign one
```

Scottish lip cancer data

```
# the ordering of the rows is different in the INLA
# data as compared to the imported polygons
scot_shape$ID
##  [1] 12 13 19 2  17 16 21 50 15 25 26 29 43 39 40 52 42 51 34
## [26] 38 44 30 45 48 47 35 28 4  20 33 31 24 55 18 56 14 32 27
## [51] 5  11 1  7  23 37
## 56 Levels: 1 10 11 12 13 14 15 16 17 18 19 2 20 21 22 23 24 2
Scotland$Region
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
## [26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
## [51] 51 52 53 54 55 56
ID_D <- match(scot_shape$ID, Scotland$Region)
Scotland2 <- Scotland[ID_D, ]
row.names(Scotland2) <- row.names(scot_shape)
library(maptools)
scot_dat_shape <- spCbind(scot_shape, Scotland2)
names(scot_dat_shape)
## [1] "NAME"   "ID"     "Counts" "E"      "X"      "Region"
all.equal(scot_dat_shape$ID, scot_dat_shape$Region)
## [1] "'current' is not a factor"
```

Scottish lip cancer data

```
library(spdep)
Obs <- scot_dat_shape$Counts
Exp <- scot_dat_shape$E
scot_dat_shape$SMR <- Obs/Exp
# Ordnance Survey of Great Britain grid system
scot_OS <- spTransform(scot_dat_shape, CRS("+init=epsg:27700"))
# What type of object are we dealing with?
class(scot_OS)
## [1] "SpatialPolygonsDataFrame"
## attr(,"package")
## [1] "sp"
# Assessing attributes
scot_OS[["Counts"]]
##  [1]  5  3  9 39  2  9 16  6 17 19 15 16  2  6  4  1  8  1  8
## [26]  8  6 11 19  3  2 11 10  9  7  7  5  7  0  7  0  8  3  7
## [51] 15 13  9 26 11 11
# Dimension of attributes
dim(scot_OS)
## [1] 56  7
# Spatial dimensions
dimensions(scot_OS)
```

Scottish lip cancer data

```
spplot(scot_OS, c("SMR"), at = c(0, 0.25, 0.5, 0.8,
    1, 1.5, 2.5, 4.5, 7), col.regions = rev(brewer.pal(8,
    "RdBu")))
```

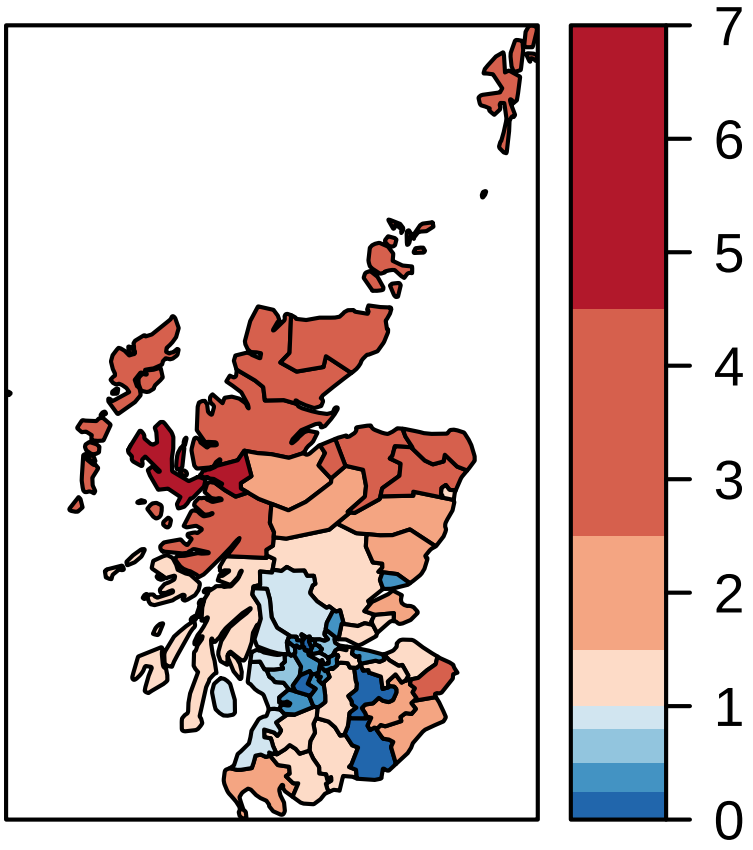


Figure 4: SMRs for Scottish lip cancer data, version 2

Scottish lip cancer data

```
spplot(scot_OS, c("X"), at = c(0, 5, 10, 15, 20, 25),
    col.regions = rev(brewer.pal(8, "RdBu")))
```

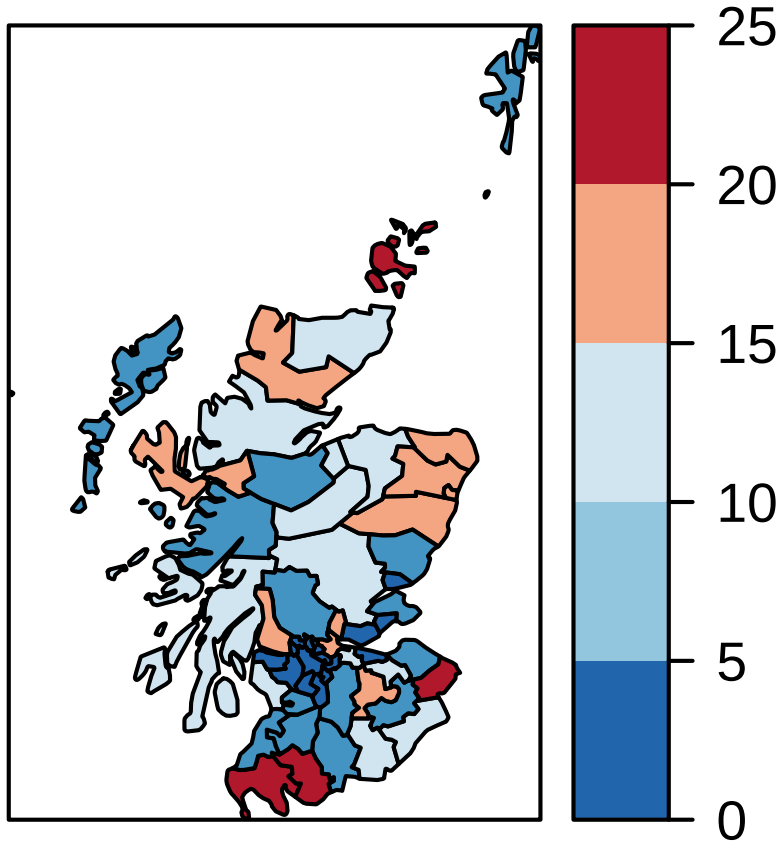


Figure 5: AFF (as a percentage) for Scottish lip cancer data.

INLA for lognormal model

We fit the Poisson-Lognormal model for Scotland.

```
# Fit Poisson-lognormal model in INLA:
pcprec <- list(theta = list(prior = "pc.prec",
    param = c(1, 0.05)))
scotland.fit1 <- inla(Counts ~ 1 + f(Region,
    model = "iid", hyper = pcprec), data = Scotland,
    family = "poisson", E = E, control.predictor = list(compute
```

Note the specification of the penalized complexity prior for the precision $\tau_\epsilon = \sigma_\epsilon^{-2}$. Here we specify that there is a 5% chance that the standard deviation σ_ϵ is greater than 1.

The default prior for β_0 (the intercept) is a zero mean normal with a large standard deviation.

In the `f()` function it is implicit that all random effects are normal.

INLA for lognormal model

```
names(scotland.fit1)
##  [1] "names.fixed"                "summary.fixed"
##  [3] "marginals.fixed"            "summary.lincomb"
##  [5] "marginals.lincomb"          "size.lincomb"
##  [7] "summary.lincomb.derived"    "marginals.lincomb.derived"
##  [9] "size.lincomb.derived"       "mlik"
## [11] "cpo"                        "po"
## [13] "waic"                       "model.random"
## [15] "summary.random"             "marginals.random"
## [17] "size.random"                "summary.linear.predictor"
## [19] "marginals.linear.predictor" "summary.fitted.values"
## [21] "marginals.fitted.values"    "size.linear.predictor"
## [23] "summary.hyperpar"           "marginals.hyperpar"
## [25] "internal.summary.hyperpar"  "internal.marginals.hyperpar"
## [27] "offset.linear.predictor"    "model.spde2.blc"
## [29] "summary.spde2.blc"          "marginals.spde2.blc"
## [31] "size.spde2.blc"             "model.spde3.blc"
## [33] "summary.spde3.blc"          "marginals.spde3.blc"
## [35] "size.spde3.blc"             "logfile"
## [37] "misc"                       "dic"
## [39] "mode"                       "neffp"
## [41] "joint.hyper"                "nhyper"
## [43] "version"                    "Q"
## [45] "graph"                      "ok"
## [47] "cpu.used"                   "all.hyper"
## [49] ".args"                      "call"
## [51] "model.matrix"
```

INLA for lognormal model

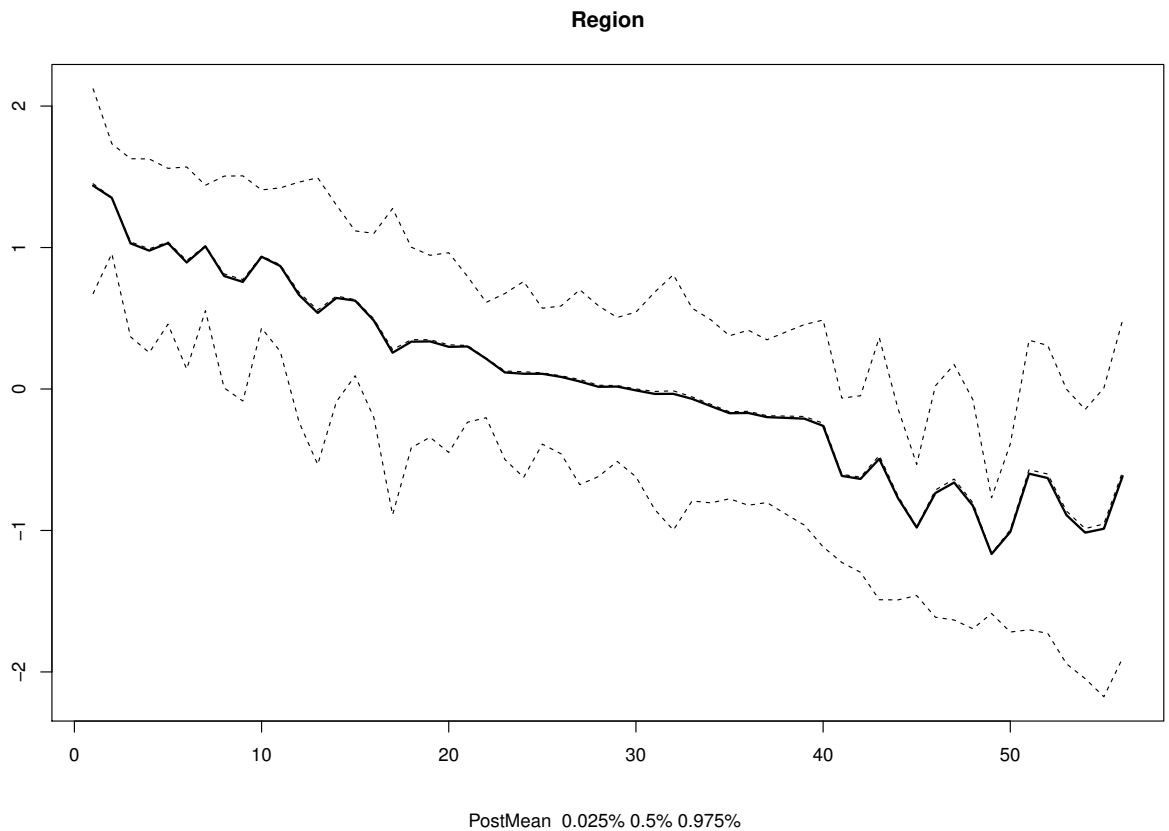
```
summary(scotland.fit1)
##
## Call:
##    c("inla(formula = Counts ~ 1 + f(Region, model = \"iid\", hyper =
##    pcprec), ", " family = \"poisson\", data = Scotland, E = E,
##    control.predictor = list(compute = TRUE))" )
## Time used:
##     Pre = 2.47, Running = 0.495, Post = 0.153, Total = 3.12
## Fixed effects:
##               mean    sd 0.025quant 0.5quant 0.975quant  mode kld
## (Intercept) 0.081 0.117     -0.154    0.082      0.307 0.085   0
##
## Random effects:
##   Name     Model
##     Region IID model
##
## Model hyperparameters:
##                       mean   sd 0.025quant 0.5quant 0.975quant mode
## Precision for Region 1.80 0.45       1.06     1.75       2.82 1.65
##
## Expected number of effective parameters(stdev): 43.78(2.06)
## Number of equivalent replicates : 1.28
##
## Marginal log-Likelihood:  -185.45
## Posterior marginals for the linear predictor and
##  the fitted values are computed
expbeta0med <- scotland.fit1$summary.fixed[4]  # intercept
sdmed <- 1/sqrt(scotland.fit1$summary.hyperpar[4])  # sd
```

The posterior median for β_0 is 0.082221207490197 2 and for σ_ϵ is 0.75653241773178.

Lognormal model: posterior marginals

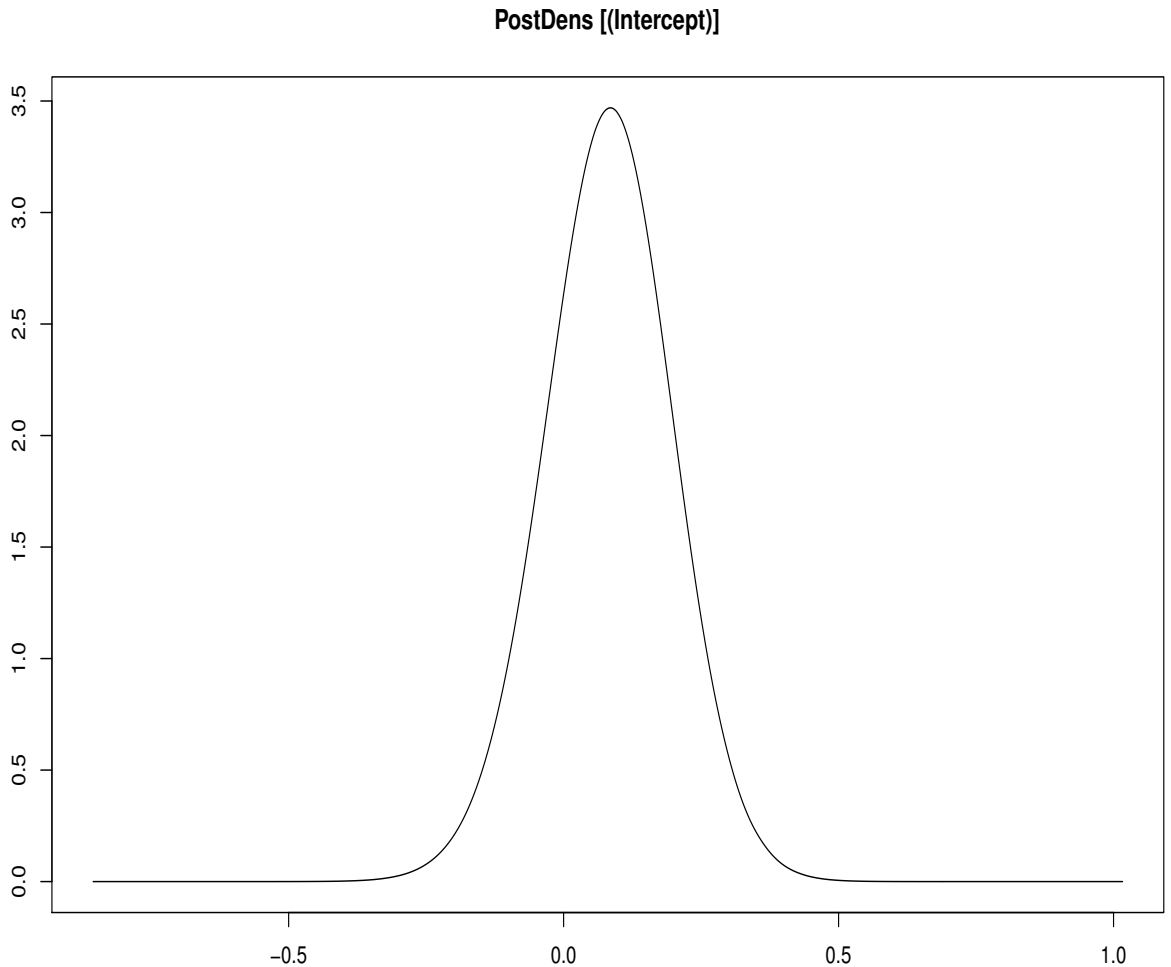
Note that the inla plot function and knitr don't play nicely so we create a postscript file and then include.

```
plot(scotland.fit1, plot.hyperparameter = FALSE, plot.random.eff
    plot.fixed.effects = FALSE, prefix = "logmodplot1",
    postscript = T)
```



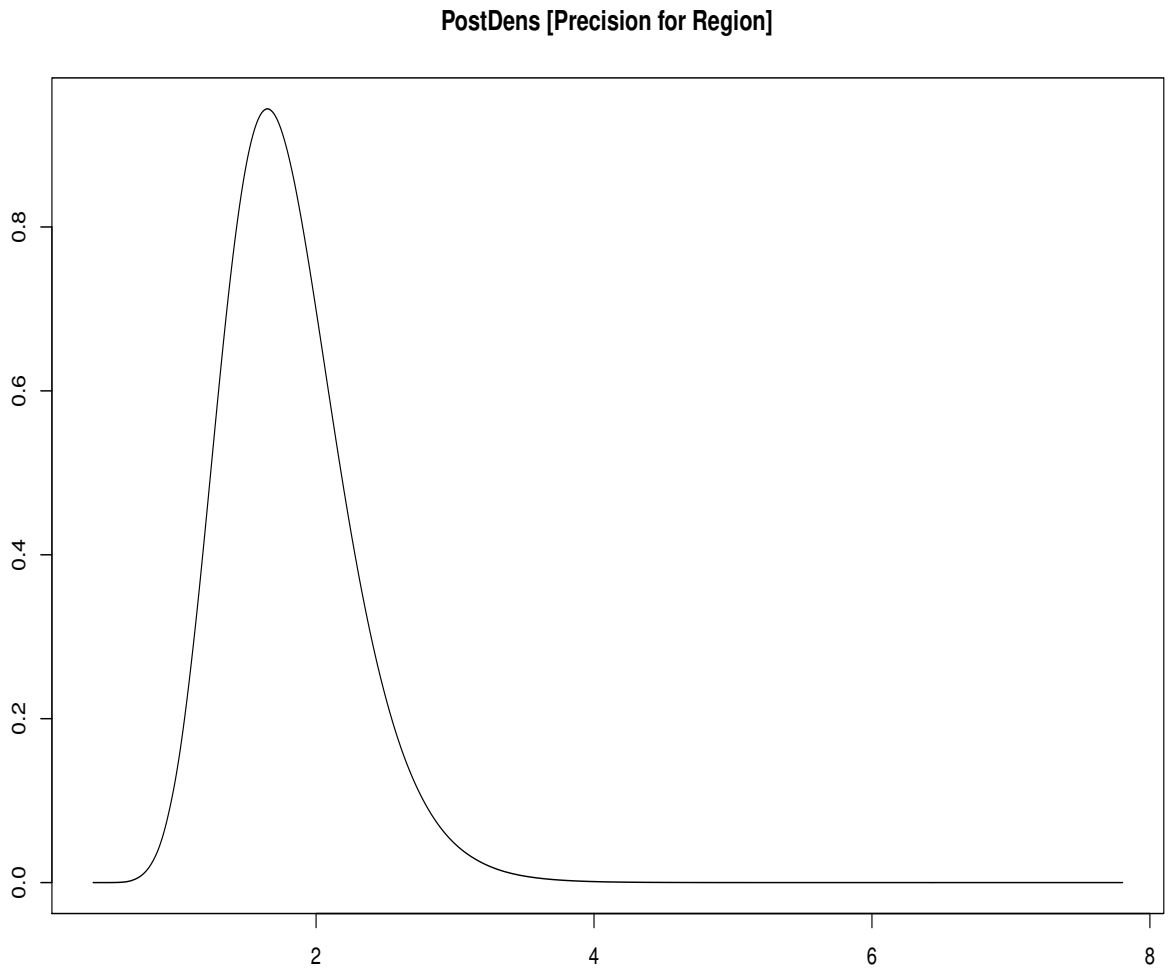
Lognormal model: posterior marginals

```
plot(scotland.fit1, plot.hyperparameter = FALSE, plot.random.eff
    plot.fixed.effects = TRUE, prefix = "logmodplot2",
    postscript = T)
```



Lognormal model: posterior marginals

```
plot(scotland.fit1, plot.hyperparameter = TRUE, plot.random.effe
    plot.fixed.effects = FALSE, prefix = "logmodplot3",
    postscript = T)
```



Comparison of lognormal and gamma models

```
scotland.fit1$summary.fixed
##                      mean         sd 0.025quant   0.5quant 0.975q
## (Intercept) 0.08074233 0.1167741 -0.1536124 0.08222121  0.306
##                         kld
## (Intercept) 6.902654e-06
# First illustrate how to extract the intercept
# (beta0)
lnorminter <- scotland.fit1$summary.fixed[4]
# Now extract the medians of the random effects
# (which are centered around alpha)
lnormREs <- exp(scotland.fit1$summary.random$Region[5])
```

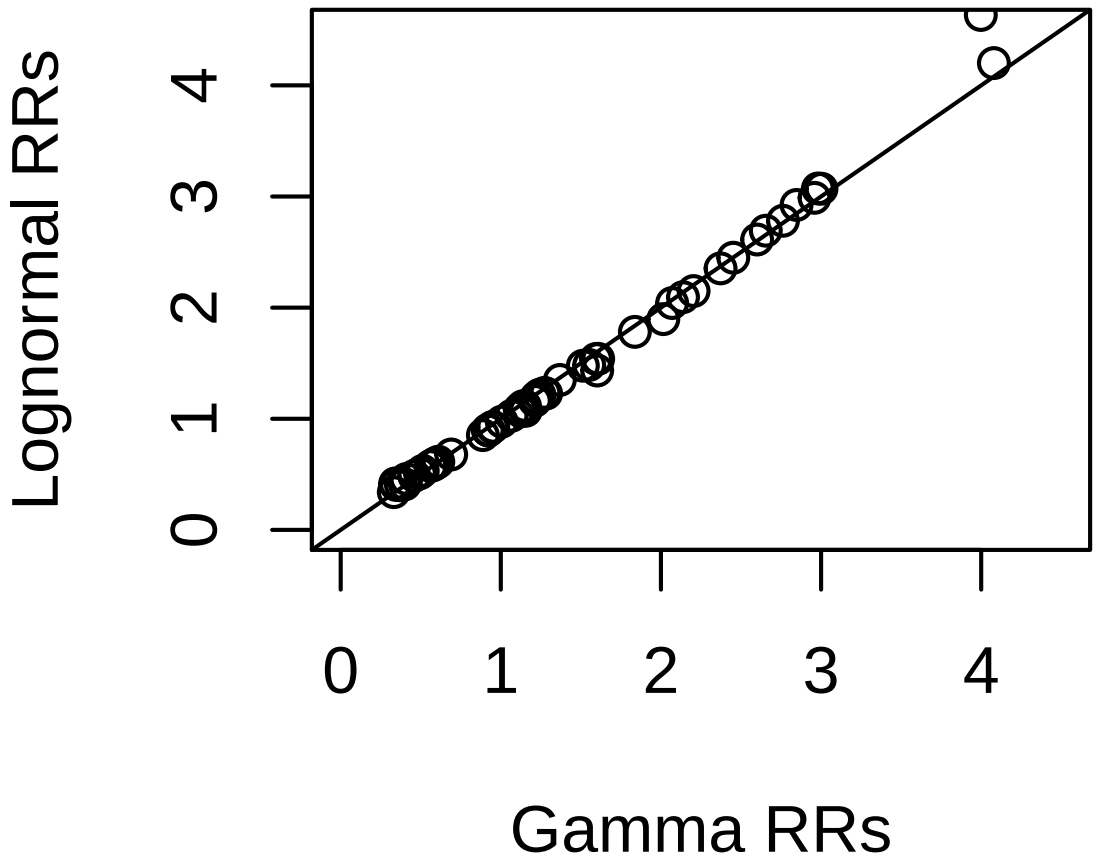
Comparison of lognormal and gamma models

We now compare the gamma and lognormal analyses.

The figure and maps below show the results: the analyses provide very similar estimates, which is reassuring.

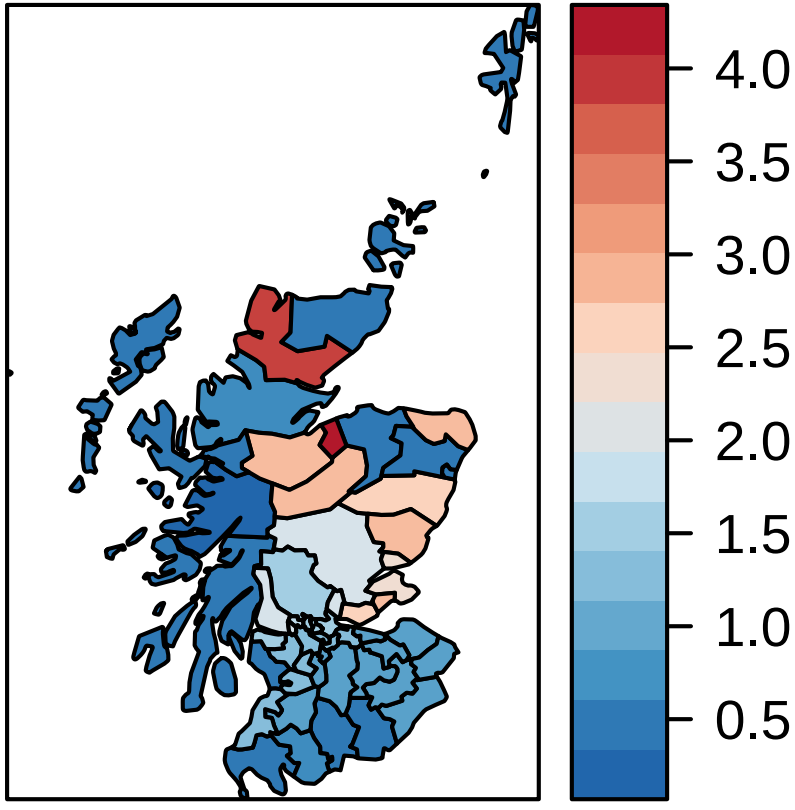
Comparison of lognormal and gamma models

```
lnormRRs <- as.double(exp(lnorminter)) * lnormREs[,
    1]
plot(ebresults$RR, lnormRRs, xlim = c(0, 4.5), ylim = c(0,
    4.5), xlab = "Gamma RRs", ylab = "Lognormal RRs")
abline(0, 1)
```



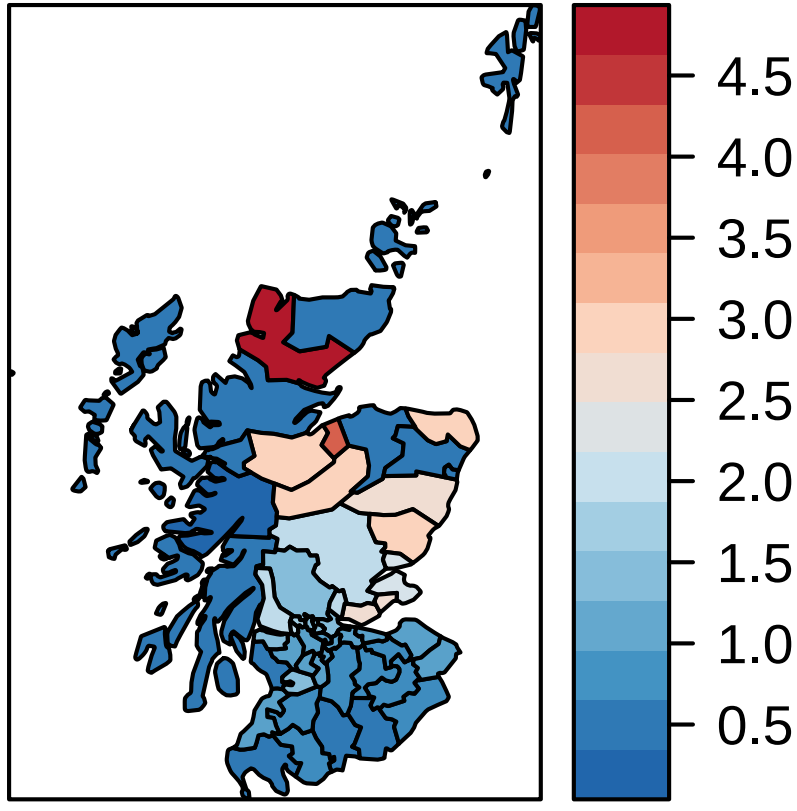
Mapped gamma relative risk estimates

```
scot_OS$RRgam <- ebresults$RR
spplot(scot_OS, c("RRgam"), col.regions = colorRampPalette(rev(brewer.pal(8,
    "RdBu")))(50))
```



Mapped lognormal relative risk estimates

```
scot_OS$RRlnorm <- lnormRRs
spplot(scot_OS, c("RRlnorm"), col.regions = colorRampPalette(rev(brewer.pal(8,
    "RdBu")))(50))
```




```
# par(mar=c(1,1,1,1)) An alternative way to plot
# mapvariable(lnormRRs,scotland.map,lower=0,upper=4.5)
```

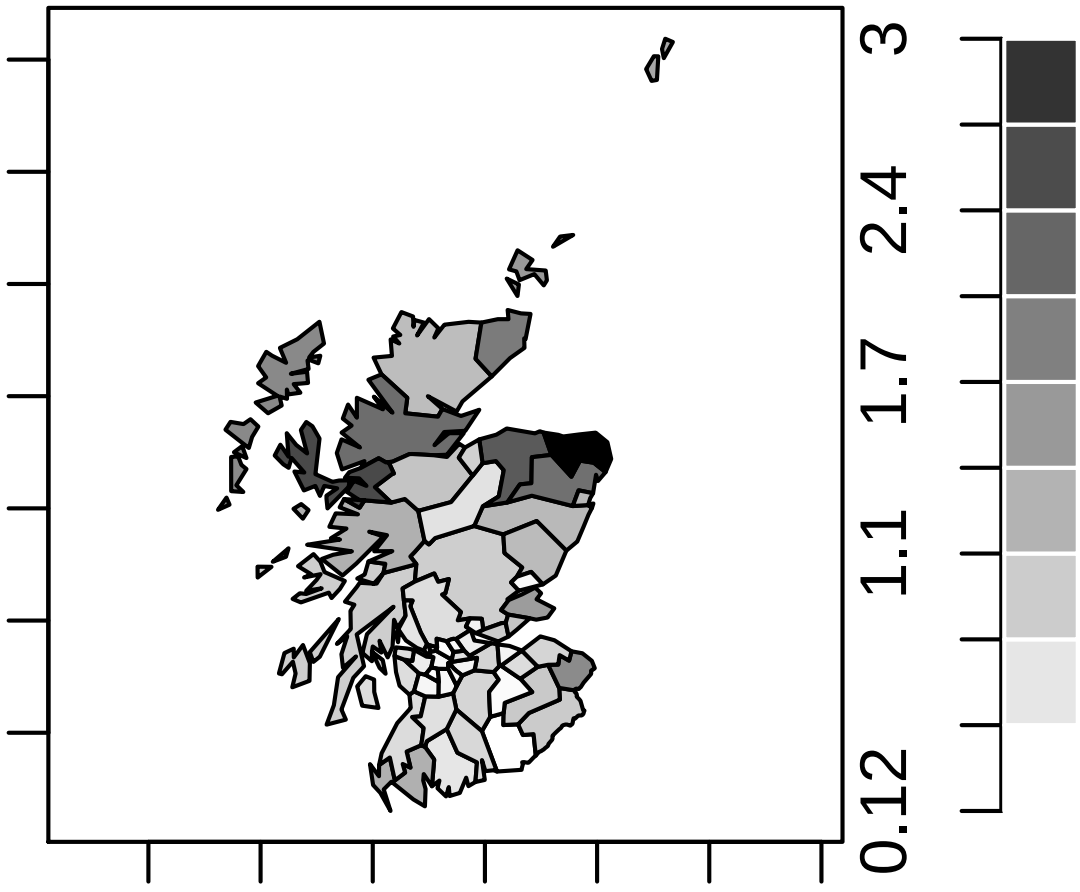
Uncertainty estimates of Lognormal estimates

We calculate the 2.5% and 97.5% points of the gamma posteriors of the relative risks.

```
LNlower <- exp(scotland.fit1$summary.linear.predictor["0.025quan
LNupper <- exp(scotland.fit1$summary.linear.predictor["0.975quan
LNwidth <- LNupper[, 1] - LNlower[, 1]
```

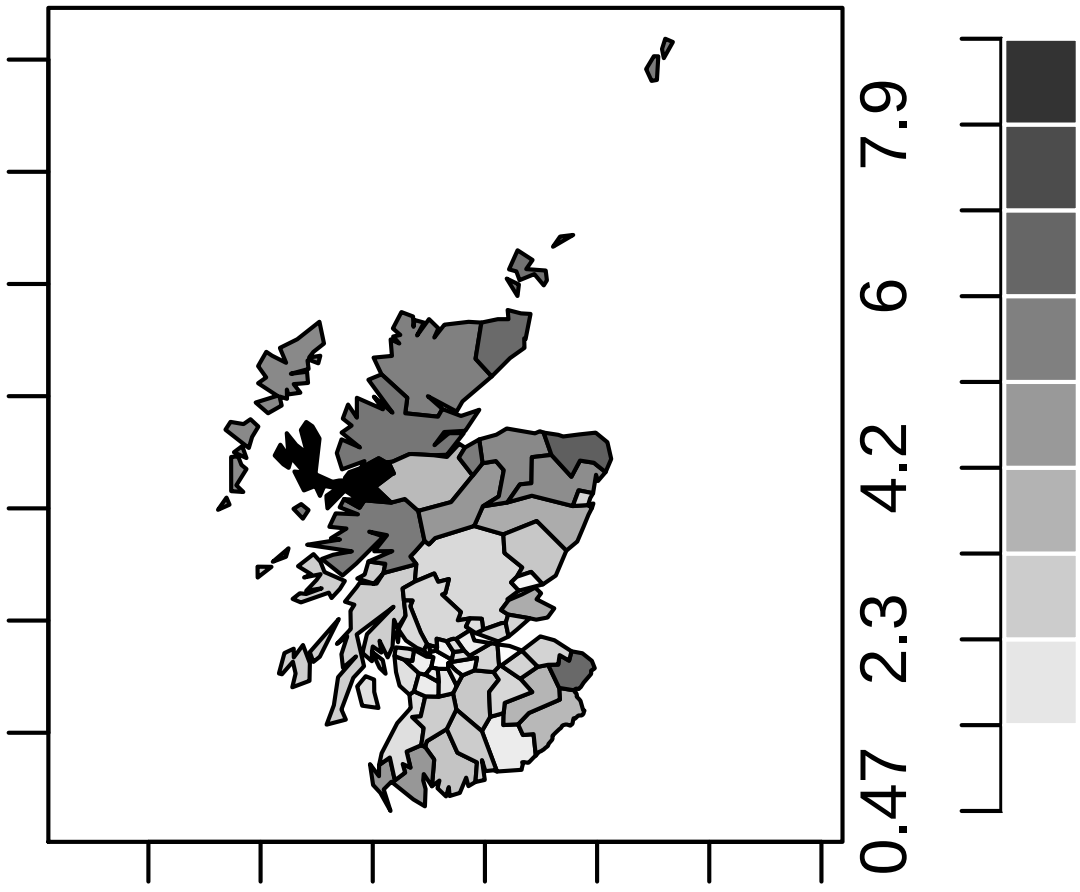
Lower 2.5% point of Lognormal estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(LNlower[, 1], scotland.map)
```



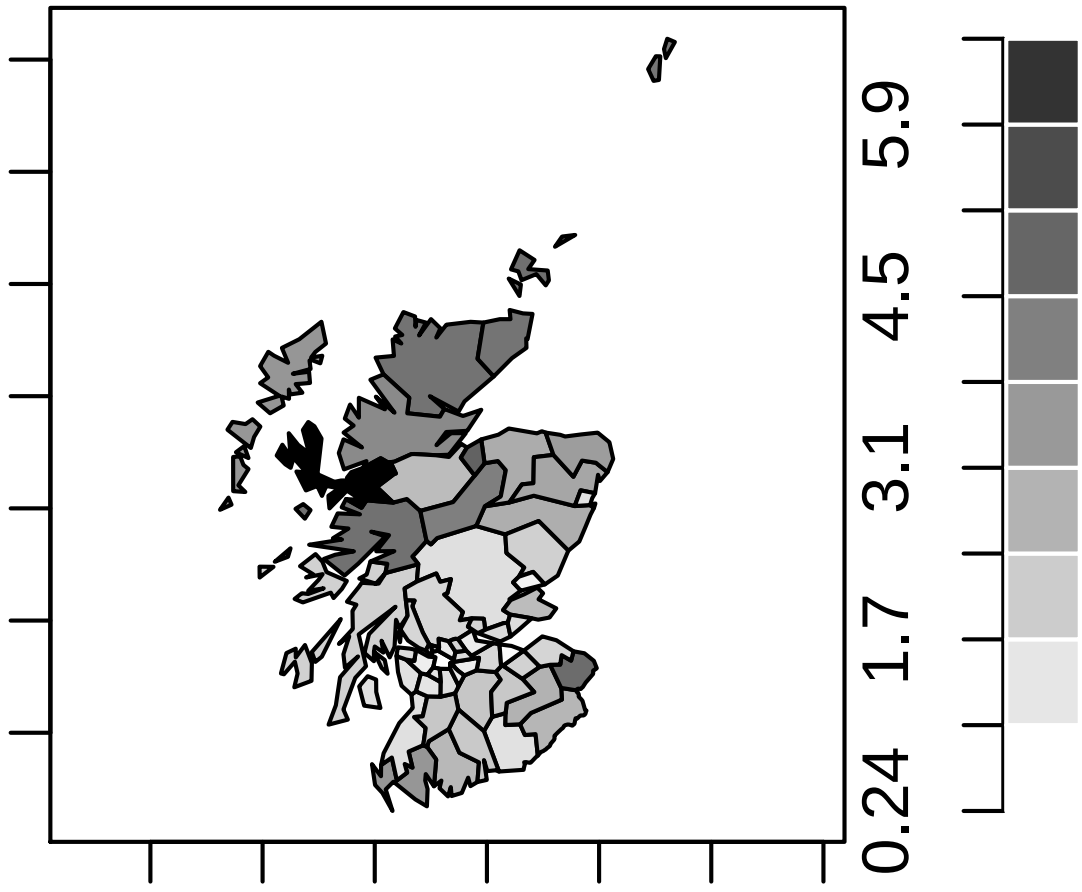
Lower 97.5% point of Lognormal estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(LNupper[, 1], scotland.map)
```



Width of Lognormal estimate

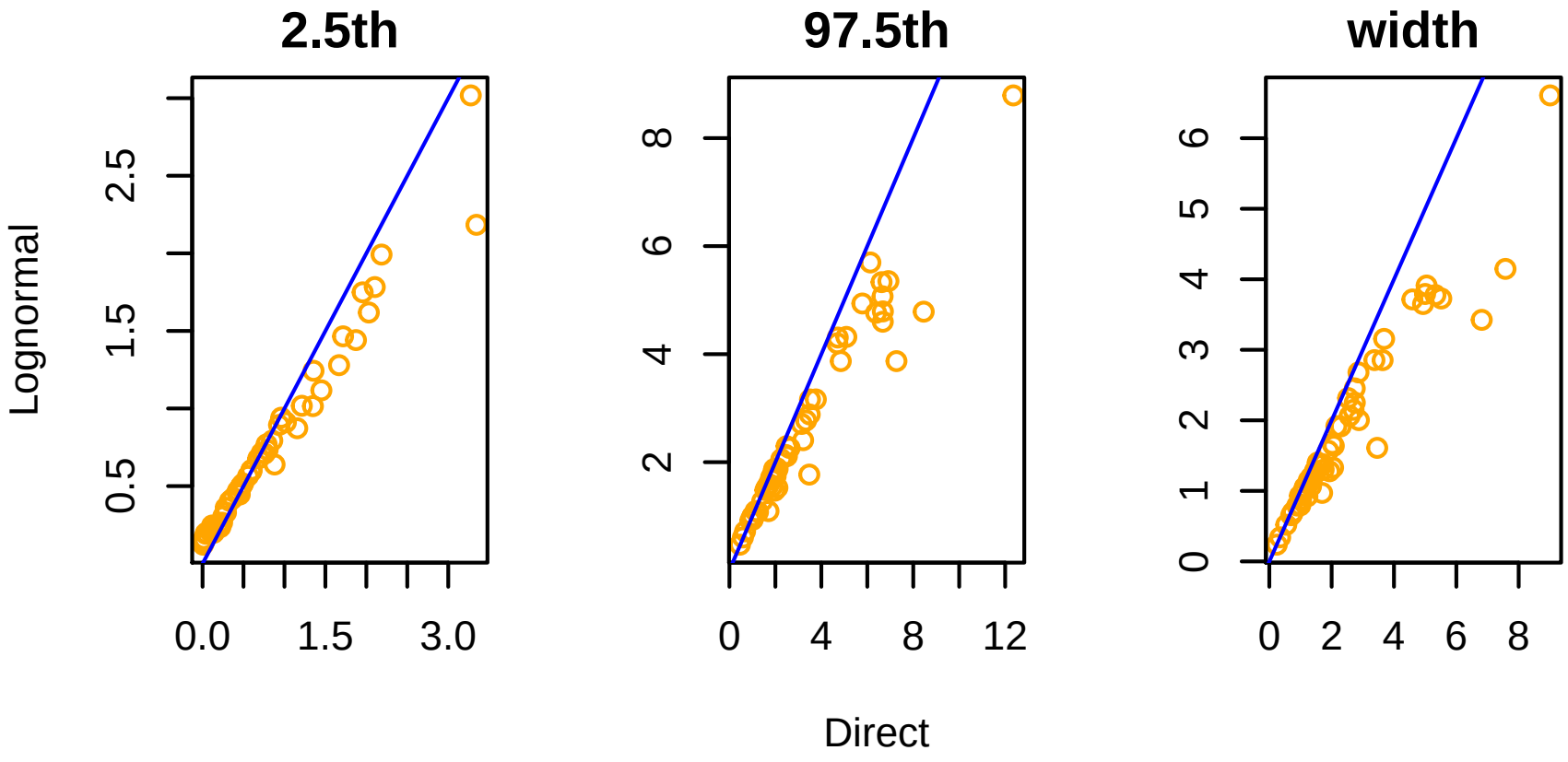
```
par(mar = c(1, 1, 1, 1))
mapvariable(LNwidth, scotland.map)
```



Comparison of interval widths

```
par(mfrow = c(1, 3), mar = c(5, 4, 2, 1))
plot(LNlower[, 1] ~ SMRlower, col = "orange", ylab = "Lognormal"
    xlab = "", main = "2.5th")
abline(0, 1, col = "blue")
plot(LNupper[, 1] ~ SMRupper, col = "orange", ylab = "",
    xlab = "Direct", main = "97.5th")
abline(0, 1, col = "blue")
plot(LNwidth ~ SMRwidth, col = "orange", ylab = "",
    xlab = "", main = "width")
abline(0, 1, col = "blue")
```

Comparison of interval widths



Scottish lip cancer

We now fit the three-stage model:

Stage 1: The Likelihood $Y_i|\theta_i \sim \text{Poisson}(E_i\theta_i)$, $i = 1, \ldots, n$ with

$$\log \theta_i = \beta_0 + x_i\beta_1 + \epsilon_i$$

where x_i is the AFF in area i.

Stage 2: The random effects (prior distribution) is $\epsilon_i|\sigma^2_\epsilon \sim_{iid} N(0, \sigma^2_\epsilon)$.

Stage 3: The hyperprior on the hyperparameters $\beta_0, \beta_1, \sigma^2_\epsilon$:

$$p(\beta_0, \beta_1, \sigma^2_\epsilon) = p(\beta_0)p(\beta_1)p(\sigma^2_\epsilon)$$

so that here we have assumed independent priors.

Lognormal non-spatial model with covariates

```
# No spatial effects with covariate
scotland.fit1X <- inla(Counts ~ 1 + I(X) + f(Region,
    model = "iid", hyper = pcprec), data = Scotland,
    family = "poisson", E = E)
summary(scotland.fit1X)
##
## Call:
##    c("inla(formula = Counts ~ 1 + I(X) + f(Region, model = \"iid\", hyper
##    = pcprec), ", " family = \"poisson\", data = Scotland, E = E)")
## Time used:
##     Pre = 2.14, Running = 0.338, Post = 0.104, Total = 2.58
## Fixed effects:
##               mean    sd 0.025quant 0.5quant 0.975quant   mode kld
## (Intercept) -0.492 0.160     -0.812   -0.490     -0.182 -0.486   0
## I(X)         0.068 0.014      0.040    0.068      0.096  0.068   0
##
## Random effects:
##   Name     Model
##     Region IID model
##
## Model hyperparameters:
##                       mean   sd 0.025quant 0.5quant 0.975quant mode
## Precision for Region 2.94 0.84       1.63     2.82       4.90 2.61
##
## Expected number of effective parameters(stdev): 39.61(2.83)
## Number of equivalent replicates : 1.41
##
## Marginal log-Likelihood:  -183.31
```

Lognormal non-spatial model with covariates: inference

If we are interested in the association with the AFF variable we can examine the posterior summaries, on the original (to give a log RR) or exponentiated (to give a RR) scale.

From these summaries we might extract the posterior median as a point estimate, or take the 2.5% and 97.5% points as a 95% credible interval.

```
scotland.fit1X$summary.fixed[2, ]
##              mean          sd 0.025quant   0.5quant 0.975quant       mode
## I(X) 0.06836914 0.01425896 0.04026373 0.06836309 0.09647358 0.06835543
##                kld
## I(X) 1.396388e-06
exp(scotland.fit1X$summary.fixed[2, ])
##          mean       sd 0.025quant 0.5quant 0.975quant     mode      kld
## I(X) 1.07076 1.014361   1.041085 1.070754    1.10128 1.070746 1.000001
```

Parameter interpretation

```
scotland.fit1X$summary.fixed
##                    mean          sd  0.025quant    0.5quant  0.975quant
## (Intercept) -0.49185747 0.15973539 -0.81194384 -0.49003576 -0.18215318
## I(X)         0.06836914 0.01425896  0.04026373  0.06836309  0.09647358
##                    mode          kld
## (Intercept) -0.48647992 2.692218e-06
## I(X)         0.06835543 1.396388e-06
```

The posterior mean for the intercept is $E[\beta_0|y] = -0.49$.

The posterior median for the relative risk associated with a 1 unit increase in X is median$(\exp(\beta_1)|y) = \exp(0.068) = 1.07$. This latter calculation exploits the fact that we can transform quantiles[1]

Similarly a 95% credible interval for the relative risk $\exp(\beta_1)$ is

$$[\ \exp(0.040), \exp(0.096)\ ] = [\ 1.04, 1.10\ ].$$

Examination of such intervals is a common way of determining whether the association is "significant" – here we have strong evidence that the relative risk associated with AFF is significant.

[1] unlike means since, for example, $E[\exp(\beta_1)|y] \neq \exp(E[\beta_1|y])$.

Scottish Lip Cancer: Parameter Interpretation

```
scotland.fit1X$summary.fixed
##                    mean          sd  0.025quant    0.5quant  0.975quant
## (Intercept) -0.49185747 0.15973539 -0.81194384 -0.49003576 -0.18215318
## I(X)         0.06836914 0.01425896  0.04026373  0.06836309  0.09647358
##                    mode          kld
## (Intercept) -0.48647992 2.692218e-06
## I(X)         0.06835543 1.396388e-06
```

The posterior median of σ_ϵ is $1/\sqrt{2.8} = 0.582$ and a 95% interval is

$$[1/\sqrt{5.13}, 1/\sqrt{1.70}] = [0.44, 0.766].$$

A more interpretable quantity is an interval on the residual relative risk (RRR). The latter follow a lognormal distribution LogNormal$(0, \sigma_\epsilon^2)$ so a 95% interval is $\exp(\pm 1.96 \times \sigma_\epsilon)$.

Scottish Lip Cancer: Parameter Interpretation

A posterior median of a 95% RRR interval is

$$\begin{aligned}&[\exp(-1.96 \times \text{median}(\sigma_\epsilon)), \exp(1.96 \times \text{median}(\sigma_\epsilon)] \\ = &[\exp(-1.96 \times 0.582), \exp(1.96 \times 0.582)] = [0.320, 3.13]\end{aligned}$$

which is quite wide.

This also explains the disparity with the prior – large residual relative risks such as this are quite surprising.

A more in depth analysis would examine the prior sensitivity to the prior on τ_ϵ.

Variances are in general more difficult to estimate than regression coefficients so there is often sensitivity (unless the number of areas is very large.

Lognormal Spatial Smoothing Model

Lognormal spatial model with one covariate

We now add spatial (ICAR) random effects to the model. We parameterize in terms of total variance and propotion that is spatial. We place a penalized complexity prior on these two parameters.

We need a graph file containing the neighbors.

```
# Spatial effects with covariate
formula <- Counts ~ 1 + I(X) + f(Region, model = "bym2",
    graph = "./examples/scotland.graph", scale.model = T,
    constr = T, hyper = list(phi = list(prior = "pc",
        param = c(0.5, 0.5), initial = 1), prec = list(pri
        param = c(0.5/0.31, 0.01), initial = 5)))
scotland.fit2 <- inla(formula, data = Scotland, family = "p
    E = E, control.predictor = list(compute = TRUE))
```

Lognormal spatial model with covariates

```
scotland.fit2$summary.hyperpar
##                                mean          sd 0.025quant 0.5quant
## Precision for Region 4.6604059 1.44297810  2.4472807 4.454379
## Phi for Region       0.9332262 0.07705094  0.7166428 0.960391
##                                  mode
## Precision for Region 4.070135
## Phi for Region       0.997625
```

The posterior median of the total standard deviation (on the log relative risk scale) is $1/\sqrt{(4.45)} = 0.47$.

The posterior median for the proportion of the residual variation that is spatial is 0.96.

Lognormal spatial model with covariates

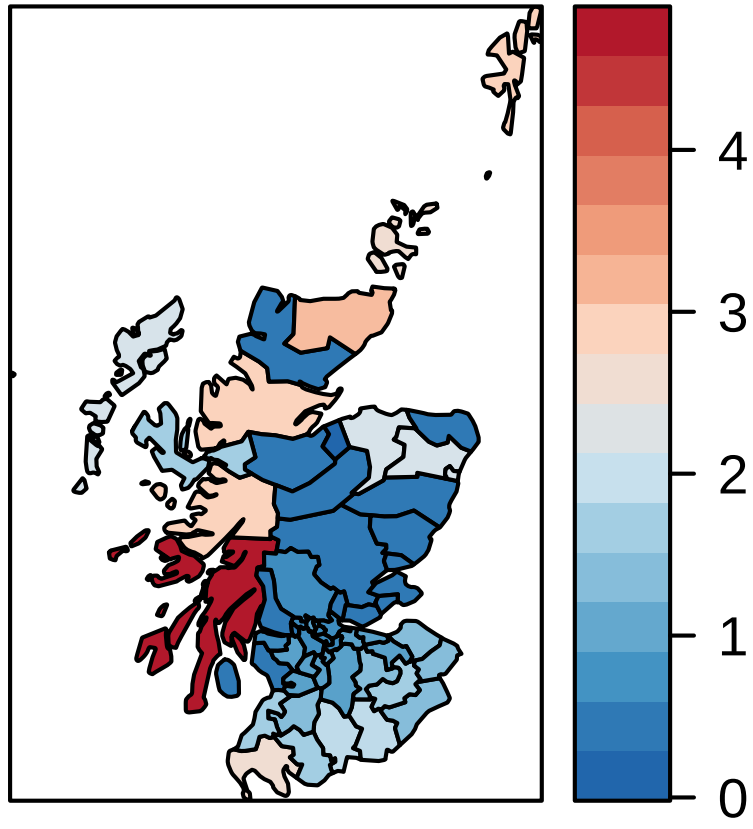
Now we provide maps of the non-spatial and spatial random effects.

Estimates of residual relative risk (posterior medians), of the non-spatial e^{ϵ_i} and the spatial contributions e^{S_i}.

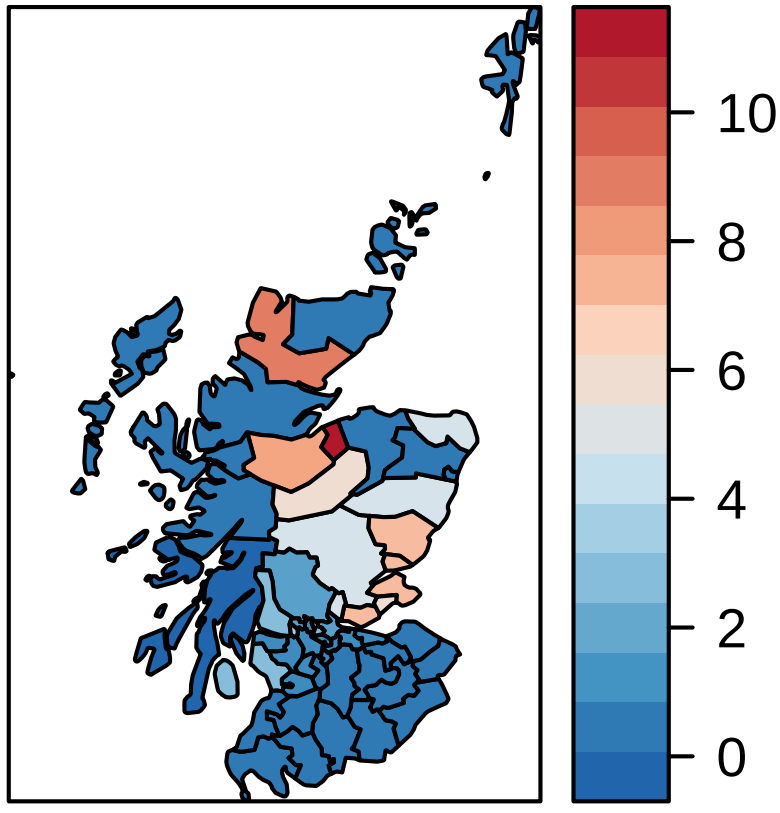
Note the differences in the scales: the spatial random effects dominate here.

```
Diff <- scotland.fit2$summary.random$Region[1:56, 2] -
    scotland.fit2$summary.random$Region[57:112, 2]
REsnonspat <- exp(Diff)
REsspat <- exp(scotland.fit2$summary.random$Region[57:112,
    5])
scot_OS$REsnonspat <- REsnonspat
scot_OS$REsspat <- REsspat
spplot(scot_OS, c("REsnonspat"), col.regions = colorRampPalette(
    "RdBu")))(50))
# par(mfrow=c(1,1),mar=c(.1,.1,.1,.1))
# mapvariable(REsnonspat[,1],scotland.map)
```

Lognormal spatial model with covariates: non-spatial random effects

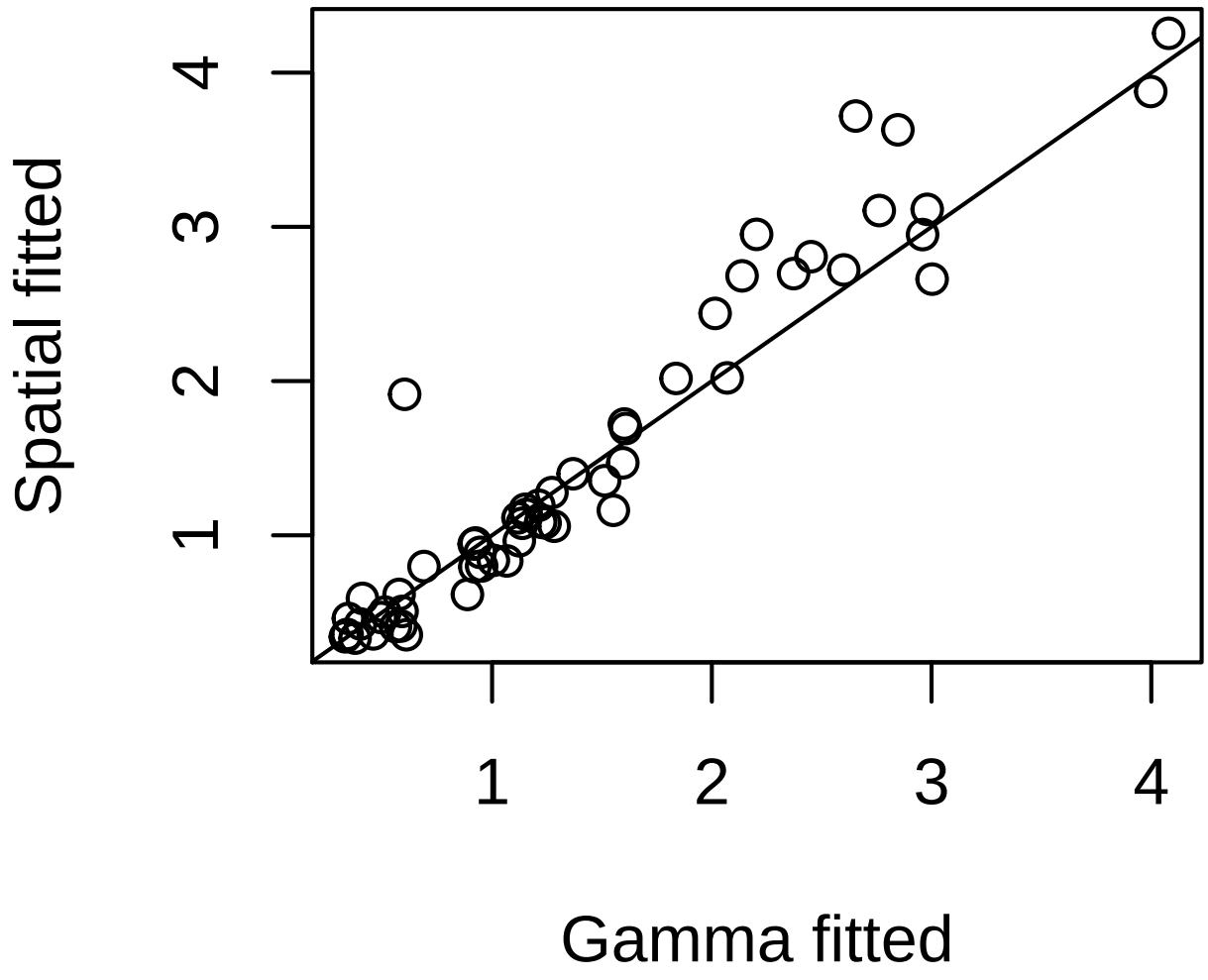


Lognormal spatial model with covariates: spatial random effects



Comparison of spatial lognormal and gamma fits: some differences

```
plot(ebresults$RR, scotland.fit2$summary.fitted.values[,
    4], xlab = "Gamma fitted", ylab = "Spatial fitted")
abline(0, 1)
```



INLA representation

Recall the spatial model is $Y_i|\theta_i \sim \text{Poisson}(E_i\theta_i)$ with

$$\log\theta_i = \beta_0 + \beta_1 x_i + \epsilon_i + S_i.$$

The relationship between the statistical model and the `inla` specification is:

$$\begin{aligned}
\log\theta_i &= \underbrace{\beta_0}_{\texttt{1}} + \underbrace{x_i\beta_1}_{\texttt{I(X)}} \\
&+ \underbrace{\epsilon_i}_{\texttt{f(Region,model="iid")}} \\
&+ \underbrace{S_i}_{\texttt{f(Region2,model="besag",graph="scotland.graph")}}
\end{aligned}$$

Spatial model: confounding by location

The command `plot(scotland.fit2)` provides plots of: marginal posterior distributions of β_0, β_1, σ_ϵ^{-2}, σ_S^{-2} and summaries of the random effects ϵ_i, S_i and the linear predictors and fitted values, all by area.

Note that the posterior mean estimate of β_1 associated with AFF goes from $0.068 \rightarrow 0.026$ when moving from the non-spatial to spatial model.

This is known as confounding by location.

The model attributes spatial variability in risk to either the covariate or to the spatial random effects.

Scotland

The posterior median estimate of σ_ϵ decreases from $1/\sqrt{2.9475} = 0.58$ to $1/\sqrt{94.986} = 0.10$ when the spatial random effect is added.

The posterior median estimate of σ_s is $1/\sqrt{1.125} = 0.94$ but, as already noted, this value is not directly comparable to the estimate of σ_ϵ.

However, the scales on the figures shows that the spatial component dominates for these data.

A rough estimate of the standard deviation of the spatial component can be determined by empirically calculating the standard deviation of the random effect estimates $\widehat{S}_i$.

A more complete analysis would address the sensitivity to the prior specifications on σ_ϵ and σ_s.

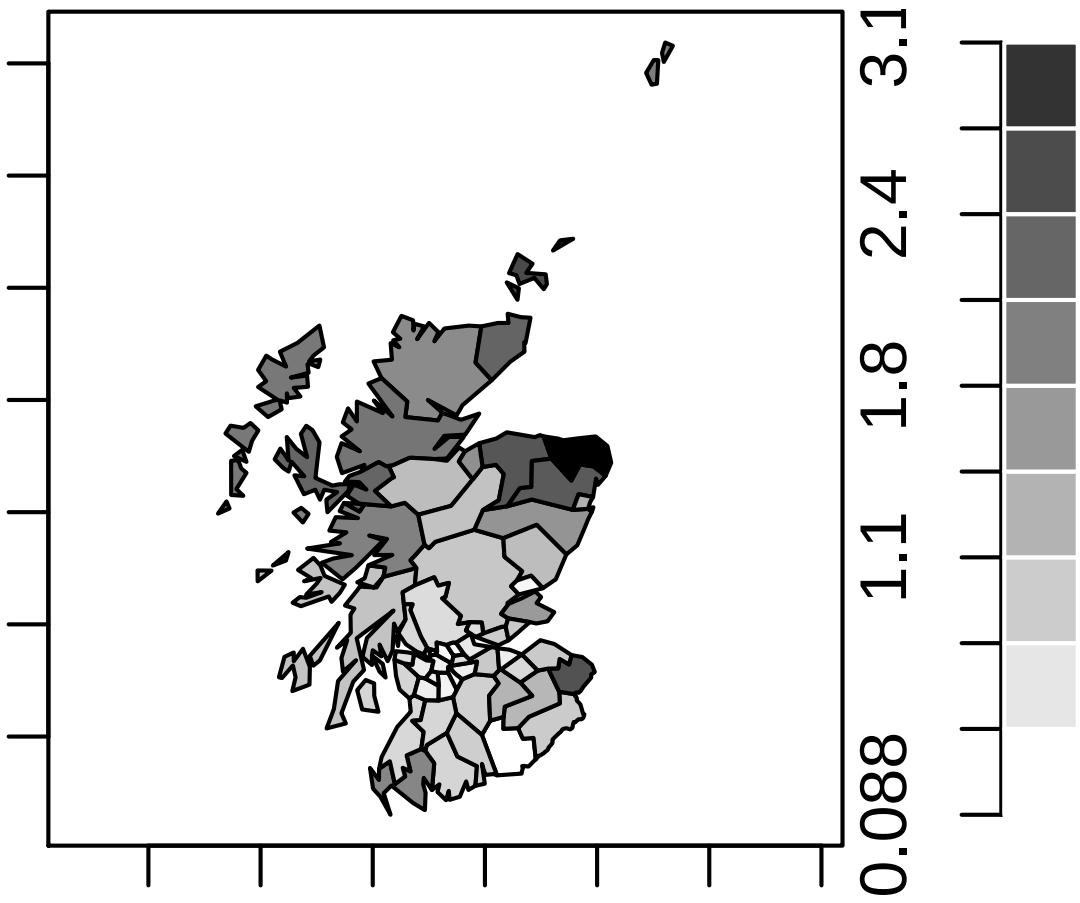
Uncertainty estimates of spatial estimates

We calculate the 2.5% and 97.5% points of the gamma posteriors of the relative risks.

```
LN2lower <- exp(scotland.fit2$summary.linear.predictor["0.025qua
LN2upper <- exp(scotland.fit2$summary.linear.predictor["0.975qua
LN2width <- LN2upper[, 1] - LN2lower[, 1]
```

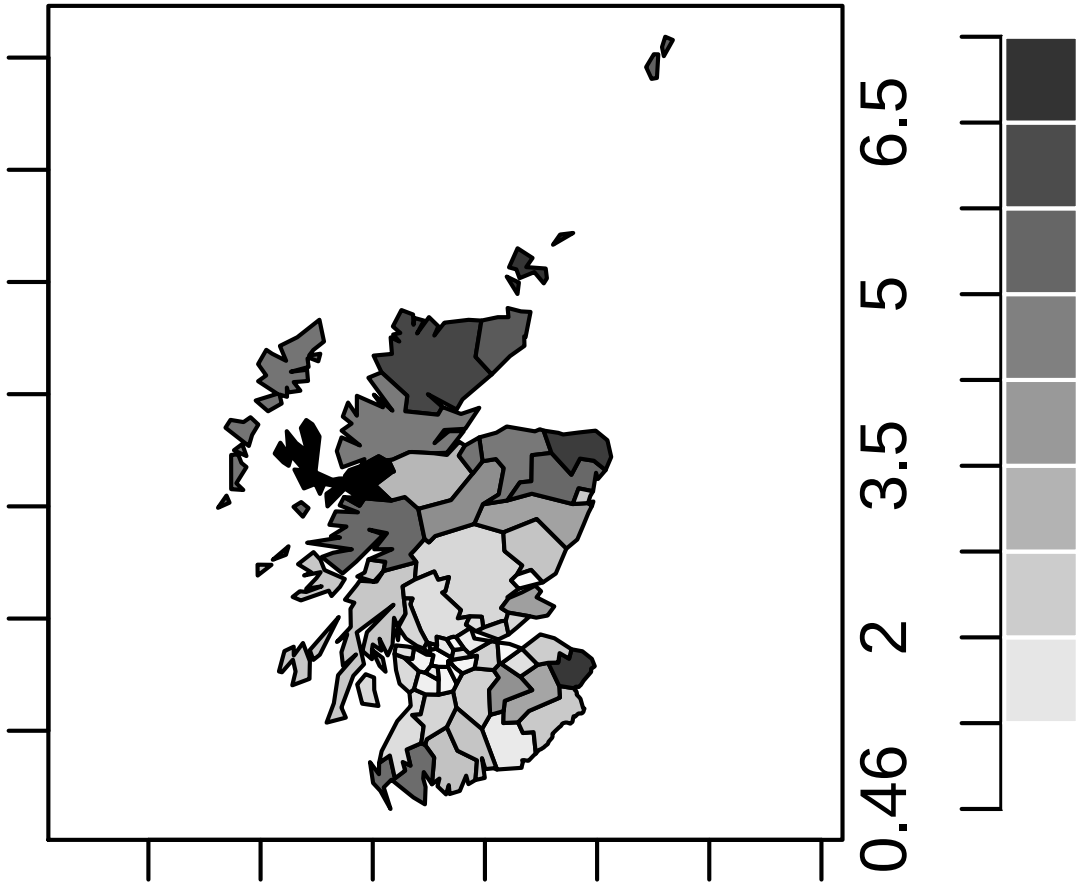
Lower 2.5% point of Lognormal estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(LN2lower[, 1], scotland.map)
```



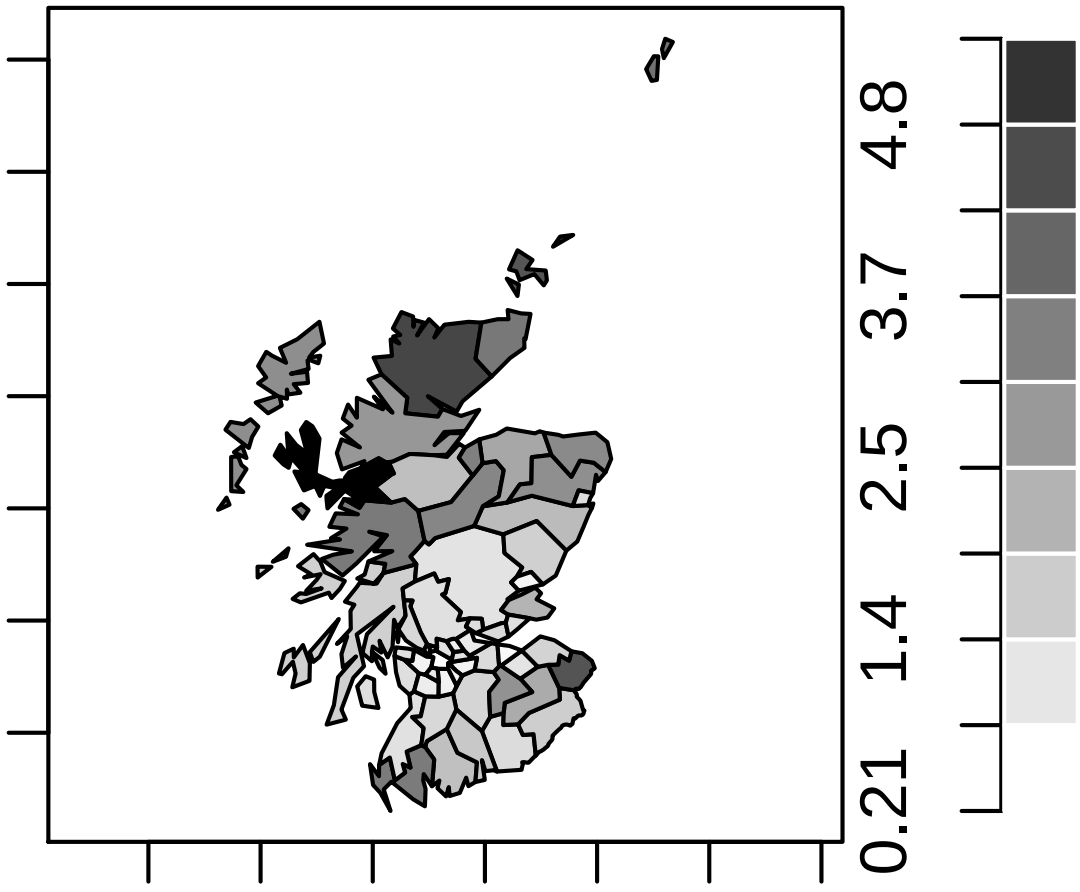
Lower 97.5% point of spatial estimate

```
par(mar = c(1, 1, 1, 1))
mapvariable(LN2upper[, 1], scotland.map)
```



Width of spatial estimate

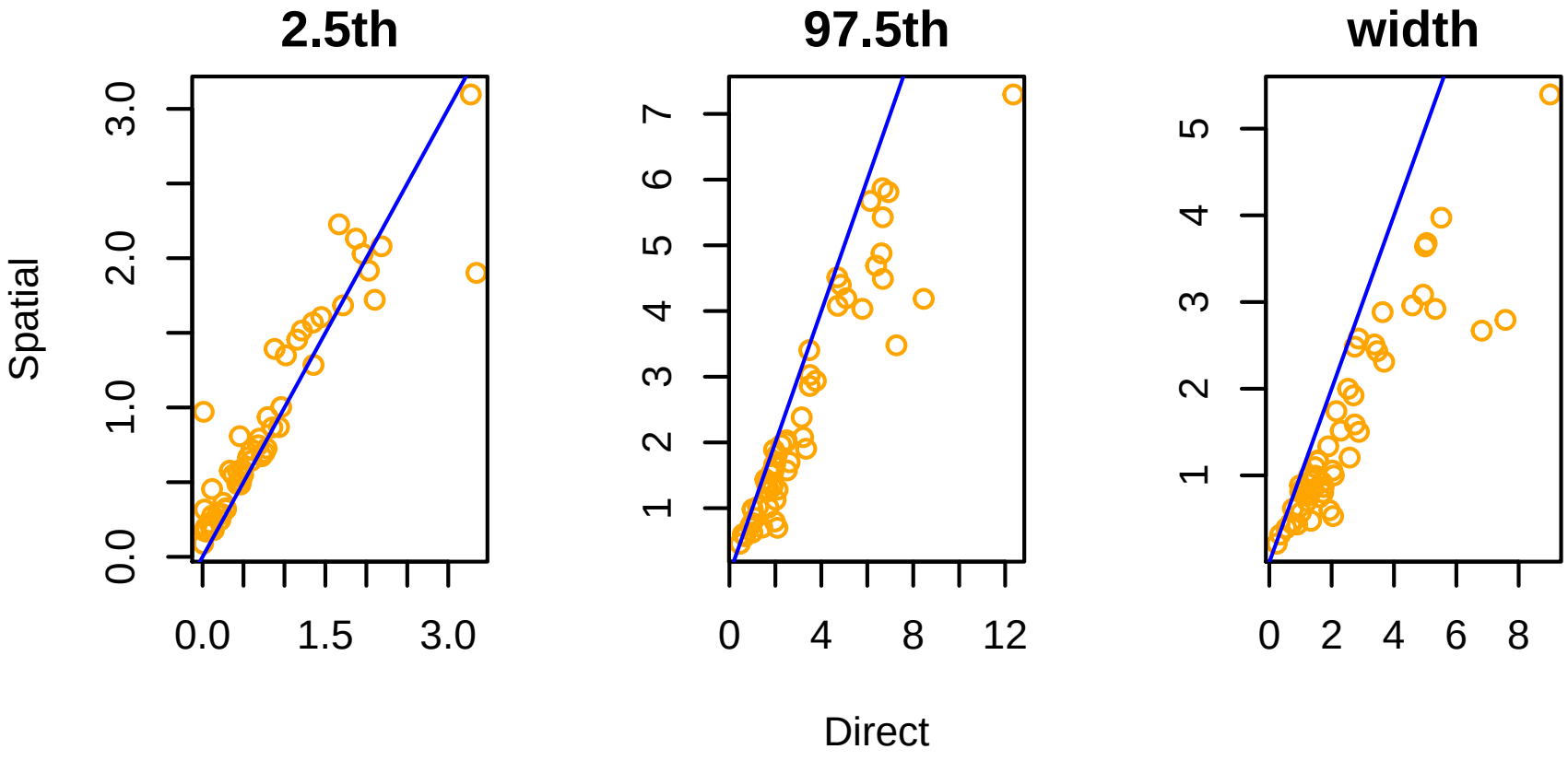
```
par(mar = c(1, 1, 1, 1))
mapvariable(LN2width, scotland.map)
```



Comparison of interval widths

```
par(mfrow = c(1, 3), mar = c(5, 4, 2, 1))
plot(LN2lower[, 1] ~ SMRlower, col = "orange", ylab = "Spatial",
    xlab = "", main = "2.5th")
abline(0, 1, col = "blue")
plot(LN2upper[, 1] ~ SMRupper, col = "orange", ylab = "",
    xlab = "Direct", main = "97.5th")
abline(0, 1, col = "blue")
plot(LN2width ~ SMRwidth, col = "orange", ylab = "",
    xlab = "", main = "width")
abline(0, 1, col = "blue")
```

Comparison of interval widths



Lognormal Spatial Smoothing Model with Relative Risk as Outcome

Relative risks as the outcome variable

Often the data arise in the form of observed rates or observed relative risks. For illustration, we imagine we had received the latter for Scotland, rather than the full data.

We model the log relative risk directly assuming they have a Gaussian distribution. We define $Z_i = \log \widehat{\theta}_i$ to emphasize that the observed data are now taken to be the log relative risks.

Recall that if any of the counts $Y_i = 0$ (which would result in a relative risk of zero and a standard error of zero), we can use the approximations $Y_i^\star = Y_i + 0.5$ and $E_i^\star = E_i + 0.5$ to calculate $\widehat{\theta}_i^\star = \frac{Y_i^\star}{E_i^\star}$. In these cases, $Z_i = \log \widehat{\theta}_i^\star$. For simplcity, we assume this has been done.

Relative risks as the outcome variable

In INLA, we can fit the model (with *'s if necessary)

$$Z_i = \log\left(\frac{Y_i}{E_i}\right) \sim \mathrm{N}(\mu_i, \sigma^2)$$

where $\mu_i = E[Z_i]$.

Relative risks as the outcome variable

INLA estimates the precision for the Gaussian observations, $1/\sigma^2$. We evaluate the variance of the observed "data" by using the Poisson variance assumption —the mean equals the variance (in general the standard error of the rate can be estimated in a variety of ways, including the jackknife).

Therefore, the Z_i have "known" variances that we can approximate using the Delta method (as we did previously) as

$$\widehat{\text{var}}(Z_i) = [E_i \exp(Z_i)]^{-1} = \frac{1}{E_i \widehat{\theta}_i}$$

with *'s if needed.

Scotland: modeling the relative risk

We calculate the log relative risks as $Z_i = \log \widehat{\theta^\star}_i$ and their variances for the Scotland lip cancer data.

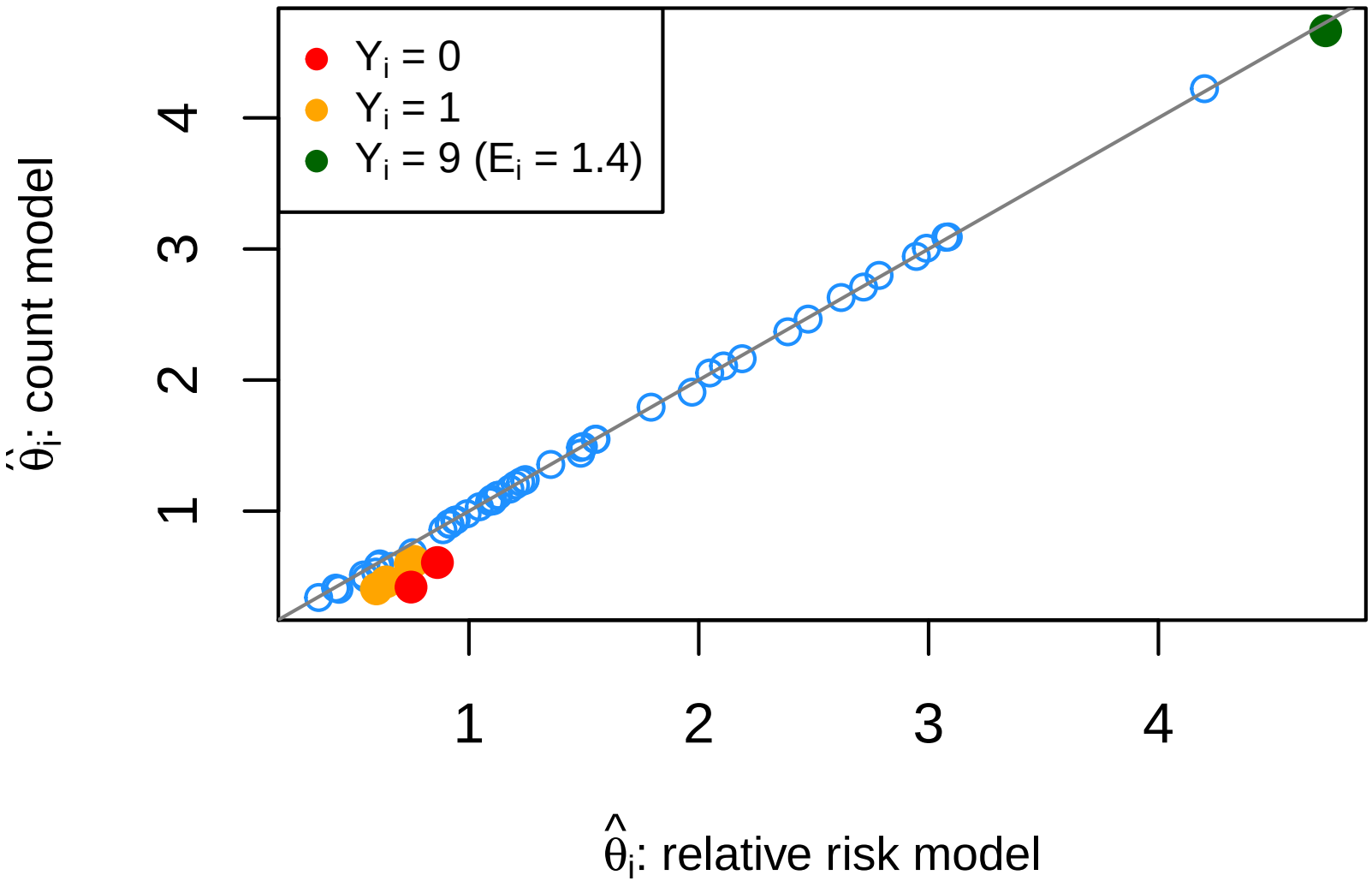
```
Scotland$Ystar <- ifelse(Scotland$Counts == 0, Scotland$Counts +
    0.5, Scotland$Counts)
Scotland$Estar <- ifelse(Scotland$Counts == 0, Scotland$E +
    0.5, Scotland$E)
Scotland$thetastar <- Scotland$Ystar/Scotland$Estar
Scotland$Z <- log(Scotland$thetastar)
Scotland$varZ <- 1/(Scotland$Estar * Scotland$thetastar)
Scotland$precZ <- 1/Scotland$varZ
```

Gaussian relative risk model with IID Normal random effects

We can fit a Gaussian model for log relative risks in INLA with fixed Gaussian precisions (equivalent to known variance) using the following code.

```
pcprec <- list(theta = list(prior = "pc.prec", param = c(1,
    0.05)))
scotland.fit3a <- inla(Z ~ 1 + f(Region, model = "iid",
    hyper = pcprec), data = Scotland, family = "gaussian",
    control.predictor = list(compute = TRUE), control.famil
        fixed = TRUE))), scale = precZ)
```

Comparison of Poisson-lognormal count outcome model and relative risk outcome model fits



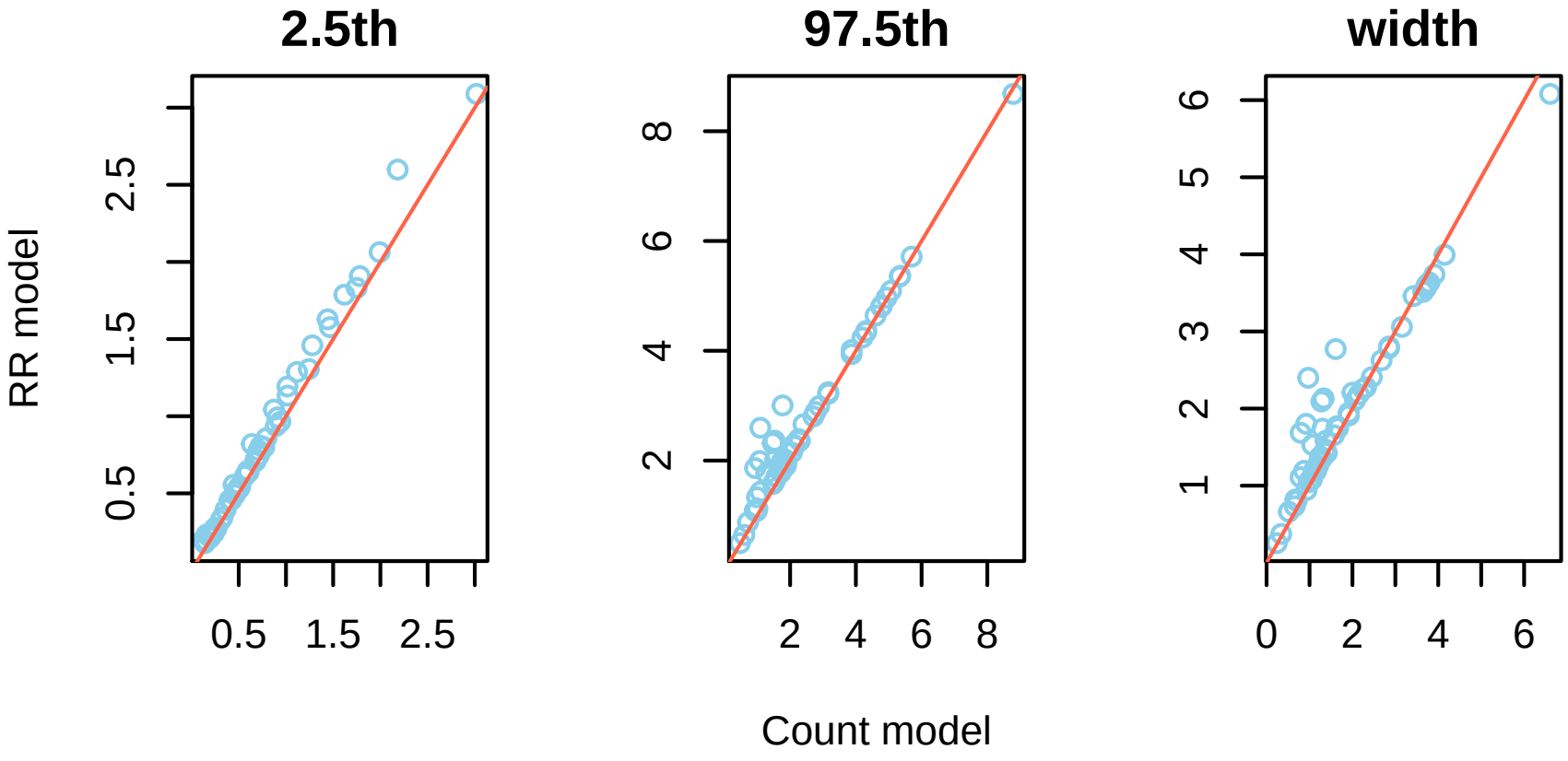
Comparison of interval widths

We calculate the 2.5% and 97.5% quantiles of the posteriors of the relative risks to compare with the count outcome model.

```
LN3alower <- exp(scotland.fit3a$summary.linear.predictor["0.025q
LN3aupper <- exp(scotland.fit3a$summary.linear.predictor["0.975q
LN3awidth <- LN3aupper[, 1] - LN3alower[, 1]

par(mfrow = c(1, 3), mar = c(5, 4, 2, 1))
plot(LN3alower[, 1] ~ LNlower[, 1], col = "skyblue",
    ylab = "RR model", xlab = "", main = "2.5th")
abline(0, 1, col = "tomato")
plot(LN3aupper[, 1] ~ LNupper[, 1], col = "skyblue",
    ylab = "", xlab = "Count model", main = "97.5th")
abline(0, 1, col = "tomato")
plot(LN3awidth ~ LNwidth, col = "skyblue", ylab = "",
    xlab = "", main = "width")
abline(0, 1, col = "tomato")
```

Comparison of interval widths



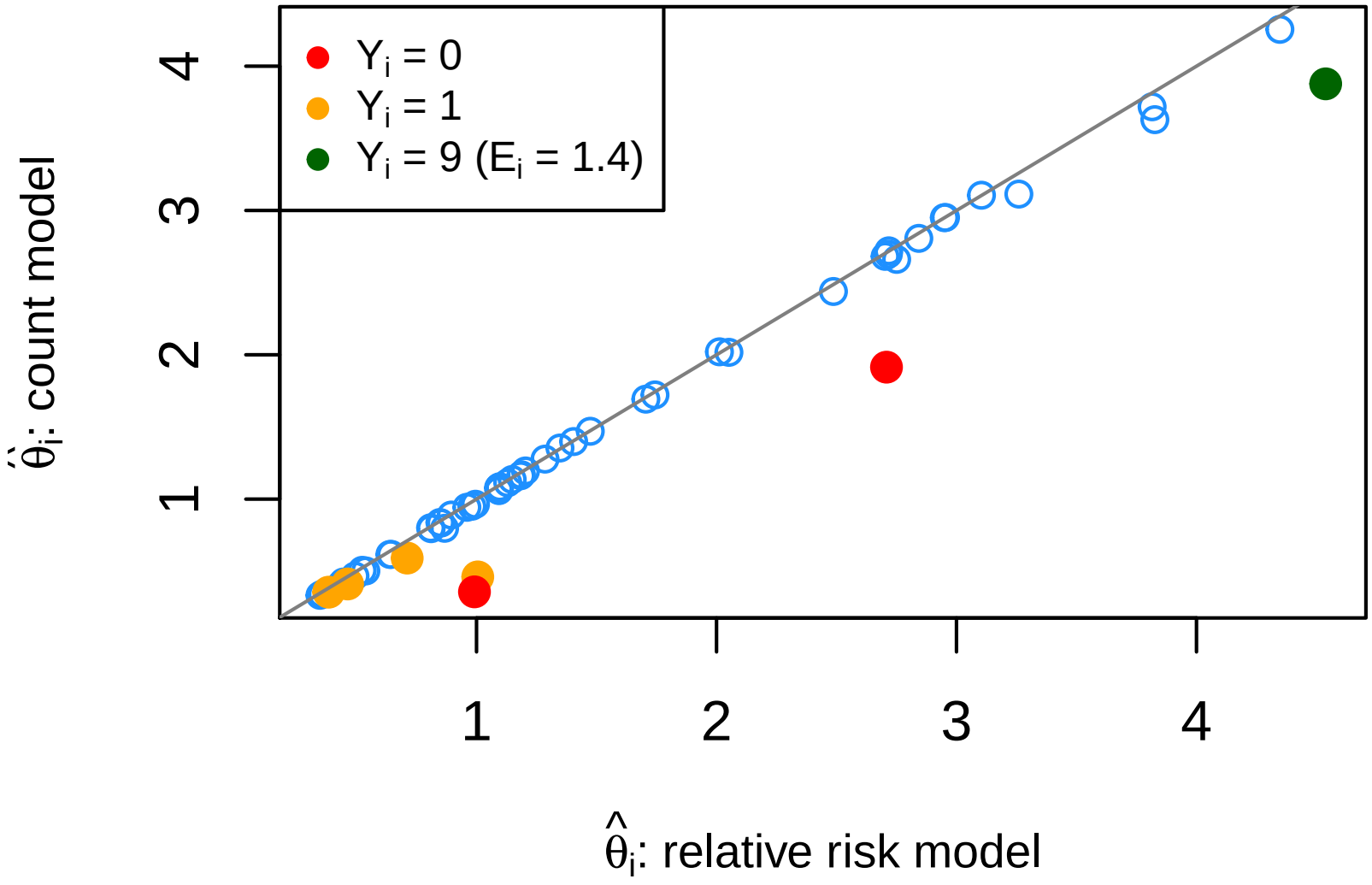
Spatial Gaussian relative risk model with IID Normal random effects and a covariate

Here, we add in a covariate and the spatial random effects

```
formula <- Z ~ 1 + I(X) + f(Region, model = "bym2",
    graph = "./examples/scotland.graph", scale.model = T,
    constr = T, hyper = list(phi = list(prior = "pc",
        param = c(0.5, 0.5), initial = 1), prec = list(prior = "pc.prec",
        param = c(0.5/0.31, 0.01), initial = 5)))

scotland.fit3 <- inla(formula, data = Scotland, family = "gaussian",
    control.predictor = list(compute = TRUE), control.family = list(hyper = lis
        fixed = TRUE))), scale = precZ)
```

Comparison of spatial Poisson-lognormal count outcome and relative risk outcome fits: differences in low and high extremes



Regression coefficient comparison

Very similar estimates (and posterior uncertainty estimates) of the regresion coefficents:

Count model:

```
##                     mean         sd   0.025quant    0.5quant 0.975quant
## (Intercept) -0.11714359 0.11175635 -0.338785561 -0.11651854 0.10075606
## I(X)         0.02644697 0.01160508  0.003279193  0.02656581 0.04894246
##                     mode          kld
## (Intercept) -0.11533089 8.686284e-06
## I(X)         0.02680688 5.581667e-06
```

Relative risk model:

```
##                     mean         sd   0.025quant    0.5quant 0.975quant
## (Intercept) -0.01520115 0.11101082 -0.233796177 -0.01509666 0.20255059
## I(X)         0.02504156 0.01141191  0.002390797  0.02511510 0.04727136
##                     mode          kld
## (Intercept) -0.01493117 4.496613e-06
## I(X)         0.02526635 9.078526e-06
```

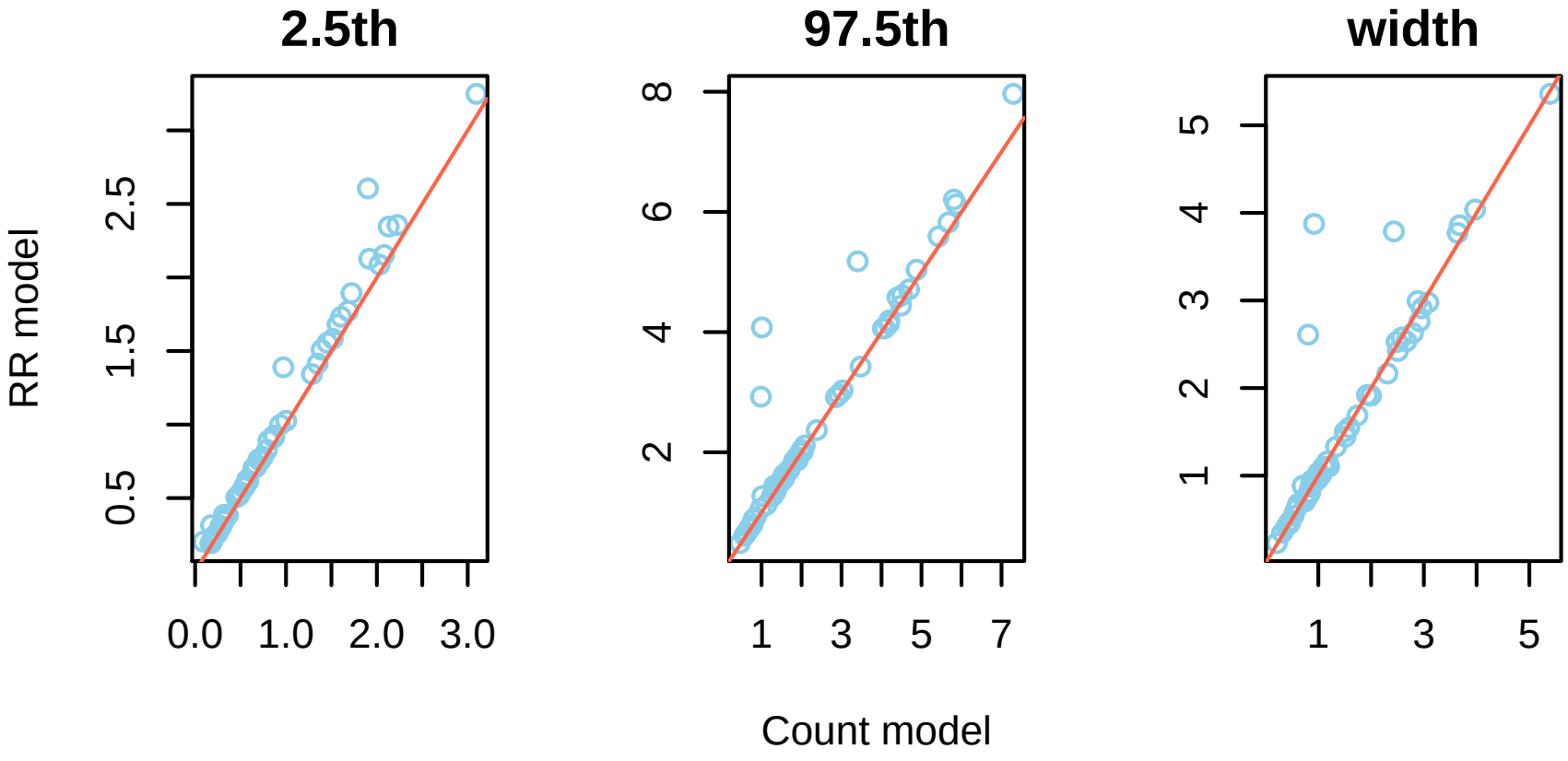
Comparison of interval widths

We calculate the 2.5% and 97.5% quantiles of the posteriors of the relative risks to compare with the count outcome model.

```
LN3lower <- exp(scotland.fit3$summary.linear.predictor["0.025qua
LN3upper <- exp(scotland.fit3$summary.linear.predictor["0.975qua
LN3width <- LN3upper[, 1] - LN3lower[, 1]

par(mfrow = c(1, 3), mar = c(5, 4, 2, 1))
plot(LN3lower[, 1] ~ LN2lower[, 1], col = "skyblue",
    ylab = "RR model", xlab = "", main = "2.5th")
abline(0, 1, col = "tomato")
plot(LN3upper[, 1] ~ LN2upper[, 1], col = "skyblue",
    ylab = "", xlab = "Count model", main = "97.5th")
abline(0, 1, col = "tomato")
plot(LN3width ~ LN2width, col = "skyblue", ylab = "",
    xlab = "", main = "width")
abline(0, 1, col = "tomato")
```

Comparison of interval widths



Missing Area Data

Missing area data in Scotland

As an illustration we suppose that for the last area Y_{56} is unobserved – it is coded as `NA` (its value is zero in the data).

The missing value can be imputed with the spatial ICAR model helping in this respect.

If the count was missing because low (e.g., not released because less than 5) then this is informative and the following analysis is not approprtiate.

```
Scotland$CountsNA <- Scotland$Counts
Scotland$CountsNA[56] <- NA
scotland.fitNA <- inla(CountsNA ~ 1 + I(X) + f(Region,
    model = "bym2", graph = "examples/scotland.graph",
    scale.model = T, constr = T, hyper = list(phi = list(prior =
        param = c(0.5, 0.5), initial = 1), prec = list(prior = "
        param = c(0.3/0.31, 0.01), initial = 5))),
    data = Scotland, family = "poisson", E = E, control.predicto
        link = 1))
```

Missing areas

```
summary(scotland.fitNA)
##
## Call:
##    c("inla(formula = CountsNA ~ 1 + I(X) + f(Region, model = \"bym2\", ",
##    " graph = \"examples/scotland.graph\", scale.model = T, constr = T, ",
##    " hyper = list(phi = list(prior = \"pc\", param = c(0.5, 0.5), ", "
##    initial = 1), prec = list(prior = \"pc.prec\", param = c(0.3/0.31, ", "
##    0.01), initial = 5))), family = \"poisson\", data = Scotland, ", " E =
##    E, control.predictor = list(compute = TRUE, link = 1))" )
## Time used:
##     Pre = 14.4, Running = 1.2, Post = 0.111, Total = 15.7
## Fixed effects:
##               mean    sd 0.025quant 0.5quant 0.975quant   mode kld
## (Intercept) -0.083 0.110     -0.301   -0.083      0.133 -0.082   0
## I(X)         0.024 0.011      0.002    0.025      0.047  0.025   0
##
## Random effects:
##   Name     Model
##     Region BYM2 model
##
## Model hyperparameters:
##                          mean    sd 0.025quant 0.5quant 0.975quant  mode
## Precision for Region 4.985 1.487       2.69    4.779      8.479 4.392
## Phi for Region       0.946 0.065       0.76    0.969      0.999 0.999
##
## Expected number of effective parameters(stdev): 27.05(3.02)
## Number of equivalent replicates : 2.03
##
## Marginal log-Likelihood:  -126.82
## Posterior marginals for the linear predictor and
##  the fitted values are computed
```

Missing areas

From the graph file we see that area 56 has areas 2,3,4,5 as neighbors – we look at these values and see the SMRs are high, which explains why the predictive mean is high.

We include the prediction of the rate from the model in which the data are observed and see it is much lower.

We look at the last line of the graph file (for area 56):

```
56 4 2 3 4 5
```

We examine the SMRs from the 4 neighboring areas, and they are high.

The covariate value for area 56 (which is also used in the prediction) is mid-range.

Missing area prediction of mean response

```
scotland.fitNA$summary.fitted.values[56, ]
##                            mean       sd 0.025quant 0.5quant 0.975quant     mode
## fitted.Predictor.56 3.268757 1.258445   1.489198 3.048772   6.336699 2.687021
Scotland$E[56]
## [1] 1.8
Scotland$X[56]
## [1] 10
set56 <- c(2, 3, 4, 5)
Scotland$Counts[set56]
## [1] 39 11  9 15
Scotland$E[set56]
## [1] 8.7 3.0 2.5 4.3
Scotland$Counts[set56]/Scotland$E[set56]
## [1] 4.482759 3.666667 3.600000 3.488372
# Compare with fit from model in which Y[56]=0 is
# used as observed
scotland.fit2$summary.fitted.values[56, ]
##                            mean       sd 0.025quant 0.5quant 0.975quant     mode
## fitted.Predictor.56 1.985036 0.6242881   0.971217 1.914376   3.404683 1.78178
```

Much lower predictor because the zero pulls down.

Missing area prediction of mean response

In the model

$$Y_i|\theta_i \sim Poisson(E_i\theta_i),$$

the fitted values/predictions are for θ_i, so no expected number and no Poisson sampling (so we're predicting the the relative risk).

We confirm that the quantiles of the fitted values are the exponentiated predictions.

```
scotland.fitNA$summary.fitted.values[56, c(3:5)]
##                       0.025quant 0.5quant 0.975quant
## fitted.Predictor.56   1.489198 3.048772   6.336699
exp(scotland.fitNA$summary.linear.predictor[56, c(3:5)])
##                 0.025quant 0.5quant 0.975quant
## Predictor.56   1.489198 3.048772   6.336699
```

Creating an INLA Graph File from a Shapefile

INLA Graph File

The code below creates a neighborhood filefor INLA that looks like:

39

1 4 11 13 22 38 2 2 12 38 3 5 11 13 20 36 39 4 6 9 17 19 24 29 31

...

38 7 1 2 7 11 12 22 32

39 8 3 13 17 19 20 21 27 30

Creating an INLA graph file from a shapefile

```
library(rgdal)  # for readOGR
library(spdep)  # for poly2nb and nb2inla
countymap = readOGR(dsn = "examples/wacounty.shp",
    layer = "wacounty")
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/austin/Dropbox/555-Materials/555-2020/5
## with 39 features
## It has 6 fields
nb.map <- poly2nb(countymap)
nb2INLA("wacounty.graph", nb.map)
```

PC Priors

PC prior details (only for those)

For a precision in the model $x|\tau \sim N(0, 1/\tau)$, the PC prior is obtained via the following rationale:

- The prior on the sd is exponential with rate λ, which we need to specify
- The exponential leads to a type-2 Gumbel on the precision (change of variables)
- Hence we have the model:

$$\begin{aligned} x|\tau &\sim N(0, 1/\tau) \\ \tau &\sim \text{Gumbel}(\lambda) \end{aligned}$$

- If we integrate out τ, we can find the marginal sd of x
- For more details see Simpson et al (2017, p. 9, top of right column) and Bakka et al (2018).

PC prior details: Code to scale

```
## simulate iid's with the PC prior for the
## precision
riid = function(n, u, alpha) {
    ## CDF for the prec (using the PC prior) is
    ## exp(-lambda/sqrt(prec))
    lambda = -log(alpha)/u
    ## first sample the precisions
    prec = (lambda/log(runif(n)))^2
    x = rnorm(n, mean = 0, sd = 1/sqrt(prec))
    return(x)
}
##
u.s = seq(0.1, 10, len = 1000)
s.s = numeric(length(u.s))
TAIL <- 0.5
for (k in seq_along(u.s)) {
    x = riid(1e+05, u.s[k], TAIL)
    s.s[k] = sd(x)
}
# plot(u.s,s.s)
coef(lm(s.s ~ -1 + u.s))
##      u.s
## 2.040486
```