An introduction to spatial and spatio-temporal modelling of small area disease rates

Professor Trevor Bailey

School of Engineering, Computing and Mathematics

University of Exeter, UK

Campinas, 8-10 April, 2008

Acknowledgements for data, WinBUGS code, and other material to:

N. Best *et al* (Imperial College, UK) and to M. Carvalho *et al*

(Fundação Oswaldo Cruz, Brazil)

My objectives in this course are to:

My objectives in this course are to:

- Describe in **outline** a **selection** of the statistical modelling methods currently employed in analysing **small area disease rates** (including a basic introduction to Bayesian modelling and 'Markov Chain Monte Carlo' (MCMC) methods)

My objectives in this course are to:

- Describe in **outline** a **selection** of the statistical modelling methods currently employed in analysing **small area disease rates** (including a basic introduction to Bayesian modelling and 'Markov Chain Monte Carlo' (MCMC) methods)
- Discuss **software** available to implement these modelling methods using a number of illustrative case studies (including a short basic introduction to the use of the public domain **WinBUGS** package for Bayesian modelling and links between WinBUGS and the public domain statistical language **R**)

My objectives in this course are to:

- Describe in **outline** a **selection** of the statistical modelling methods currently employed in analysing **small area disease rates** (including a basic introduction to Bayesian modelling and 'Markov Chain Monte Carlo' (MCMC) methods)
- Discuss **software** available to implement these modelling methods using a number of illustrative case studies (including a short basic introduction to the use of the public domain **WinBUGS** package for Bayesian modelling and links between WinBUGS and the public domain statistical language **R**)
- Briefly discuss some **spatio-temporal extensions** to these models

My objectives in this course are to:

- Describe in **outline** a **selection** of the statistical modelling methods currently employed in analysing **small area disease rates** (including a basic introduction to Bayesian modelling and 'Markov Chain Monte Carlo' (MCMC) methods)
- Discuss **software** available to implement these modelling methods using a number of illustrative case studies (including a short basic introduction to the use of the public domain **WinBUGS** package for Bayesian modelling and links between WinBUGS and the public domain statistical language **R**)
- Briefly discuss some **spatio-temporal extensions** to these models
- Provide **references** to more details about these methods and to extensions and additional approaches

- Before I embark on that agenda **first note** that my focus is only on **modelling disease data in the form of small area disease rates**.

The context of this course

- Before I embark on that agenda **first note** that my focus is only on **modelling disease data in the form of small area disease rates**.
  - i.e. data on health events and related measures aggregated or averaged to the level of census tracts or other kinds of administrative districts (e.g. counts of cases and population at risk grouped in areas along with corresponding socio-economic descriptors and exposure measures for these groups)

The context of this course

- Before I embark on that agenda **first note** that my focus is only on **modelling disease data in the form of small area disease rates**.
  - i.e. data on health events and related measures aggregated or averaged to the level of census tracts or other kinds of administrative districts (e.g. counts of cases and population at risk grouped in areas along with corresponding socio-economic descriptors and exposure measures for these groups)
- I am **not going** to discuss **modelling methods appropriate to case event data**

The context of this course

- Before I embark on that agenda **first note** that my focus is only on **modelling disease data in the form of small area disease rates**.
  - i.e. data on health events and related measures aggregated or averaged to the level of census tracts or other kinds of administrative districts (e.g. counts of cases and population at risk grouped in areas along with corresponding socio-economic descriptors and exposure measures for these groups)
- I am **not going** to discuss **modelling methods appropriate to case event data**
  - i.e. data involving locations (usually residential) and covariate measures on individual cases of a disease and of individual members of a suitable control group

The context of this course

- Area-level data on disease incidence remains more readily available than case-event data. However, the latter is now increasingly becoming accessible and many studies in spatial epidemiology may now be involved with either of these data types and sometimes a mixture of both.

The context of this course

- Area-level data on disease incidence remains more readily available than case-event data. However, the latter is now increasingly becoming accessible and many studies in spatial epidemiology may now be involved with either of these data types and sometimes a mixture of both.
- However, there are important distinctions between the statistical modelling of area data as opposed to that on individual cases and throughout this course we confine ourselves **solely to models for data at a group level within geographical areas**

The context of this course

- Area-level data on disease incidence remains more readily available than case-event data. However, the latter is now increasingly becoming accessible and many studies in spatial epidemiology may now be involved with either of these data types and sometimes a mixture of both.
- However, there are important distinctions between the statistical modelling of area data as opposed to that on individual cases and throughout this course we confine ourselves **solely to models for data at a group level within geographical areas**
- In doing so we must of course remain aware of the problems involved in examining associations between disease incidence and risk factors measured on groups (the so-called **ecological fallacy**)—we take this as 'gospel' throughout the course

- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)

- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:

- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:
  - An understanding of the relevant **Epidemiology**

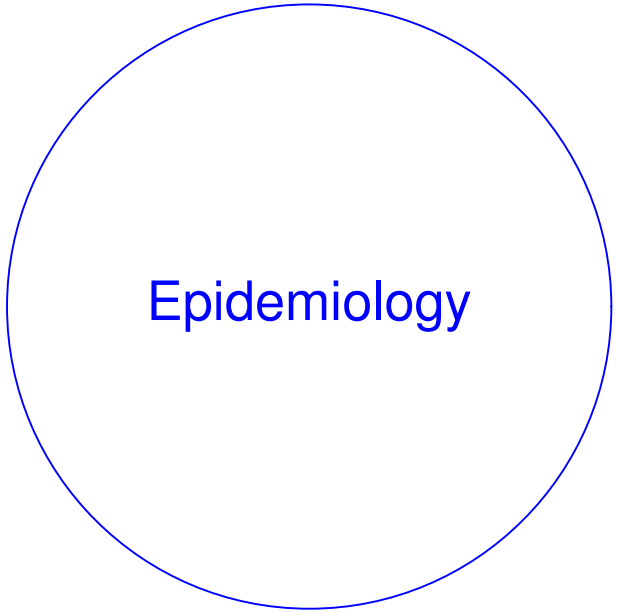
- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:
  - An understanding of the relevant **Epidemiology**
  - Access to appropriate **Data** (good survey design included)

- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:
  - An understanding of the relevant **Epidemiology**
  - Access to appropriate **Data** (good survey design included)
  - Suitable **Analytical Tools** — this includes not only statistical models and methodology, but also 'geoprocessing' more widely (computer software, GIS, algorithms etc.)

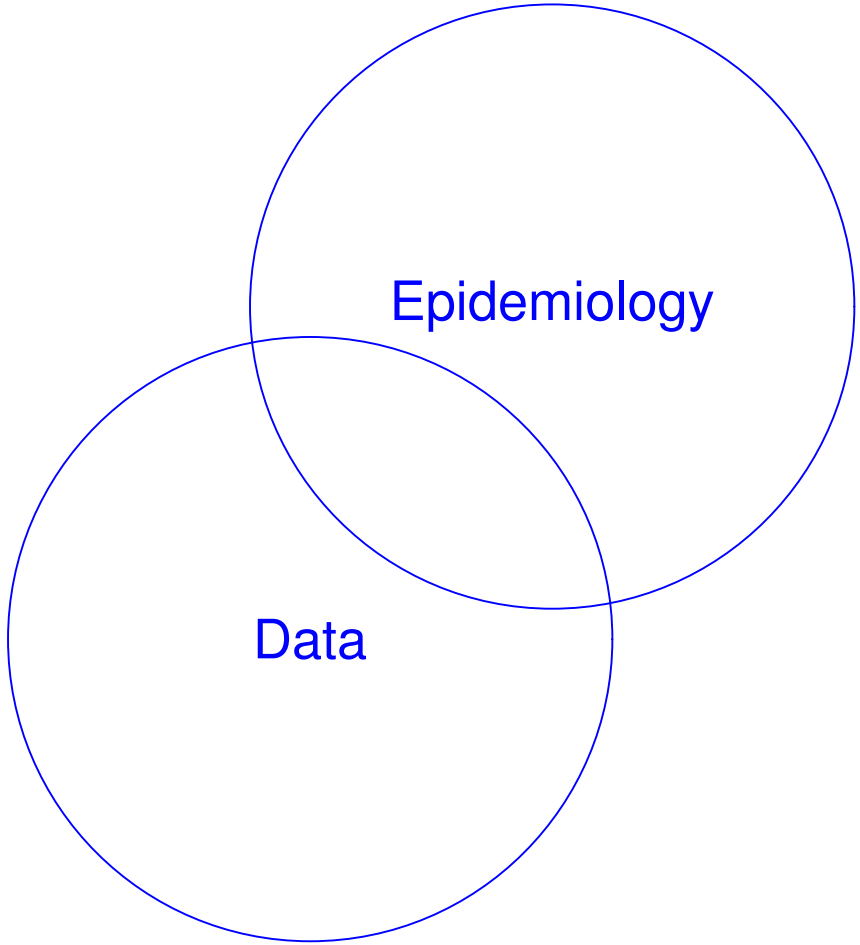
- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:
  - An understanding of the relevant **Epidemiology**
  - Access to appropriate **Data** (good survey design included)
  - Suitable **Analytical Tools** — this includes not only statistical models and methodology, but also 'geoprocessing' more widely (computer software, GIS, algorithms etc.)
- So let me state from the outset that I appreciate that this course is about only part of an overall process

- **Second note** that my **focus will be on statistical methods** and I fully acknowledge spatial epidemiology involves **much more than statistical models** (however sophisticated)
- It is a blend of essentially three factors:
  - An understanding of the relevant **Epidemiology**
  - Access to appropriate **Data** (good survey design included)
  - Suitable **Analytical Tools** — this includes not only statistical models and methodology, but also 'geoprocessing' more widely (computer software, GIS, algorithms etc.)
- So let me state from the outset that I appreciate that this course is about only part of an overall process —**statistics is easy, spatial epidemiology is hard.**

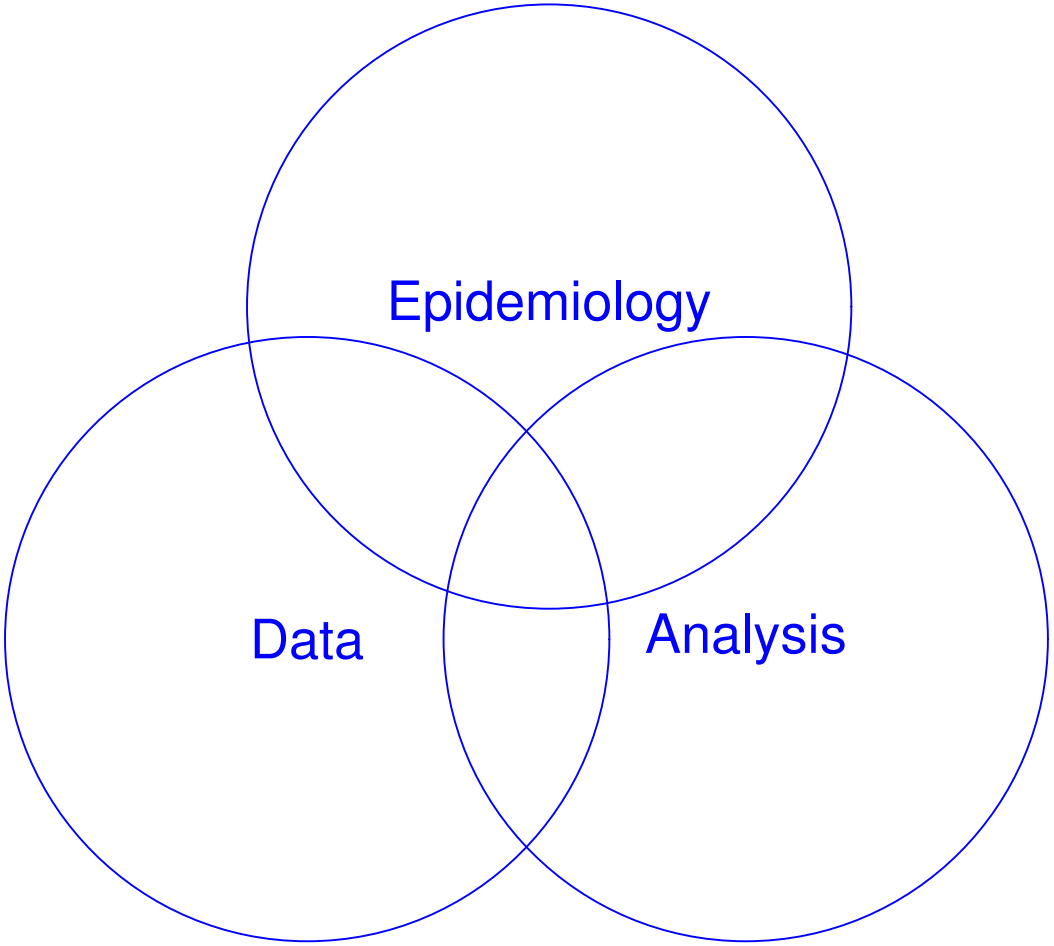
Components of Spatial Epidemiology



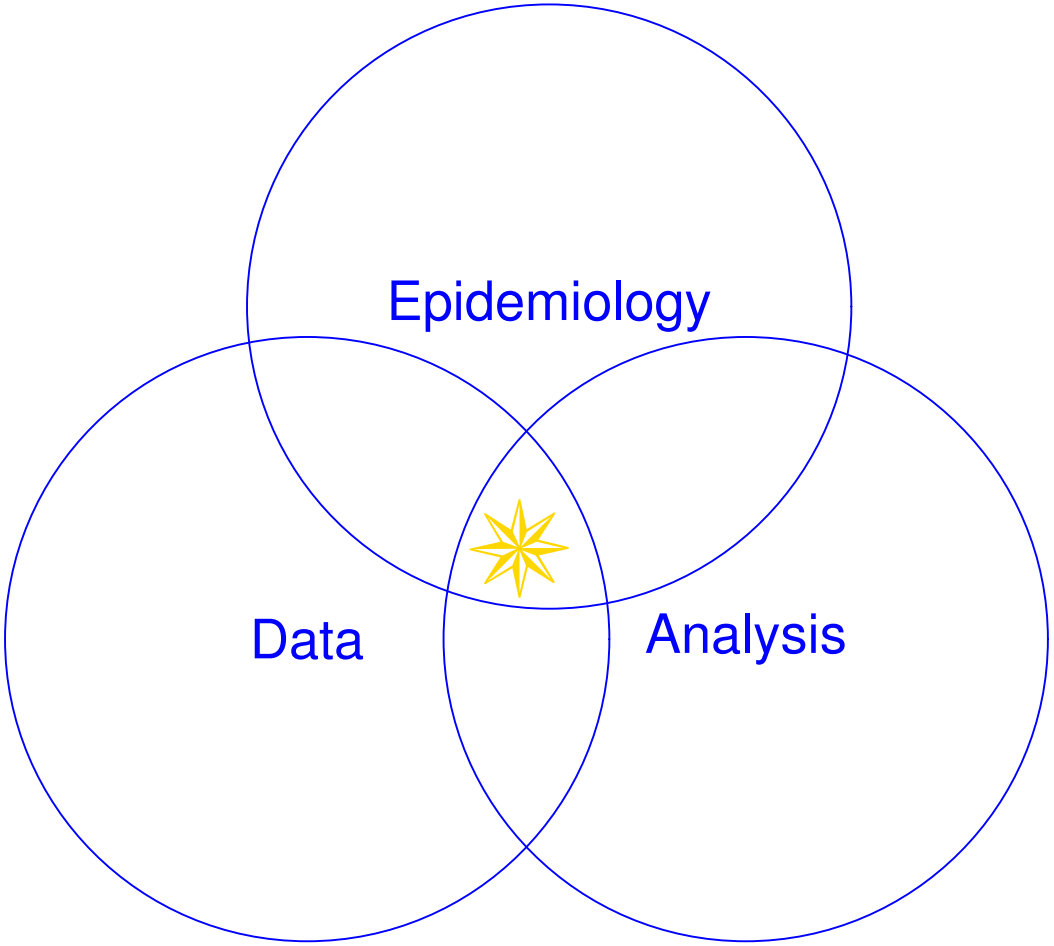
Components of Spatial Epidemiology



Components of Spatial Epidemiology



Components of Spatial Epidemiology



Finally note that we shall focus on two key areas in which statistical models have proved useful in the analysis of area-level data on disease incidence:

Finally note that we shall focus on two key areas in which statistical models have proved useful in the analysis of area-level data on disease incidence:

- **Disease Mapping** — Describing the underlying geographical distribution and pattern in a disease as part of general health surveillance (epidemics, increasing rates, preliminary identification of unusual elevations in risk, etc.)

Finally note that we shall focus on two key areas in which statistical models have proved useful in the analysis of area-level data on disease incidence:

- **Disease Mapping** — Describing the underlying geographical distribution and pattern in a disease as part of general health surveillance (epidemics, increasing rates, preliminary identification of unusual elevations in risk, etc.)
- **Ecological (or correlation) Studies** — investigation of the relationships between disease incidence and corresponding group risk factors i.e. covariates (so as to correct for confounding factors, or establish hypotheses to target further research or prevention)

Finally note that we shall focus on two key areas in which statistical models have proved useful in the analysis of area-level data on disease incidence:

- **Disease Mapping** — Describing the underlying geographical distribution and pattern in a disease as part of general health surveillance (epidemics, increasing rates, preliminary identification of unusual elevations in risk, etc.)
- **Ecological (or correlation) Studies** — investigation of the relationships between disease incidence and corresponding group risk factors i.e. covariates (so as to correct for confounding factors, or establish hypotheses to target further research or prevention)

Division is convenient, but blurred in practice — disease mapping commonly involves relationships with known risk factors for the disease and ecological models often incorporate spatial and/or temporal 'smoothing' effects employed in disease mapping

- We **will not** consider methods (many) and models (fewer) which have been explicitly designed to address a third key area in the analysis of area-level disease incidence—**disease clustering**

- We **will not** consider methods (many) and models (fewer) which have been explicitly designed to address a third key area in the analysis of area-level disease incidence—**disease clustering**
  - i.e. testing for **significant** space or/and time elevations in the risk of disease (either unfocussed or focussed) around known suspected (putative) hazards

- We **will not** consider methods (many) and models (fewer) which have been explicitly designed to address a third key area in the analysis of area-level disease incidence—**disease clustering**
  - i.e. testing for **significant** space or/and time elevations in the risk of disease (either unfocussed or focussed) around known suspected (putative) hazards
- There are a whole range of specialised hypothesis tests and related techniques in this area and references are provided to follow these up

The context of this course

- We **will not** consider methods (many) and models (fewer) which have been explicitly designed to address a third key area in the analysis of area-level disease incidence—**disease clustering**
  - i.e. testing for **significant** space or/and time elevations in the risk of disease (either unfocussed or focussed) around known suspected (putative) hazards
- There are a whole range of specialised hypothesis tests and related techniques in this area and references are provided to follow these up
- However note that the two areas we intend to cover do indirectly relate to disease clustering — good disease incidence maps often play an important preliminary role in such studies and putative hazards are now sometimes usefully viewed as particular kinds of covariate in models which are similar to those used in correlation studies

Some illustrative examples

Having set the scene for the course, it may be useful to briefly introduce three case studies that we can use throughout to illustrate the methods discussed. These concern:

Some illustrative examples

Having set the scene for the course, it may be useful to briefly introduce three case studies that we can use throughout to illustrate the methods discussed. These concern:

- Leprosy surveillance in the period 1991-1995 in Olinda, N.E. Brazil.

Some illustrative examples

Having set the scene for the course, it may be useful to briefly introduce three case studies that we can use throughout to illustrate the methods discussed. These concern:

- Leprosy surveillance in the period 1991-1995 in Olinda, N.E. Brazil.
- Incidence of larynx cancer diagnosed during 1982-1991, in the Mersey and West Lancashire districts of N.W England.

Some illustrative examples

Having set the scene for the course, it may be useful to briefly introduce three case studies that we can use throughout to illustrate the methods discussed. These concern:

- Leprosy surveillance in the period 1991-1995 in Olinda, N.E. Brazil.
- Incidence of larynx cancer diagnosed during 1982-1991, in the Mersey and West Lancashire districts of N.W England.
- Leptospirosis incidence in the city of Rio de Janeiro, Brazil in the period 1997-2002.

Leprosy surveillance in Olinda

- Olinda (as you know!) is a municipality of Pernambuco State in N.E. Brazil comprising (in 1991 census) 241 census tracts with approx 350,000 inhabitants.

Leprosy surveillance in Olinda

- Olinda (as you know!) is a municipality of Pernambuco State in N.E. Brazil comprising (in 1991 census) 241 census tracts with approx 350,000 inhabitants.
- Data involved are the incidence of new leprosy cases by census tract over the period 1991-1995 (1135 cases in total) along with corresponding mid-period (1993) population estimates in these tracts.

Leprosy surveillance in Olinda

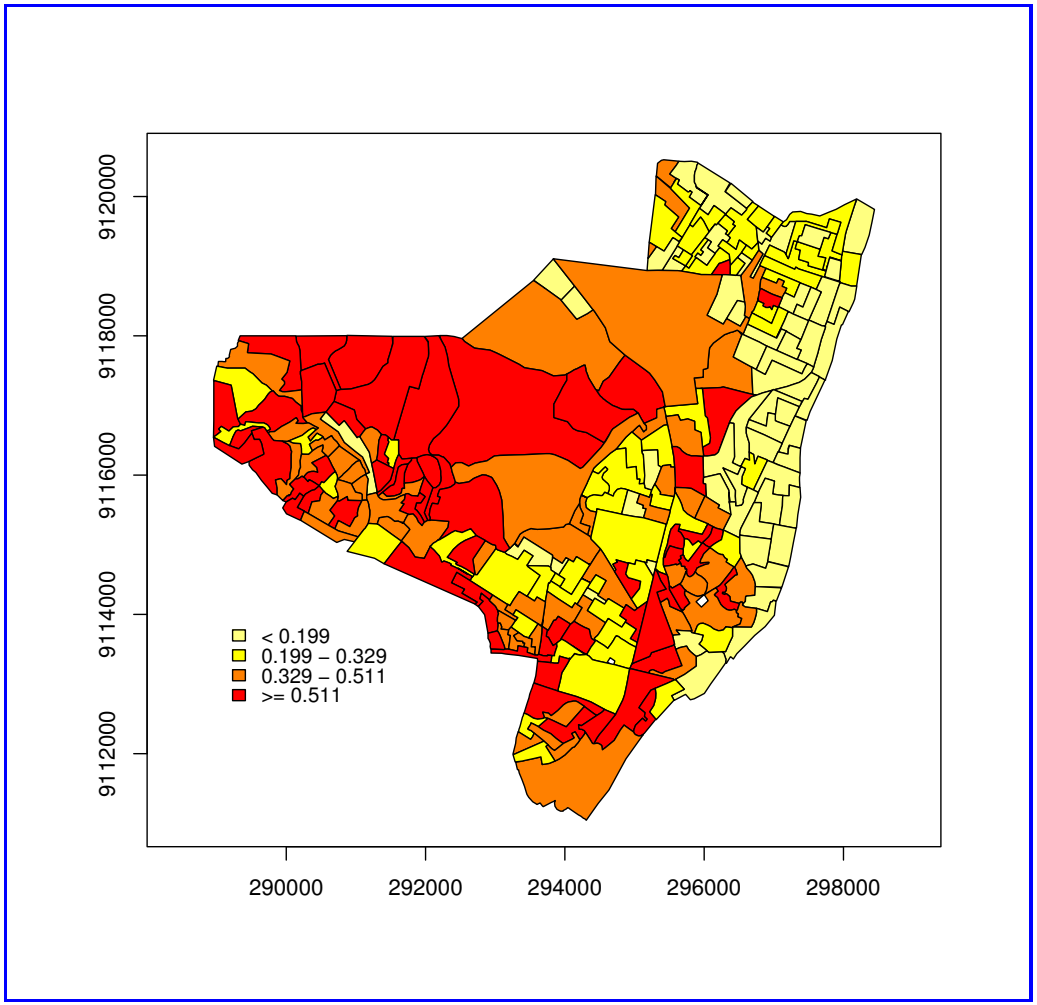
- Olinda (as you know!) is a municipality of Pernambuco State in N.E. Brazil comprising (in 1991 census) 241 census tracts with approx 350,000 inhabitants.
- Data involved are the incidence of new leprosy cases by census tract over the period 1991-1995 (1135 cases in total) along with corresponding mid-period (1993) population estimates in these tracts.
- A simple indicator of deprivation is also available — the proportion of heads of household with monthly income below one minimum legal wage (approximately US$80).

Leprosy surveillance in Olinda

- Olinda (as you know!) is a municipality of Pernambuco State in N.E. Brazil comprising (in 1991 census) 241 census tracts with approx 350,000 inhabitants.
- Data involved are the incidence of new leprosy cases by census tract over the period 1991-1995 (1135 cases in total) along with corresponding mid-period (1993) population estimates in these tracts.
- A simple indicator of deprivation is also available — the proportion of heads of household with monthly income below one minimum legal wage (approximately US$80). Census tracts where this indicator is extremely high (in excess of 60%) are primarily 'favelas' where accurate case detection is notoriously difficult.

Pernambuco State, Brazil

% households in Olinda (1993) with monthly income $<$ minimum legal wage ($\approx$ US$80)



Larynx cancer incidence in Mersey and West Lancashire

- Data consist of 876 cases of larynx cancer diagnosed during 1982-1991, in 144 electoral wards of Mersey and West Lancashire in NW England.

Larynx cancer incidence in Mersey and West Lancashire

- Data consist of 876 cases of larynx cancer diagnosed during 1982-1991, in 144 electoral wards of Mersey and West Lancashire in NW England.
- Also available are the expected numbers of cases in these wards calculated using external standardisation based on national age- and sex-specific reference rates and population counts from the 1991 census.

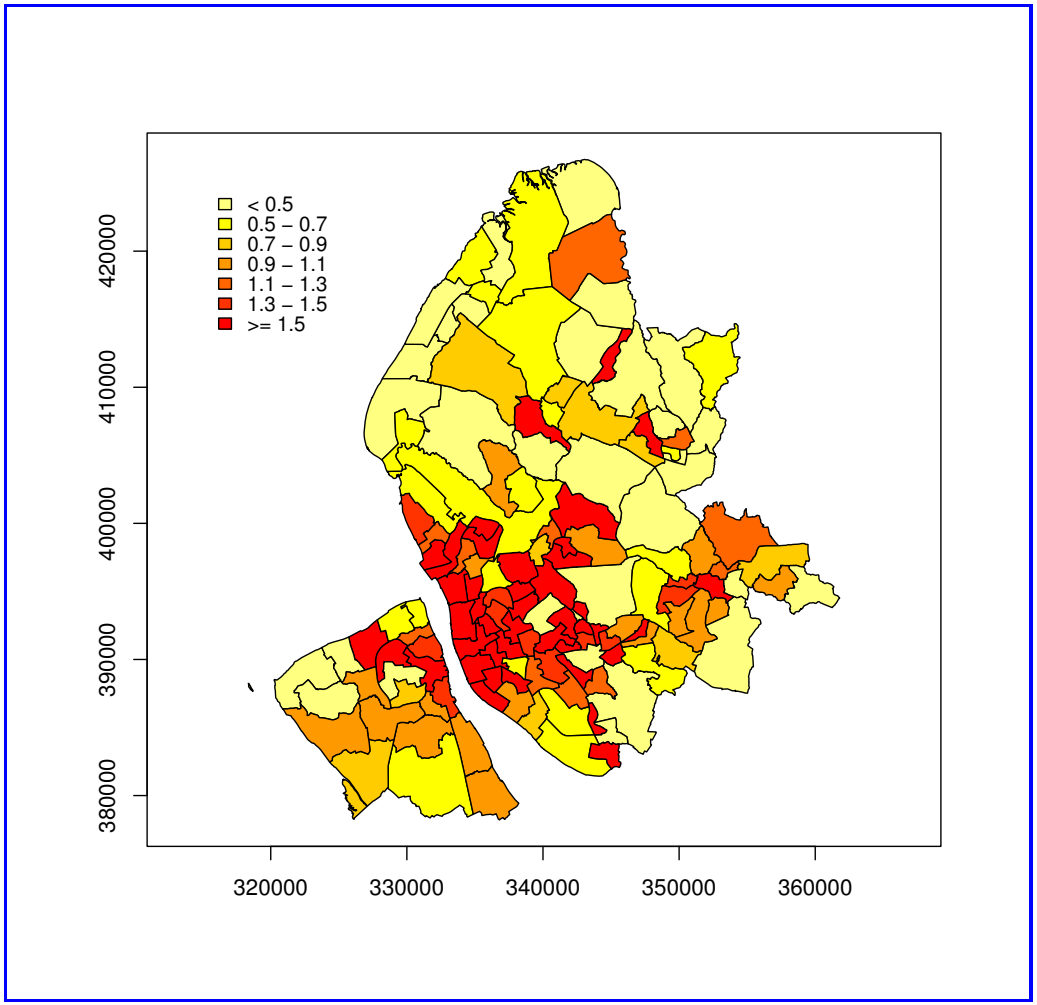
Larynx cancer incidence in Mersey and West Lancashire

- Data consist of 876 cases of larynx cancer diagnosed during 1982-1991, in 144 electoral wards of Mersey and West Lancashire in NW England.
- Also available are the expected numbers of cases in these wards calculated using external standardisation based on national age- and sex-specific reference rates and population counts from the 1991 census.
- In addition results of a survey in the Mersey and West Lancashire region which included questions on smoking habits have been used to derive an area-level smoking index for each district (‘low’, ‘medium’ or ‘high’ — i.e. a predominance of non-smokers, moderate smokers or heavy smokers)

Larynx cancer incidence in Mersey and West Lancashire

- Data consist of 876 cases of larynx cancer diagnosed during 1982-1991, in 144 electoral wards of Mersey and West Lancashire in NW England.
- Also available are the expected numbers of cases in these wards calculated using external standardisation based on national age- and sex-specific reference rates and population counts from the 1991 census.
- In addition results of a survey in the Mersey and West Lancashire region which included questions on smoking habits have been used to derive an area-level smoking index for each district ('low', 'medium' or 'high' — i.e. a predominance of non-smokers, moderate smokers or heavy smokers)
- Finally, a measure of air pollution is available in the form of annual mean levels of particulates in each district estimated from a dispersion model based upon traffic flow.

SMRs of Larynx Cancer in in Mersey and West Lancashire 1982-1991



Leptospirosis incidence in Rio de Janeiro

- Data comprise diagnosed cases of Leptospirosis by month and by year for the period 1997-2002 (total of 367 cases) in 157 districts (Bairro) of the city of Rio de Janeiro along with corresponding populations (2001 census)

Leptospirosis incidence in Rio de Janeiro

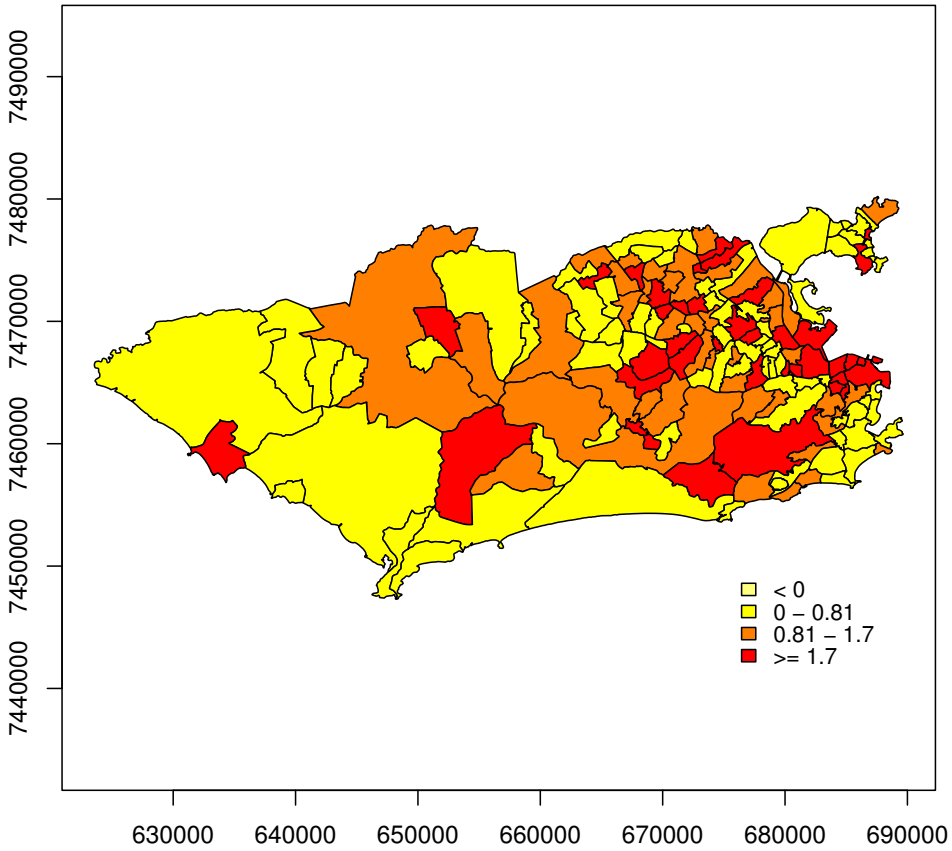
- Data comprise diagnosed cases of Leptospirosis by month and by year for the period 1997-2002 (total of 367 cases) in 157 districts (Bairro) of the city of Rio de Janeiro along with corresponding populations (2001 census)
- Information is also available on social deprivation in the districts including the proportion of the population living in favelas, mean numbers of persons per house and proportion of families with income of less than one minimum wage.

Leptospirosis incidence in Rio de Janeiro

- Data comprise diagnosed cases of Leptospirosis by month and by year for the period 1997-2002 (total of 367 cases) in 157 districts (Bairro) of the city of Rio de Janeiro along with corresponding populations (2001 census)
- Information is also available on social deprivation in the districts including the proportion of the population living in favelas, mean numbers of persons per house and proportion of families with income of less than one minimum wage.
- Mean and maximum annual rainfall in the years 1997-2002 from 32 weather stations dotted across the city provides some indication of the risk of floods in each district.

Satellite Image of Rio de Janeiro

Leptospirosis rates per 100,000 population 1997-2002



Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings using the three illustrative applications where appropriate and introducing computing ideas 'as we go':
  - Preliminaries — a tour of statistical/Bayesian modelling
  - Disease mapping
  - Ecological (correlation) studies
  - Further topics in ecological studies
  - Space-time models
  - Concluding remarks

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings using the three illustrative applications where appropriate and introducing computing ideas 'as we go':
  - Preliminaries — a tour of statistical/Bayesian modelling
  - Disease mapping
  - Ecological (correlation) studies
  - Further topics in ecological studies
  - Space-time models
  - Concluding remarks
- Time prevents much mathematical detail or anything like an exhaustive coverage. List of selected references provided to help.

Preliminaries — A tour of statistical modelling

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

  Data $\Rightarrow$ **Trend**+**Error**

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

  Data $\Rightarrow$ **Trend**+**Error** $\Rightarrow$ **model** $\Rightarrow$ **Probability distribution for the data**

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

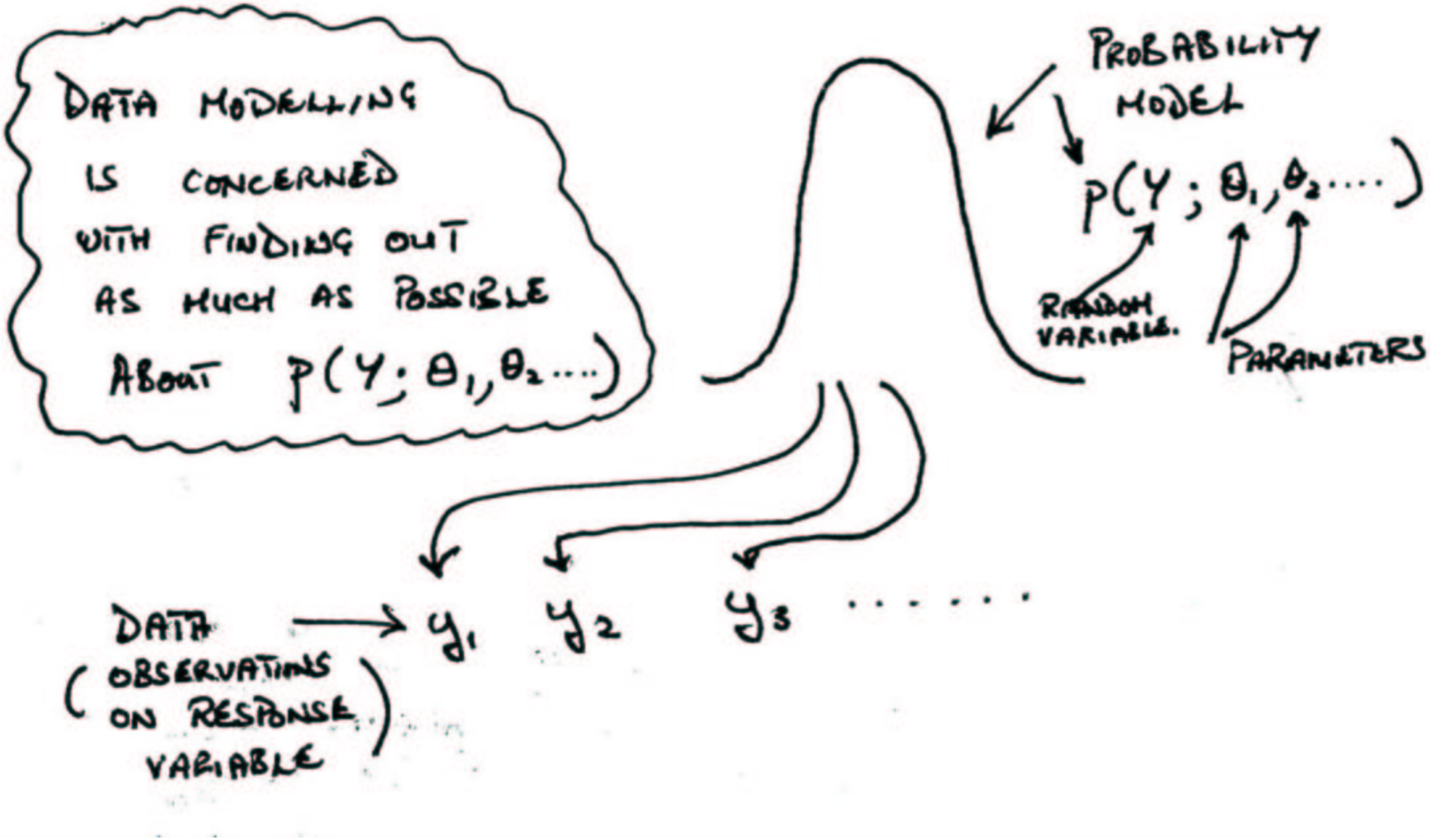
  Data $\Rightarrow$ **Trend+Error** $\Rightarrow$ **model** $\Rightarrow$ **Probability distribution for the data**
- So a statistical model for data $\boldsymbol{y} = (y_1, \ldots, y_n)$ consists of a **joint probability distribution** for these data $P(\boldsymbol{y})$

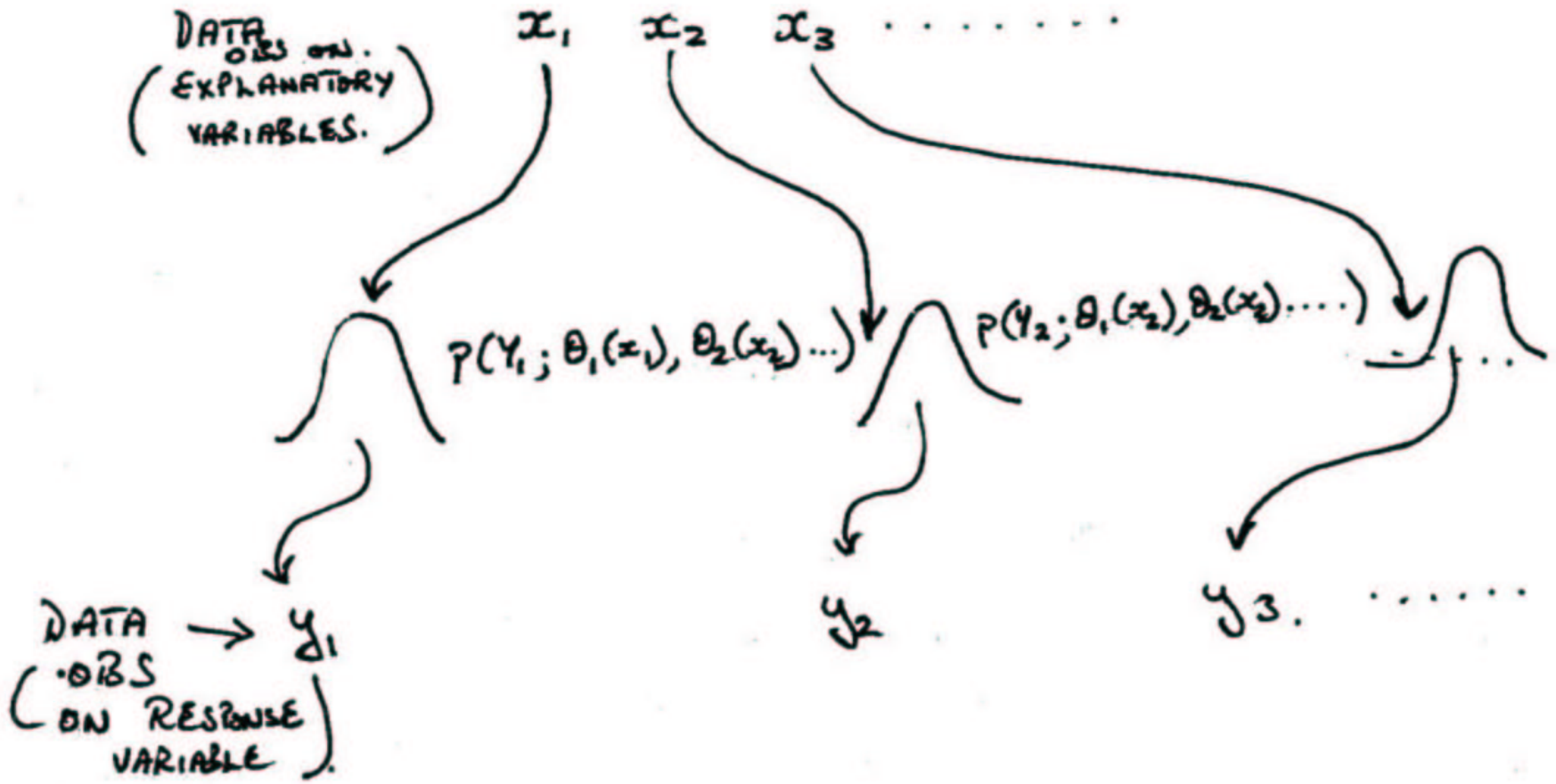
- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

  Data $\Rightarrow$ **Trend+Error** $\Rightarrow$ **model** $\Rightarrow$ **Probability distribution for the data**
- So a statistical model for data $\boldsymbol{y} = (y_1, \ldots, y_n)$ consists of a **joint probability distribution** for these data $P(\boldsymbol{y})$
- In practice the general **form** of the model will involve unknown **parameters** $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. So it is actually a joint probability distribution for $\boldsymbol{y}$ which depends on values for the parameters—$P(\boldsymbol{y}; \boldsymbol{\theta})$. When viewed as a function of $\boldsymbol{\theta}$ rather than $\boldsymbol{y}$ $P(\boldsymbol{y}; \boldsymbol{\theta})$ is known as the **likelihood** for the data

- Much of this course is about Bayesian spatial statistical models. Therefore start by reviewing some statistical background to set the stage for the these models.
- Much of modern statistics (perhaps all of it!) is about **modelling** data

  Data $\Rightarrow$ **Trend+Error** $\Rightarrow$ **model** $\Rightarrow$ **Probability distribution for the data**
- So a statistical model for data $\boldsymbol{y} = (y_1, \ldots, y_n)$ consists of a **joint probability distribution** for these data $P(\boldsymbol{y})$
- In practice the general **form** of the model will involve unknown **parameters** $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. So it is actually a joint probability distribution for $\boldsymbol{y}$ which depends on values for the parameters—$P(\boldsymbol{y}; \boldsymbol{\theta})$. When viewed as a function of $\boldsymbol{\theta}$ rather than $\boldsymbol{y}$ $P(\boldsymbol{y}; \boldsymbol{\theta})$ is known as the **likelihood** for the data
- Assuming the model form is well chosen (*the 'art' of statistics*) then the focus in statistical modelling is to obtain good estimates for values of the associated parameters (*the 'science' of statistics*)


DATA MODELLING
IS CONCERNED
WITH FINDING OUT
AS MUCH AS POSSIBLE
ABOUT $p(Y; \theta_1, \theta_2 \ldots)$
PROBABILITY
MODEL
$p(Y; \theta_1, \theta_2 \ldots)$
RANDOM
VARIABLE.
PARAMETERS
DATA
(OBSERVATIONS
ON RESPONSE
VARIABLE)
y_1 y_2 y_3 . . . . .


DATA OBS ON. EXPLANATORY VARIABLES.
x_1 x_2 x_3 .....
$p(Y_1; \theta_1(x_1), \theta_2(x_2) \ldots)$
$p(Y_2; \theta_1(x_2), \theta_2(x_2) \ldots)$
DATA → y_1
OBS ON RESPONSE VARIABLE
y_2
y_3. .....

- The key general purpose conventional method used in estimating the parameters of a statistical model is **maximum likelihood**

- The key general purpose conventional method used in estimating the parameters of a statistical model is **maximum likelihood**
- The joint probability distribution $P(\boldsymbol{y};\boldsymbol{\theta})$ when viewed as a function of $\boldsymbol{\theta}$ is known as the **likelihood** because under the proposed form of the model it essentially represents *'how likely it is that you would get the particular data $\boldsymbol{y}$ that you observed, given specific values for the parameters $\boldsymbol{\theta}$'* (although note it is **not a probability distribution for $\boldsymbol{\theta}$**)

- The key general purpose conventional method used in estimating the parameters of a statistical model is **maximum likelihood**
- The joint probability distribution $P(\boldsymbol{y};\boldsymbol{\theta})$ when viewed as a function of $\boldsymbol{\theta}$ is known as the **likelihood** because under the proposed form of the model it essentially represents *'how likely it is that you would get the particular data $\boldsymbol{y}$ that you observed, given specific values for the parameters $\boldsymbol{\theta}$'* (although note it is **not a probability distribution for $\boldsymbol{\theta}$**)
- One way to find 'good' estimates for $\boldsymbol{\theta}$ is to choose values $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ that maximise the likelihood $P(\boldsymbol{y};\boldsymbol{\theta})$ — i.e. use **maximum likelihood estimates** (mle)

- The key general purpose conventional method used in estimating the parameters of a statistical model is **maximum likelihood**
- The joint probability distribution $P(\boldsymbol{y};\boldsymbol{\theta})$ when viewed as a function of $\boldsymbol{\theta}$ is known as the **likelihood** because under the proposed form of the model it essentially represents *'how likely it is that you would get the particular data $\boldsymbol{y}$ that you observed, given specific values for the parameters $\boldsymbol{\theta}$'* (although note it is **not a probability distribution for $\boldsymbol{\theta}$**)
- One way to find 'good' estimates for $\boldsymbol{\theta}$ is to choose values $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ that maximise the likelihood $P(\boldsymbol{y};\boldsymbol{\theta})$ — i.e. use **maximum likelihood estimates** (mle)
- In general the accuracy of such estimates (i.e. their standard errors) may then be assessed by 'how peaked' the likelihood is at the maximum — i.e. by a function of the second derivative of the likelihood evaluated at the maximum ($\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$). Hypothesis tests may be performed by looking at likelihood ratios — ratio of maximised likelihood under null hypothesis to maximised likelihood without it.

- If all this sounds a bit abstract, then rest assured that most of you have been doing it for years!

- If all this sounds a bit abstract, then rest assured that most of you have been doing it for years!
  - The least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ of the intercept and slope parameters in the simple regression model $\mu = \alpha + \beta x$ under the assumption that $y \sim N(\mu, \sigma^2)$ (i.e. normally distributed errors) are in fact the maximum likelihood likelihood estimates of these parameters
  - The residual sum of squares is closely related to the value of the likelihood at the maximum
  - All the usual calculations for the standard errors of the regression coefficients, t-tests, F-tests and the like, are essentially equivalent to the same quantities derived from the general maximum likelihood approach.

- Let's take a simple example (non-spatial) to clarify these ideas

- Let's take a simple example (non-spatial) to clarify these ideas
- Suppose you want to know the incidence of a disease in the general population. You take a random sample of n individuals, test each for the disease, and let y be the number who test $+$ve.

- Let's take a simple example (non-spatial) to clarify these ideas
- Suppose you want to know the incidence of a disease in the general population. You take a random sample of n individuals, test each for the disease, and let y be the number who test $+$ve.
- Let θ be the probability an individual has the disease, then the simplest model (likelihood) for the data y is a binomial distribution: $P(y;\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$

- Let's take a simple example (non-spatial) to clarify these ideas
- Suppose you want to know the incidence of a disease in the general population. You take a random sample of n individuals, test each for the disease, and let y be the number who test $+$ve.
- Let θ be the probability an individual has the disease, then the simplest model (likelihood) for the data y is a binomial distribution: $P(y;\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
- The log likelihood is therefore proportional to $y\log(\theta) + (n-y)\log(1-\theta)$ and differentiating with respect to θ and setting this derivative equal to zero (for a maximum) gives:

$$\frac{y}{\hat{\theta}} = \frac{(n-y)}{(1-\hat{\theta})} \qquad \text{or} \qquad \hat{\theta} = \frac{y}{n}$$

- Let's take a simple example (non-spatial) to clarify these ideas
- Suppose you want to know the incidence of a disease in the general population. You take a random sample of n individuals, test each for the disease, and let y be the number who test $+$ve.
- Let θ be the probability an individual has the disease, then the simplest model (likelihood) for the data y is a binomial distribution: $P(y;\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
- The log likelihood is therefore proportional to $y\log(\theta) + (n-y)\log(1-\theta)$ and differentiating with respect to θ and setting this derivative equal to zero (for a maximum) gives:

$$\frac{y}{\hat{\theta}} = \frac{(n-y)}{(1-\hat{\theta})} \qquad \text{or} \qquad \hat{\theta} = \frac{y}{n}$$

- so the maximum likelihood estimate of theta is just the sample proportion who test $+$ve (as you would expect!!)

➢ A simple (non spatial) model for counts of cases of a rare disease (aren't they all!) in n areas also serves as another illustration of the maximum likelihood idea.

- A simple (non spatial) model for counts of cases of a rare disease (aren't they all!) in n areas also serves as another illustration of the maximum likelihood idea.
- The most basic model is that observed counts of cases $\boldsymbol{y} = (y_1, \ldots, y_n)$ are Poisson distributed $\left(P(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right)$ with means $\mu_i = e_i \rho_i$, where e_i is the *'expected'* number of cases (based on some global reference rate) and ρ_i is the *'relative disease risk'* for observation (area) i.

- A simple (non spatial) model for counts of cases of a rare disease (aren't they all!) in n areas also serves as another illustration of the maximum likelihood idea.
- The most basic model is that observed counts of cases $\boldsymbol{y} = (y_1, \ldots, y_n)$ are Poisson distributed $\left(P(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right)$ with means $\mu_i = e_i \rho_i$, where e_i is the *'expected'* number of cases (based on some global reference rate) and ρ_i is the *'relative disease risk'* for observation (area) i.
- The 'expected' cases are assumed *known* and taken as $e_i = r\pi_i$ where r is an known overall reference rate for the disease and π_i is the population at risk for each observation. Often this reference rate is stratified for known confounders, such as age and sex i.e. $e_i = \sum_j r_j \pi_{ij}$ (where j is age/sex etc. group)

- A simple (non spatial) model for counts of cases of a rare disease (aren't they all!) in n areas also serves as another illustration of the maximum likelihood idea.
- The most basic model is that observed counts of cases $\boldsymbol{y} = (y_1, \ldots, y_n)$ are Poisson distributed $\left(P(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right)$ with means $\mu_i = e_i \rho_i$, where e_i is the *'expected'* number of cases (based on some global reference rate) and ρ_i is the *'relative disease risk'* for observation (area) i.
- The 'expected' cases are assumed *known* and taken as $e_i = r\pi_i$ where r is an known overall reference rate for the disease and π_i is the population at risk for each observation. Often this reference rate is stratified for known confounders, such as age and sex i.e. $e_i = \sum_j r_j \pi_{ij}$ (where j is age/sex etc. group)
- So the model can be summarised as: $y_i \sim$ Poisson$\big(e_i \rho_i\big)$ where ρ_i is the relative disease risk for observation i compared to the chosen reference rate.

- Here the parameters of the model are $\boldsymbol{\theta} = (\rho_1, \ldots, \rho_n)$ and the likelihood for the data i.e. $P(\boldsymbol{y}; \boldsymbol{\theta})$ or $P(y_1, \ldots, y_n; \rho_1, \ldots, \rho_n)$ is therefore:

$$\prod_{i=1}^{n} \left[\frac{(e_i \rho_i)^{y_i}}{y_i!} \exp(-e_i \rho_i) \right]$$

So the log likelihood is: $\sum_{i=1}^{n} \left[y_i \left(\log e_i + \log \rho_i \right) - e_i \rho_i - \log(y_i!) \right]$

- Here the parameters of the model are $\boldsymbol{\theta} = (\rho_1, \ldots, \rho_n)$ and the likelihood for the data i.e. $P(\boldsymbol{y}; \boldsymbol{\theta})$ or $P(y_1, \ldots, y_n; \rho_1, \ldots, \rho_n)$ is therefore:

$$\prod_{i=1}^{n} \left[\frac{(e_i \rho_i)^{y_i}}{y_i!} \exp(-e_i \rho_i) \right]$$

  So the log likelihood is: $\sum_{i=1}^{n} \left[y_i \left(\log e_i + \log \rho_i \right) - e_i \rho_i - \log(y_i!) \right]$

- Differentiating wrt ρ_i and setting this derivative equal to zero (for a maximum) gives: $\frac{y_i}{\rho_i} - e_i = 0$ so the maximum likelihood estimate of ρ_i is: $\hat{\rho}_i = \frac{y_i}{e_i}$

- Here the parameters of the model are $\boldsymbol{\theta} = (\rho_1, \ldots, \rho_n)$ and the likelihood for the data i.e. $P(\boldsymbol{y}; \boldsymbol{\theta})$ or $P(y_1, \ldots, y_n; \rho_1, \ldots, \rho_n)$ is therefore:

$$\prod_{i=1}^{n} \left[\frac{(e_i \rho_i)^{y_i}}{y_i!} \exp(-e_i \rho_i) \right]$$

  So the log likelihood is: $\sum_{i=1}^{n} \left[y_i \left(\log e_i + \log \rho_i \right) - e_i \rho_i - \log(y_i!) \right]$

- Differentiating wrt ρ_i and setting this derivative equal to zero (for a maximum) gives: $\frac{y_i}{\rho_i} - e_i = 0$ so the maximum likelihood estimate of ρ_i is: $\hat{\rho}_i = \frac{y_i}{e_i}$
  i.e. the familiar **standardised morbidity ratio** (SMR) for the ith observation.

- Here the parameters of the model are $\boldsymbol{\theta} = (\rho_1, \ldots, \rho_n)$ and the likelihood for the data i.e. $P(\boldsymbol{y}; \boldsymbol{\theta})$ or $P(y_1, \ldots, y_n; \rho_1, \ldots, \rho_n)$ is therefore:

$$\prod_{i=1}^{n} \left[\frac{(e_i \rho_i)^{y_i}}{y_i!} \exp(-e_i \rho_i) \right]$$

  So the log likelihood is: $\sum_{i=1}^{n} \left[y_i \left(\log e_i + \log \rho_i \right) - e_i \rho_i - \log(y_i!) \right]$

- Differentiating wrt ρ_i and setting this derivative equal to zero (for a maximum) gives: $\frac{y_i}{\rho_i} - e_i = 0$ so the maximum likelihood estimate of ρ_i is: $\hat{\rho}_i = \frac{y_i}{e_i}$ i.e. the familiar **standardised morbidity ratio** (SMR) for the ith observation.

- Can then go on to show (using the second derivative of the likelihood) that $\text{Var}(\hat{\rho}_i) = \frac{\rho_i}{e_i}$ which may be estimated by $\frac{\hat{\rho}_i}{e_i}$ or alternatively $\frac{y_i}{e_i^2}$. (i.e. extreme SMRs are subject to large standard errors)

- Maximum likelihood is fine as a general approach, but if the form of the likelihood $P(\boldsymbol{y};\boldsymbol{\theta})$ is complex and/or the number of individual parameters involved in $\boldsymbol{\theta}$ is large then the approach may prove either very difficult or infeasible to implement

- Maximum likelihood is fine as a general approach, but if the form of the likelihood $P(\boldsymbol{y};\boldsymbol{\theta})$ is complex and/or the number of individual parameters involved in $\boldsymbol{\theta}$ is large then the approach may prove either very difficult or infeasible to implement
- If so then a **Bayesian approach** to parameter estimation may prove useful

- Maximum likelihood is fine as a general approach, but if the form of the likelihood $P(\boldsymbol{y};\boldsymbol{\theta})$ is complex and/or the number of individual parameters involved in $\boldsymbol{\theta}$ is large then the approach may prove either very difficult or infeasible to implement
- If so then a **Bayesian approach** to parameter estimation may prove useful
- In the Bayesian approach we also think of the parameters as 'random quantities' (rather than fixed constants)
- The statistical model then becomes a **joint probability distribution for both the data and the parameters**: $P(\boldsymbol{y},\boldsymbol{\theta})$ (the likelihood is now the **conditional distribution** of $\boldsymbol{y}$ 'given' the parameter values – $P(\boldsymbol{y}|\boldsymbol{\theta})$)

➢ Elementary probability theory then allows us to relate $P(\boldsymbol{y}, \boldsymbol{\theta})$ to the likelihood:

$$P(\boldsymbol{y}, \boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

➢ Elementary probability theory then allows us to relate $P(\boldsymbol{y}, \boldsymbol{\theta})$ to the likelihood:

$$P(\boldsymbol{y}, \boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

where $P(\boldsymbol{\theta})$ is called the **prior** probability distribution for the parameters. This prior expresses our **uncertainty about $\boldsymbol{\theta}$ before taking the data into account**. It will usually be chosen to be 'non-informative'.

- Elementary probability theory then allows us to relate $P(\boldsymbol{y}, \boldsymbol{\theta})$ to the likelihood:

$$P(\boldsymbol{y}, \boldsymbol{\theta}) = P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

where $P(\boldsymbol{\theta})$ is called the **prior** probability distribution for the parameters. This prior expresses our **uncertainty about $\boldsymbol{\theta}$ before taking the data into account**. It will usually be chosen to be 'non-informative'.

- Bayes Theorem then allows derivation of a **posterior** probability distribution for the parameters given the observed data:

$$P(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{y})} = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

i.e. **'posterior' is proportional to 'likelihood'** $\times$ **'prior'** — the denominator is just a **normalising constant** independent of the parameters (but unfortunately difficult to calculate because it involves a 'nasty' multi-dimensional integral)

- The posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$ expresses our **uncertainty about θ after taking the data into account**.

- The posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$ expresses our **uncertainty about θ after taking the data into account**.
- So any characteristics of interest concerning the parameters (e.g. mean, standard deviation, mode, median, quantiles etc.) may be derived from the corresponding characteristics of the posterior.

- The posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$ expresses our **uncertainty about $\boldsymbol{\theta}$ after taking the data into account**.
- So any characteristics of interest concerning the parameters (e.g. mean, standard deviation, mode, median, quantiles etc.) may be derived from the corresponding characteristics of the posterior.
- For example an obvious choice of a point estimate, $\hat{\boldsymbol{\theta}}$, for the parameter values is the **posterior mean** of the parameters:

$$\hat{\boldsymbol{\theta}} = \mathsf{E}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} P(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} = \frac{\int_{\boldsymbol{\theta}} \boldsymbol{\theta} P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

- The posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$ expresses our **uncertainty about $\boldsymbol{\theta}$ after taking the data into account**.
- So any characteristics of interest concerning the parameters (e.g. mean, standard deviation, mode, median, quantiles etc.) may be derived from the corresponding characteristics of the posterior.
- For example an obvious choice of a point estimate, $\hat{\boldsymbol{\theta}}$, for the parameter values is the **posterior mean** of the parameters:

$$\hat{\boldsymbol{\theta}} = \mathsf{E}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} P(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} = \frac{\int_{\boldsymbol{\theta}} \boldsymbol{\theta} P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} P(\boldsymbol{y}|\boldsymbol{\theta}) P(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

- But it's important to stress that Bayes gives us a full posterior distribution for $\boldsymbol{\theta}$ and thus allows us to examine **any aspect** of $\boldsymbol{\theta}$ we choose and make associated probability statements

- Let's return to our earlier simple example to clarify these ideas.

- Let's return to our earlier simple example to clarify these ideas.
- Recall we were estimating the incidence of a disease in the general population from a random sample of n individuals, each of whom are tested for the disease, and y is the number who test $+$ve.

- Let's return to our earlier simple example to clarify these ideas.
- Recall we were estimating the incidence of a disease in the general population from a random sample of n individuals, each of whom are tested for the disease, and y is the number who test $+$ve.
- As before, let θ be probability an individual has the disease, then the likelihood for y is a binomial distribution:

$$P(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

- Let's return to our earlier simple example to clarify these ideas.
- Recall we were estimating the incidence of a disease in the general population from a random sample of n individuals, each of whom are tested for the disease, and y is the number who test $+$ve.
- As before, let θ be probability an individual has the disease, then the likelihood for y is a binomial distribution:

$$P(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

- Suppose take a prior for θ as U$(0,1)$ i.e. $P(\theta) = 1$ for $0 \leq \theta \leq 1$ (this says θ equally likely to be anywhere in the $(0,1)$ range)

➢ Then posterior is given by:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_\theta P(y|\theta)P(\theta)\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\frac{1}{(n+1)}}$$

➢ Then posterior is given by:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_\theta P(y|\theta)P(\theta)\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\frac{1}{(n+1)}}$$

➢ A reasonable point estimate for θ is the mean of the posterior i.e.

$$\hat{\theta} = \mathsf{E}\left[\theta|y\right] = \int_0^1 \theta(n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta = \frac{y+1}{n+2}$$

➢ Then posterior is given by:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_\theta P(y|\theta)P(\theta)\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\frac{1}{(n+1)}}$$

➢ A reasonable point estimate for θ is the mean of the posterior i.e.

$$\hat{\theta} = \mathsf{E}\left[\theta|y\right] = \int_0^1 \theta(n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta = \frac{y+1}{n+2}$$

➢ Recall the mle estimate of θ was $\frac{y}{n}$. For this prior that is the mode of the posterior or **MAP** estimate (rather than the posterior mean). $\frac{y}{n}$ is actually the posterior mean when prior is taken to be uniform for the log odds i.e. for $\log\frac{\theta}{1-\theta}$.

- Then posterior is given by:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_\theta P(y|\theta)P(\theta)\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\frac{1}{(n+1)}}$$

- A reasonable point estimate for θ is the mean of the posterior i.e.

$$\hat{\theta} = \mathsf{E}\left[\theta|y\right] = \int_0^1 \theta(n+1)\binom{n}{y}\theta^y(1-\theta)^{n-y}\,d\theta = \frac{y+1}{n+2}$$

- Recall the mle estimate of θ was $\frac{y}{n}$. For this prior that is the mode of the posterior or **MAP** estimate (rather than the posterior mean). $\frac{y}{n}$ is actually the posterior mean when prior is taken to be uniform for the log odds i.e. for $\log\frac{\theta}{1-\theta}$.
- Illustrates that prior choice can be tricky — which is most sensible estimate of θ?

- In summary the Bayesian modelling approach is:

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$
  - Derive the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$
  - Derive the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$
  - Estimate any characteristic of interest involving one or more of the parameters by the corresponding characteristic of the posterior. E.g the posterior mean for a point estimate or the posterior standard deviation for a standard error.

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$
  - Derive the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$
  - Estimate any characteristic of interest involving one or more of the parameters by the corresponding characteristic of the posterior. E.g the posterior mean for a point estimate or the posterior standard deviation for a standard error.
- This is a very general and flexible approach to statistical modelling capable of handling very complex modelling frameworks.

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$
  - Derive the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$
  - Estimate any characteristic of interest involving one or more of the parameters by the corresponding characteristic of the posterior. E.g the posterior mean for a point estimate or the posterior standard deviation for a standard error.
- This is a very general and flexible approach to statistical modelling capable of handling very complex modelling frameworks. **Problem is that you have to be able to integrate to find the posterior distribution in order to use the method!**

- In summary the Bayesian modelling approach is:
  - Choose (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
  - Choose (joint) prior for the parameters—$P(\boldsymbol{\theta})$
  - Derive the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$
  - Estimate any characteristic of interest involving one or more of the parameters by the corresponding characteristic of the posterior. E.g the posterior mean for a point estimate or the posterior standard deviation for a standard error.
- This is a very general and flexible approach to statistical modelling capable of handling very complex modelling frameworks. **Problem is that you have to be able to integrate to find the posterior distribution in order to use the method!** So why is it any more useful than maximum likelihood?

- It's true that until relatively recently the integrations involved in determining the posterior have presented practical difficulties in Bayesian modelling, especially when large numbers of parameters are involved.
- In many applications mathematical evaluation of the posterior is impossible because of the multidimensional integration involved in determining the normalising denominator

- It's true that until relatively recently the integrations involved in determining the posterior have presented practical difficulties in Bayesian modelling, especially when large numbers of parameters are involved.
- In many applications mathematical evaluation of the posterior is impossible because of the multidimensional integration involved in determining the normalising denominator
- But now the 'engineering' approach of **Monte Carlo integration** can be used.

- It's true that until relatively recently the integrations involved in determining the posterior have presented practical difficulties in Bayesian modelling, especially when large numbers of parameters are involved.
- In many applications mathematical evaluation of the posterior is impossible because of the multidimensional integration involved in determining the normalising denominator
- But now the 'engineering' approach of **Monte Carlo integration** can be used.
- This evaluates any characteristic of the posterior by **simulating** many sample values from it and then approximating any characteristic of it by the corresponding characteristic of these samples. If samples are numerous and representative of the posterior then they can provide virtually complete information about it.

- For example suppose $f(\boldsymbol{\theta})$ is some function of the parameters of interest (e.g. a prediction from the model)

- For example suppose $f(\boldsymbol{\theta})$ is some function of the parameters of interest (e.g. a prediction from the model)
- Let $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(n)}$ where $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$, denote n samples from the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$.

- For example suppose $f(\boldsymbol{\theta})$ is some function of the parameters of interest (e.g. a prediction from the model)
- Let $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(n)}$ where $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$, denote n samples from the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$.
- Then if n is large enough:

$$\hat{f}(\boldsymbol{\theta}) = \mathsf{E}\left[f(\boldsymbol{\theta})|\boldsymbol{y}\right] \approx \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{\theta}^{(i)})$$

➢ That's nice! **But** the problem is then how to simulate samples from the posterior? *Direct* sampling from $P(\boldsymbol{\theta}|\boldsymbol{y})$ is difficult (because you don't know what it is!). But *indirect* sampling from a **Markov Chain** (MC) with $P(\boldsymbol{\theta}|\boldsymbol{y})$ as its stationary (equilibrium) distribution is feasible.

- That's nice! **But** the problem is then how to simulate samples from the posterior? *Direct* sampling from $P(\boldsymbol{\theta}|\boldsymbol{y})$ is difficult (because you don't know what it is!). But *indirect* sampling from a **Markov Chain** (MC) with $P(\boldsymbol{\theta}|\boldsymbol{y})$ as its stationary (equilibrium) distribution is feasible.
- Sequence $\{\boldsymbol{\theta}^{(i)}\}$ is an MC if $P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(i)}) = P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})$ i.e. next value $\boldsymbol{\theta}^{(i+1)}$ depends only on current value $\boldsymbol{\theta}^{(i)}$ and not previous values.

- That's nice! **But** the problem is then how to simulate samples from the posterior? *Direct* sampling from $P(\boldsymbol{\theta}|\boldsymbol{y})$ is difficult (because you don't know what it is!). But *indirect* sampling from a **Markov Chain** (MC) with $P(\boldsymbol{\theta}|\boldsymbol{y})$ as its stationary (equilibrium) distribution is feasible.
- Sequence $\{\boldsymbol{\theta}^{(i)}\}$ is an MC if $P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(i)}) = P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})$ i.e. next value $\boldsymbol{\theta}^{(i+1)}$ depends only on current value $\boldsymbol{\theta}^{(i)}$ and not previous values. Subject to certain conditions, MCs gradually 'forget' their initial value and converge to a **stationary distribution** (overall probability of taking any value remains same and this is independent of the original starting value). Subsequent values of the chain are then samples from this stationary distribution

- That's nice! **But** the problem is then how to simulate samples from the posterior? *Direct* sampling from $P(\boldsymbol{\theta}|\boldsymbol{y})$ is difficult (because you don't know what it is!). But *indirect* sampling from a **Markov Chain** (MC) with $P(\boldsymbol{\theta}|\boldsymbol{y})$ as its stationary (equilibrium) distribution is feasible.
- Sequence $\{\boldsymbol{\theta}^{(i)}\}$ is an MC if $P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(i)}) = P(\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\theta}^{(i)})$ i.e. next value $\boldsymbol{\theta}^{(i+1)}$ depends only on current value $\boldsymbol{\theta}^{(i)}$ and not previous values. Subject to certain conditions, MCs gradually 'forget' their initial value and converge to a **stationary distribution** (overall probability of taking any value remains same and this is independent of the original starting value). Subsequent values of the chain are then samples from this stationary distribution
- Hence construct an MC with a stationary distribution identical to the posterior and use values from that MC chain after a sufficiently long *burn in* as simulated samples from the posterior. This is called **Markov Chain Monte Carlo** (MCMC)

- The reason that this is so useful in practice is because it is surprisingly easy to construct a Markov Chain which has a given stationary distribution.

- The reason that this is so useful in practice is because it is surprisingly easy to construct a Markov Chain which has a given stationary distribution.
- This is achieved via the general **Metropolis-Hastings** algorithm.

- The reason that this is so useful in practice is because it is surprisingly easy to construct a Markov Chain which has a given stationary distribution.
- This is achieved via the general **Metropolis-Hastings** algorithm.
- Furthermore this algorithm only requires the stationary distribution (in our case the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$) to be specified up to the normalising constant—i.e. we just need the product of the prior and the likelihood — no nasty integration required!

- The reason that this is so useful in practice is because it is surprisingly easy to construct a Markov Chain which has a given stationary distribution.
- This is achieved via the general **Metropolis-Hastings** algorithm.
- Furthermore this algorithm only requires the stationary distribution (in our case the posterior $P(\boldsymbol{\theta}|\boldsymbol{y})$) to be specified up to the normalising constant—i.e. we just need the product of the prior and the likelihood — no nasty integration required!
- Brooks (1998) or Gilks *et al* (1996) provide excellent accounts of MCMC methodology

Markov Chain Monte Carlo (MCMC) methods

- The Metropolis-Hastings algorithm constructs a Markov Chain to converge to the **target distribution** by sampling a **candidate** for the next value of the chain from a **proposal distribution** and then either accepting it or rejecting it according to a **acceptance probability** which depends upon the proposal distribution, the target distribution, the current state of the chain and the candidate value

Markov Chain Monte Carlo (MCMC) methods

- The Metropolis-Hastings algorithm constructs a Markov Chain to converge to the **target distribution** by sampling a **candidate** for the next value of the chain from a **proposal distribution** and then either accepting it or rejecting it according to a **acceptance probability** which depends upon the proposal distribution, the target distribution, the current state of the chain and the candidate value
- The proposal distribution can have any form subject to certain regularity conditions. It will be chosen to be appropriate to the particular target distribution required and so that it is easy to sample from

➢ Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:

- Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:
  - Set $i = 0$ and choose arbitrary starting values $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta}$

- Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:
  - Set $i = 0$ and choose arbitrary starting values $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta}$
  - Sample a candidate, $\boldsymbol{\theta}^{(*)}$, for the next state of the chain given the current state $\boldsymbol{\theta}^{(i)}$ from a pre-selected proposal distribution, $q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})$

- Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:
  - Set $i = 0$ and choose arbitrary starting values $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta}$
  - Sample a candidate, $\boldsymbol{\theta}^{(*)}$, for the next state of the chain given the current state $\boldsymbol{\theta}^{(i)}$ from a pre-selected proposal distribution, $q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})$
  - Compute an **acceptance probability**

$$\alpha\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(*)}\right) = \min\left\{1, \frac{p(\boldsymbol{\theta}^{(*)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(*)})}{p(\boldsymbol{\theta}^{(i)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})}\right\}$$

(note *target distribution appears only as a ratio, so unknown constant of proportionality involved in* $p(\boldsymbol{\theta}|\boldsymbol{y})$ *cancels*)

- Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:
  - Set $i = 0$ and choose arbitrary starting values $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta}$
  - Sample a candidate, $\boldsymbol{\theta}^{(*)}$, for the next state of the chain given the current state $\boldsymbol{\theta}^{(i)}$ from a pre-selected proposal distribution, $q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})$
  - Compute an **acceptance probability**

$$\alpha\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(*)}\right) = \min\left\{1, \frac{p(\boldsymbol{\theta}^{(*)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(*)})}{p(\boldsymbol{\theta}^{(i)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})}\right\}$$

    (note *target distribution appears only as a ratio, so unknown constant of proportionality involved in* $p(\boldsymbol{\theta}|\boldsymbol{y})$ *cancels*)
  - Sample u such that $u \sim \mathrm{U}(0, 1)$. If $u \leq \alpha(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(*)})$ then set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(*)}$, else set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$

- Given a target distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, the Metropolis-Hastings proceeds as follows:
  - Set $i = 0$ and choose arbitrary starting values $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta}$
  - Sample a candidate, $\boldsymbol{\theta}^{(*)}$, for the next state of the chain given the current state $\boldsymbol{\theta}^{(i)}$ from a pre-selected proposal distribution, $q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})$
  - Compute an **acceptance probability**

$$\alpha\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(*)}\right) = \min\left\{1, \frac{p(\boldsymbol{\theta}^{(*)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(*)})}{p(\boldsymbol{\theta}^{(i)}|\boldsymbol{y})q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(i)})}\right\}$$

    (note *target distribution appears only as a ratio, so unknown constant of proportionality involved in* $p(\boldsymbol{\theta}|\boldsymbol{y})$ *cancels*)
  - Sample u such that $u \sim \mathrm{U}(0, 1)$. If $u \leq \alpha(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(*)})$ then set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(*)}$, else set $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$
  - Set $i = i + 1$ and return to step 2 for a new candidate (repeat 1000's of times)

- In general, particular versions of the algorithm need to be 'hand crafted' to fit different applications so as to obtain:
  - a good **rate of convergence** (short burn-in needed to achieve stationary distribution)
  - and good **rate of mixing** (fast movement around the support of the stationary distribution once it is achieved)

- In general, particular versions of the algorithm need to be 'hand crafted' to fit different applications so as to obtain:
  - a good **rate of convergence** (short burn-in needed to achieve stationary distribution)
  - and good **rate of mixing** (fast movement around the support of the stationary distribution once it is achieved)
- But all this is generally easier than maximising the equivalent likelihoods and you get a full distribution for the parameters from it, rather than just point estimates and standard errors

- One variant of the Metropolis-Hastings algorithm is **Gibbs sampling**. This is convenient when conditional posterior distributions of each parameter given values of all the others and the data are available

- One variant of the Metropolis-Hastings algorithm is **Gibbs sampling**. This is convenient when conditional posterior distributions of each parameter given values of all the others and the data are available
- Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ then in Gibbs sampling we:

- One variant of the Metropolis-Hastings algorithm is **Gibbs sampling**. This is convenient when conditional posterior distributions of each parameter given values of all the others and the data are available
- Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ then in Gibbs sampling we:
  - Set $i = 0$ and choose arbitrary starting values $(\theta_1^{(0)}, \ldots, \theta_p^{(0)})$

- One variant of the Metropolis-Hastings algorithm is **Gibbs sampling**. This is convenient when conditional posterior distributions of each parameter given values of all the others and the data are available
- Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ then in Gibbs sampling we:
  - Set $i = 0$ and choose arbitrary starting values $(\theta_1^{(0)}, \ldots, \theta_p^{(0)})$
  - Sample $\theta_1^{(i+1)}$ from $P(\theta_1 | \theta_2^{(i)}, \ldots, \theta_p^{(i)}, \boldsymbol{y})$
    Sample $\theta_2^{(i+1)}$ from $P(\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_p^{(i)}, \boldsymbol{y})$
    $\vdots$
    Sample $\theta_p^{(i+1)}$ from $P(\theta_p | \theta_1^{(i+1)}, \ldots, \theta_{p-1}^{(i+1)}, \boldsymbol{y})$

- One variant of the Metropolis-Hastings algorithm is **Gibbs sampling**. This is convenient when conditional posterior distributions of each parameter given values of all the others and the data are available
- Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ then in Gibbs sampling we:
  - Set $i = 0$ and choose arbitrary starting values $(\theta_1^{(0)}, \ldots, \theta_p^{(0)})$
  - Sample $\theta_1^{(i+1)}$ from $P(\theta_1 | \theta_2^{(i)}, \ldots, \theta_p^{(i)}, \boldsymbol{y})$
    Sample $\theta_2^{(i+1)}$ from $P(\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_p^{(i)}, \boldsymbol{y})$
    $\vdots$
    Sample $\theta_p^{(i+1)}$ from $P(\theta_p | \theta_1^{(i+1)}, \ldots, \theta_{p-1}^{(i+1)}, \boldsymbol{y})$
  - Set $i = i + 1$ and repeat the last step (do this many 1000's of times)

- To do Gibbs sampling we need to be able to specify the full conditional posterior distributions of each parameter given the values of the others and the data. That is we need $P(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p, \boldsymbol{y})$

- To do Gibbs sampling we need to be able to specify the full conditional posterior distributions of each parameter given the values of the others and the data. That is we need $P(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p, \boldsymbol{y})$
- Turns out that these are relatively easy to work out for a wide range of commonly used models, and this includes many spatial models, (see Gilks *et al*, 1993)

- To do Gibbs sampling we need to be able to specify the full conditional posterior distributions of each parameter given the values of the others and the data. That is we need $P(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p, \boldsymbol{y})$
- Turns out that these are relatively easy to work out for a wide range of commonly used models, and this includes many spatial models, (see Gilks *et al*, 1993)
- We also need to be able to simulate observations from each of these distributions and this again turns out to be relatively easy since each is one-dimensional and often 'log concave'. Which means that general techniques such as *adaptive rejection sampling* can be used

- To do Gibbs sampling we need to be able to specify the full conditional posterior distributions of each parameter given the values of the others and the data. That is we need $P(\theta_j|\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p, \boldsymbol{y})$
- Turns out that these are relatively easy to work out for a wide range of commonly used models, and this includes many spatial models, (see Gilks *et al*, 1993)
- We also need to be able to simulate observations from each of these distributions and this again turns out to be relatively easy since each is one-dimensional and often 'log concave'. Which means that general techniques such as *adaptive rejection sampling* can be used
- Hence Gibbs sampling is able to be used in a wide variety of Bayesian models. It forms the basis of the MCMC method in the public domain **WinBUGS** package (**B**ayesian Inference **U**sing **G**ibbs **S**ampling) (see Spiegelhalter *et al*, 1997)

- After sufficient 'burn in' successive samples $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$ formed from general Metropolis-Hastings or some variant such as Gibbs Sampling settle down to samples from a markov chain with stationary distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$.

- After sufficient 'burn in' successive samples $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$ formed from general Metropolis-Hastings or some variant such as Gibbs Sampling settle down to samples from a markov chain with stationary distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$.
- Samples from marginal posteriors (e.g. $P(\theta_j|\boldsymbol{y})$) are approximated by simply picking out the values for one parameter from the samples ignoring the other parameters.

- After sufficient 'burn in' successive samples $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_p^{(i)})$ formed from general Metropolis-Hastings or some variant such as Gibbs Sampling settle down to samples from a markov chain with stationary distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$.
- Samples from marginal posteriors (e.g. $P(\theta_j|\boldsymbol{y})$) are approximated by simply picking out the values for one parameter from the samples ignoring the other parameters.
- Characteristics concerning a parameter are then estimated from the marginal posterior samples via their sample equivalents (e.g. mean, mode, median, standard deviation, quantiles etc.)

- As said, important issues in MCMC to ensure good estimates are **convergence** ('burn-in' required) and **mixing** (required number of samples after convergence)

- As said, important issues in MCMC to ensure good estimates are **convergence** ('burn-in' required) and **mixing** (required number of samples after convergence)
- There are formal ways to assess convergence (see references) but essential point is that samples for any parameter should be a random scatter about a stable mean value. Note convergence is to a target distribution not to a single value. Check convergence by several long runs and widely different starting values (**multiple chains**). Statistics such as **'R hat'** help to assess whether the chains have converged (Rule of thumb: its value should be < 1.2 for each parameter)

- As said, important issues in MCMC to ensure good estimates are **convergence** ('burn-in' required) and **mixing** (required number of samples after convergence)
- There are formal ways to assess convergence (see references) but essential point is that samples for any parameter should be a random scatter about a stable mean value. Note convergence is to a target distribution not to a single value. Check convergence by several long runs and widely different starting values (**multiple chains**). Statistics such as **'R hat'** help to assess whether the chains have converged (Rule of thumb: its value should be < 1.2 for each parameter)
- After convergence, sufficient samples required to ensure posterior variance is estimated accurately. Again formal techniques exist (see references). A useful statistic is the **MC standard error** for each parameter. Ideally want MC error small in relation to posterior st. dev. (Rule of thumb: run simulation until MC error for each parameter $< 5\%$ of sample (posterior) st. dev)

- Choice of suitable prior distributions in Bayesian modelling can be controversial (see references).
- **Conjugate** priors are priors which lead to the posterior being in the same family as the prior. These are useful, but unfortunately conjugate priors do not exist for all likelihoods. MCMC methods make conjugacy less important.

- Choice of suitable prior distributions in Bayesian modelling can be controversial (see references).
- **Conjugate** priors are priors which lead to the posterior being in the same family as the prior. These are useful, but unfortunately conjugate priors do not exist for all likelihoods. MCMC methods make conjugacy less important.
- In some cases the prior for the basic model parameters $P(\boldsymbol{\theta})$ will itself involve some additional parameters, $\boldsymbol{\gamma}$, i.e. the prior may be of the form $P(\boldsymbol{\theta}|\boldsymbol{\gamma})$. Then we have a **hierarchical** model. Parameters of the prior are known as **hyperparameters**.

- One beauty of the Bayesian framework is that it easily incorporates these extra unknown quantities.

- One beauty of the Bayesian framework is that it easily incorporates these extra unknown quantities.
- We just extend the same game and specify a joint **hyperprior** $P(\boldsymbol{\gamma})$ for the hyperparameters $\boldsymbol{\gamma}$ and then use:

$$P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})}{P(\boldsymbol{y})} = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})}{\int_{\boldsymbol{\theta}}\int_{\boldsymbol{\gamma}} P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})\, d\boldsymbol{\gamma}\, d\boldsymbol{\theta}}$$

- One beauty of the Bayesian framework is that it easily incorporates these extra unknown quantities.
- We just extend the same game and specify a joint **hyperprior** $P(\boldsymbol{\gamma})$ for the hyperparameters $\boldsymbol{\gamma}$ and then use:

$$P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})}{P(\boldsymbol{y})} = \frac{P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})}{\int_{\boldsymbol{\theta}}\int_{\boldsymbol{\gamma}} P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})\, d\boldsymbol{\gamma}\, d\boldsymbol{\theta}}$$

- Essentially the hyperparameters $\boldsymbol{\gamma}$ are treated on the same footing as the primary parameters $\boldsymbol{\theta}$

MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

- Choose appropriate (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$

MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

- Choose appropriate (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
- Choose appropriate (joint) prior for the parameters—$P(\boldsymbol{\theta}|\boldsymbol{\gamma})$

MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

- Choose appropriate (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
- Choose appropriate (joint) prior for the parameters—$P(\boldsymbol{\theta}|\boldsymbol{\gamma})$
- Choose appropriate (joint) hyperprior for the hyperparameters—$P(\boldsymbol{\gamma})$

MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

- Choose appropriate (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
- Choose appropriate (joint) prior for the parameters—$P(\boldsymbol{\theta}|\boldsymbol{\gamma})$
- Choose appropriate (joint) hyperprior for the hyperparameters—$P(\boldsymbol{\gamma})$
- Use MCMC (Gibbs Sampling or general Metropolis-Hastings) to generate numerous samples $(\boldsymbol{\theta}, \boldsymbol{\gamma})^{(i)}$, $i = 1, \ldots, n$ from posterior $P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y})$ using:

$$P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})$$

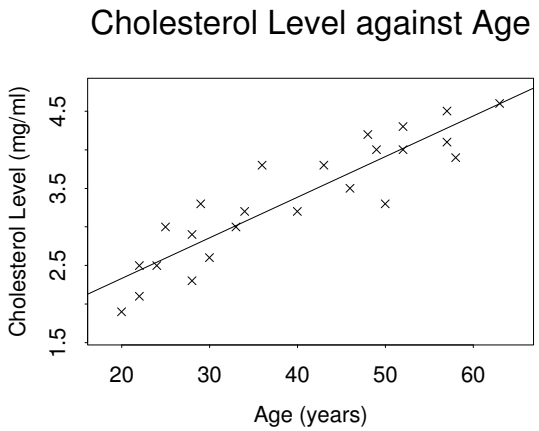
MCMC methods make Bayesian modelling of complex situations involving many parameters a practical feasibility. Use is now widespread (non spatial as well as spatial). In summary the full Bayesian MCMC approach is:

- Choose appropriate (joint) probability model (likelihood) for the data—$P(\boldsymbol{y}|\boldsymbol{\theta})$
- Choose appropriate (joint) prior for the parameters—$P(\boldsymbol{\theta}|\boldsymbol{\gamma})$
- Choose appropriate (joint) hyperprior for the hyperparameters—$P(\boldsymbol{\gamma})$
- Use MCMC (Gibbs Sampling or general Metropolis-Hastings) to generate numerous samples $(\boldsymbol{\theta}, \boldsymbol{\gamma})^{(i)}$, $i = 1, \ldots, n$ from posterior $P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y})$ using:

$$P(\boldsymbol{\theta}, \boldsymbol{\gamma}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\gamma})P(\boldsymbol{\gamma})$$

- Estimate any characteristic of interest involving one or more of the parameters or hyperparameters by the equivalent characteristic of the posterior samples

➢ Lets see all this in action on a very simple example. Cholesterol level (mg/ml) and age (years) was measured for 24 patients diagnosed with hyperlipoproteinaemia and resulted in the following scatter plot:



- Lets see all this in action on a very simple example. Cholesterol level (mg/ml) and age (years) was measured for 24 patients diagnosed with hyperlipoproteinaemia and resulted in the following scatter plot:

- Sample correlation between age and cholesterol is strong (≈ 0.9) and a standard linear regression model (indicated in the plot) results in the following model:

$$y_i \text{ (Cholesterol level)} = \alpha \text{ (1.2799)} + \beta \text{ (0.0526)} \times \text{age}_i$$

with residual standard deviation σ equal to 0.334.

➢ In the notation we have been using, we can represent this as a Bayesian model by taking the parameters as: $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$ and the data $\boldsymbol{y} = (y_1, \ldots, y_{24})$ to consist of independent observations with $P(y_i|\boldsymbol{\theta}) \sim \mathsf{N}(\alpha + \beta \mathsf{age}_i, \sigma^2)$ so that:

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{24} P(y_i|\boldsymbol{\theta})$$

- In the notation we have been using, we can represent this as a Bayesian model by taking the parameters as: $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$ and the data $\boldsymbol{y} = (y_1, \ldots, y_{24})$ to consist of independent observations with $P(y_i|\boldsymbol{\theta}) \sim \mathsf{N}(\alpha + \beta \mathsf{age}_i, \sigma^2)$ so that:

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{24} P(y_i|\boldsymbol{\theta})$$

- As priors for both α and β we use Normal distributions with zero means and large variances.

- In the notation we have been using, we can represent this as a Bayesian model by taking the parameters as: $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$ and the data $\boldsymbol{y} = (y_1, \ldots, y_{24})$ to consist of independent observations with $P(y_i|\boldsymbol{\theta}) \sim \mathsf{N}(\alpha + \beta \mathsf{age}_i, \sigma^2)$ so that:

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{24} P(y_i|\boldsymbol{\theta})$$

- As priors for both α and β we use Normal distributions with zero means and large variances.
- As a prior for σ^2 we take $\tau = \frac{1}{\sigma^2}$ (known as the **precision**) to have a Gamma distribution with mean 1 and a large variance.

- In the notation we have been using, we can represent this as a Bayesian model by taking the parameters as: $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$ and the data $\boldsymbol{y} = (y_1, \ldots, y_{24})$ to consist of independent observations with $P(y_i|\boldsymbol{\theta}) \sim \mathsf{N}(\alpha + \beta \mathsf{age}_i, \sigma^2)$ so that:

$$P(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{24} P(y_i|\boldsymbol{\theta})$$

- As priors for both α and β we use Normal distributions with zero means and large variances.
- As a prior for σ^2 we take $\tau = \frac{1}{\sigma^2}$ (known as the **precision**) to have a Gamma distribution with mean 1 and a large variance.
- These choices are pretty standard for this situation and represent minimally informative priors. Note in this case the priors do not involve hyperparameters

In **WinBUGS** we require the following specification:

Model

```
for(i in 1:N){
Y[i] ~ dnorm(mu[i], tau)   # normal distribution for data, mean mu, precision tau
mu[i] <- alpha + beta * age[i]   # linear model for mean mu
}
alpha ~ dnorm(0, 1.0E-6)   # diffuse normal prior for alpha
beta ~ dnorm(0, 1.0E-6)   # diffuse normal prior for beta
tau ~ dgamma(.001, .001)   # vague gamma prior for tau
sigma <- 1/sqrt(tau)   # st. deviation for Y derived from tau
```

Data

```
list(N = 24,Y = c(3.5,1.9,...,3.3),age = c(46, 20,...,50))
```

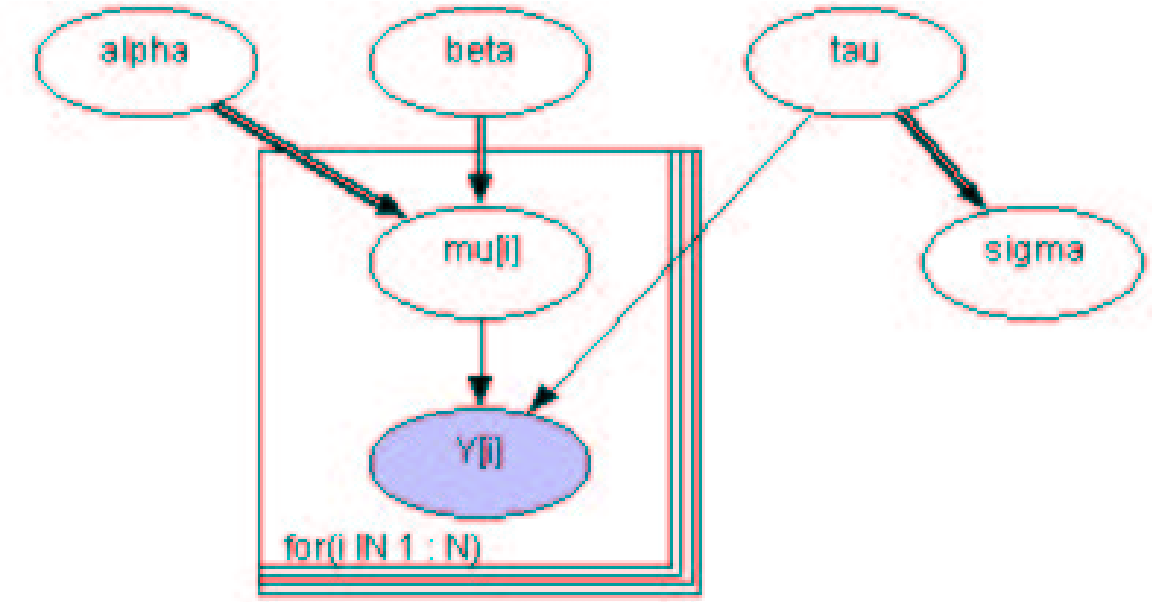
Initial values for the MCMC sampler

```
list(alpha = 0, beta = 0, tau = 1)
```

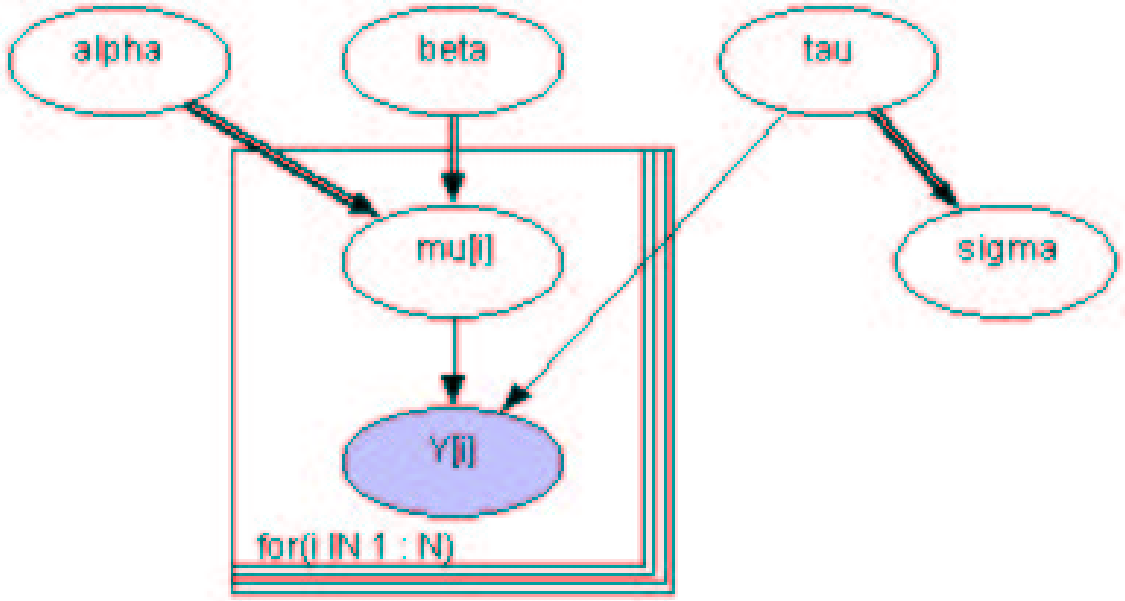
☞

- Note that **WinBUGS** provides an interface to specify models via a *directed graph* which indicates the nature of all quantities in the model and their dependencies.

➢ Note that **WinBUGS** provides an interface to specify models via a *directed graph* which indicates the nature of all quantities in the model and their dependencies. In this case a suitable graphical model would be:



- Note that **WinBUGS** provides an interface to specify models via a *directed graph* which indicates the nature of all quantities in the model and their dependencies. In this case a suitable graphical model would be:

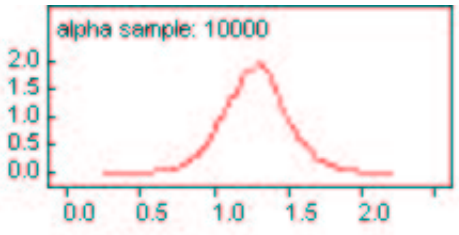


- Each of the nodes in the diagram can be edited to define the details of the corresponding part of the model

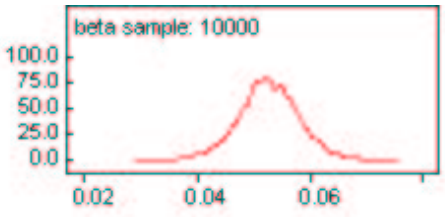
- We can now run this model to generate samples from the posterior distribution and collect summary statistics from those samples. **WinBUGS** itself derives the conditional distributions required for the Gibbs Sampling from the dependency structure specified in the model.

- We can now run this model to generate samples from the posterior distribution and collect summary statistics from those samples. **WinBUGS** itself derives the conditional distributions required for the Gibbs Sampling from the dependency structure specified in the model.

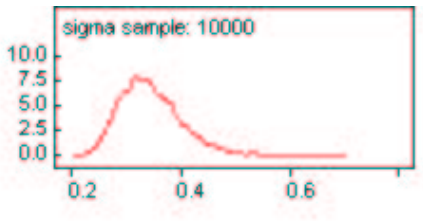
- In this case 10,000 samples with a 'burn-in' of 5000 values gave the following (kernel density) estimates for the marginal posterior distributions of each parameter:



$P(\alpha|\boldsymbol{y})$



$P(\beta|\boldsymbol{y})$



$P(\sigma|\boldsymbol{y})$

➢ The corresponding summary statistics were:

posterior	mean	sd	2.5%	median	97.5%
$P(\alpha\|\boldsymbol{y})$	1.27800	0.224300	0.83170	1.27900	1.73800
$P(\beta\|\boldsymbol{y})$	0.05265	0.005406	0.04178	0.05259	0.06324
$P(\sigma\|\boldsymbol{y})$	0.34600	0.055730	0.25660	0.33930	0.47360

- The corresponding summary statistics were:

posterior	mean	sd	2.5%	median	97.5%
$P(\alpha\|\boldsymbol{y})$	1.27800	0.224300	0.83170	1.27900	1.73800
$P(\beta\|\boldsymbol{y})$	0.05265	0.005406	0.04178	0.05259	0.06324
$P(\sigma\|\boldsymbol{y})$	0.34600	0.055730	0.25660	0.33930	0.47360

- Results from the conventional regression were:

parameter	estimate	sd
α	1.279900	0.215700
β	0.052625	0.005192
σ	0.334000	

- The corresponding summary statistics were:

posterior	mean	sd	2.5%	median	97.5%
$P(\alpha\|\boldsymbol{y})$	1.27800	0.224300	0.83170	1.27900	1.73800
$P(\beta\|\boldsymbol{y})$	0.05265	0.005406	0.04178	0.05259	0.06324
$P(\sigma\|\boldsymbol{y})$	0.34600	0.055730	0.25660	0.33930	0.47360

- Results from the conventional regression were:

parameter	estimate	sd
α	1.279900	0.215700
β	0.052625	0.005192
σ	0.334000	

- Note that changing the model from $P(y_i|\boldsymbol{\theta}) \sim \mathrm{N}(\alpha + \beta \mathrm{age}_i, \sigma^2)$ to one with a different distributional assumption, or with a mean which is a non-linear function of the parameters means that regression cannot be used (a GLM is then required). However, in the Bayesian case it means a simple adjustment to the model specification **the basic approach remains unchanged**.

- Note that good links have been developed between **WinBUGS** and the versatile statistical software environment **SPlus** (and its many add on packages). These links also exist for **R**— the public domain equivalent of **SPlus**.

- Note that good links have been developed between **WinBUGS** and the versatile statistical software environment **SPlus** (and its many add on packages). These links also exist for **R**— the public domain equivalent of **SPlus**.
- For example, **R** package **R2WinBugs** allows one to set up data and model specification in **R** and then use this to call **WinBUGS** to do the MCMC sampling and return the results to **R** for further analysis ☞

- Note that good links have been developed between **WinBUGS** and the versatile statistical software environment **SPlus** (and its many add on packages). These links also exist for **R**— the public domain equivalent of **SPlus**.
- For example, **R** package **R2WinBugs** allows one to set up data and model specification in **R** and then use this to call **WinBUGS** to do the MCMC sampling and return the results to **R** for further analysis ☞
- Note also that **WinBUGS** includes an add on package known as **GeoBUGS** which allows display of model results on maps imported by the user.
- There also exists a **maptools** package for **R** which allows for the importation of maps from GIS software (such as **ARC/INFO**/**ARC/View** or **MAPINFO**) and the plotting of such maps in conjunction with results from the **R**/**WinBUGS** interface.

Disease Mapping

Disease Mapping

- Maps of disease incidence are useful for several purposes and production of disease ‘atlases’ has a long tradition
  - Description of geographical distribution of disease
  - Hypothesis generation
  - Surveillance — to highlight areas at apparently high risk
  - Placing point source/cluster investigations in context
  - Aid to policy formation and resource allocation

Disease Mapping

- Maps of disease incidence are useful for several purposes and production of disease ‘atlases’ has a long tradition
  - Description of geographical distribution of disease
  - Hypothesis generation
  - Surveillance — to highlight areas at apparently high risk
  - Placing point source/cluster investigations in context
  - Aid to policy formation and resource allocation
- Methods are sought which produce a ‘clean’ map free of random noise and effects produced by population size/age/sex variations or other well-known risk factors (conceptual similarities to ‘filtering’ or ‘cleaning’ in image processing)

- Recall our focus is purely on data in the form of aggregated measures of disease incidence (rates in areas)
- Mapping of such data can be carried out at a variety of scales (International, National, sub-National). The models we discuss are particularly important at the sub-National or 'small-area' scale, where numbers of cases and risk populations are relatively small and observed SMRs can be highly variable (recall variance of SMRs ($\frac{y_i}{e_i^2}$) is high when risk populations or cases are small)

- Recall our focus is purely on data in the form of aggregated measures of disease incidence (rates in areas)
- Mapping of such data can be carried out at a variety of scales (International, National, sub-National). The models we discuss are particularly important at the sub-National or 'small-area' scale, where numbers of cases and risk populations are relatively small and observed SMRs can be highly variable (recall variance of SMRs ($\frac{y_i}{e_i^2}$) is high when risk populations or cases are small)
- Different models and approaches can be used (see references). I will focus here on what has emerged as the 'mainstream'—that based on a Poisson **Generalised Linear Mixed Model** or (**GLMM**)
  - Generalised $\Rightarrow$ error distribution is other than Normal (Gaussian)
  - Mixed $\Rightarrow$ model contains both **fixed** and **random effects** (parameters)

➢ We have already introduced the basic model for observed counts y_i i.e.:

$$y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$$

where e_i is the known 'expected' number of cases (based on some global reference rates within suitable population strata) and ρ_i is the unknown relative risk in area i compared to the chosen reference rates.

- We have already introduced the basic model for observed counts y_i i.e.:

$$y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$$

  where e_i is the known 'expected' number of cases (based on some global reference rates within suitable population strata) and ρ_i is the unknown relative risk in area i compared to the chosen reference rates.

- Such a model is reasonable for a relatively rare and non-infectious disease with assumed constant risk within each area $\times$ strata combination. Sometimes for less rare conditions a Binomial formulation might be appropriate, but we focus on the more common Poisson case here.

- We have already introduced the basic model for observed counts y_i i.e.:

$$y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$$

  where e_i is the known 'expected' number of cases (based on some global reference rates within suitable population strata) and ρ_i is the unknown relative risk in area i compared to the chosen reference rates.

- Such a model is reasonable for a relatively rare and non-infectious disease with assumed constant risk within each area $\times$ strata combination. Sometimes for less rare conditions a Binomial formulation might be appropriate, but we focus on the more common Poisson case here.

- Note the model for μ_i can be expressed equivalently as: $\log \mu_i = \log e_i + \theta_i$ where θ_i denotes log relative risk (i.e. $\theta_i = \log \rho_i$ or $\rho_i = \exp(\theta_i)$).

- Have already seen that if ρ_i are taken as **fixed effects** in this model then it is just a standard **Generalised Linear Model** (**GLM**) and the mles $\hat{\rho}_i$ are just the traditional SMRs $\frac{y_i}{e_i}$ (ratio of observed to expected cases)

- Have already seen that if ρ_i are taken as **fixed effects** in this model then it is just a standard **Generalised Linear Model** (**GLM**) and the mles $\hat{\rho}_i$ are just the traditional SMRs $\frac{y_i}{e_i}$ (ratio of observed to expected cases)
- But we have also noted that SMRs may be unreliable as estimates of relative risk and thus naïve use of SMRs in disease mapping can be misleading — SMR maps are unstable due to low event counts or populations at risk, with small changes in case numbers sometimes producing dramatic shifts in particular SMRs

- Have already seen that if ρ_i are taken as **fixed effects** in this model then it is just a standard **Generalised Linear Model** (**GLM**) and the mles $\hat{\rho}_i$ are just the traditional SMRs $\frac{y_i}{e_i}$ (ratio of observed to expected cases)
- But we have also noted that SMRs may be unreliable as estimates of relative risk and thus naïve use of SMRs in disease mapping can be misleading — SMR maps are unstable due to low event counts or populations at risk, with small changes in case numbers sometimes producing dramatic shifts in particular SMRs
- In devising models to counter this, one may envisage the total variability in the observed rates or SMRs as having two components:
  - **within** area variation about the true underlying area rate (due to unmeasured or unknown risk variations and/or data inaccuracies within the area)
  - **between** area variations in the true rates

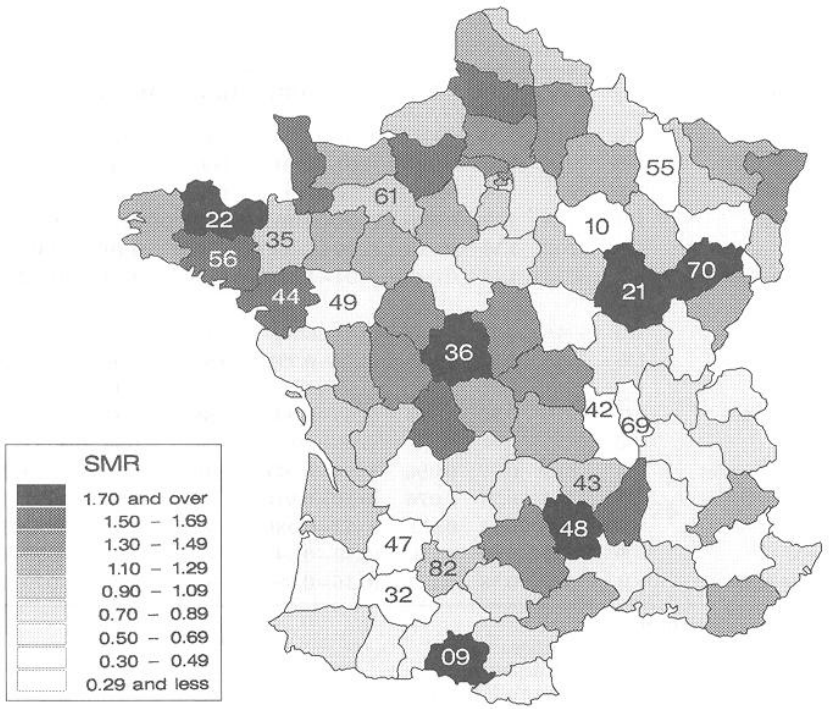
- In other words the data exhibit extra-Poisson variation or **overdispersion** (because of the within area variation the variance in disease counts is greater than that which would be expected from a Poisson distribution)

- In other words the data exhibit extra-Poisson variation or **overdispersion** (because of the within area variation the variance in disease counts is greater than that which would be expected from a Poisson distribution)
- One way to allow for **overdispersion** in the Poisson model is to take the ρ_i as **random effects**.
- Treating these parameters as random (rather than fixed) effects introduces an extra source of variability (a **latent** effect) into the model to capture the impact of unknown or unobserved confounding factors

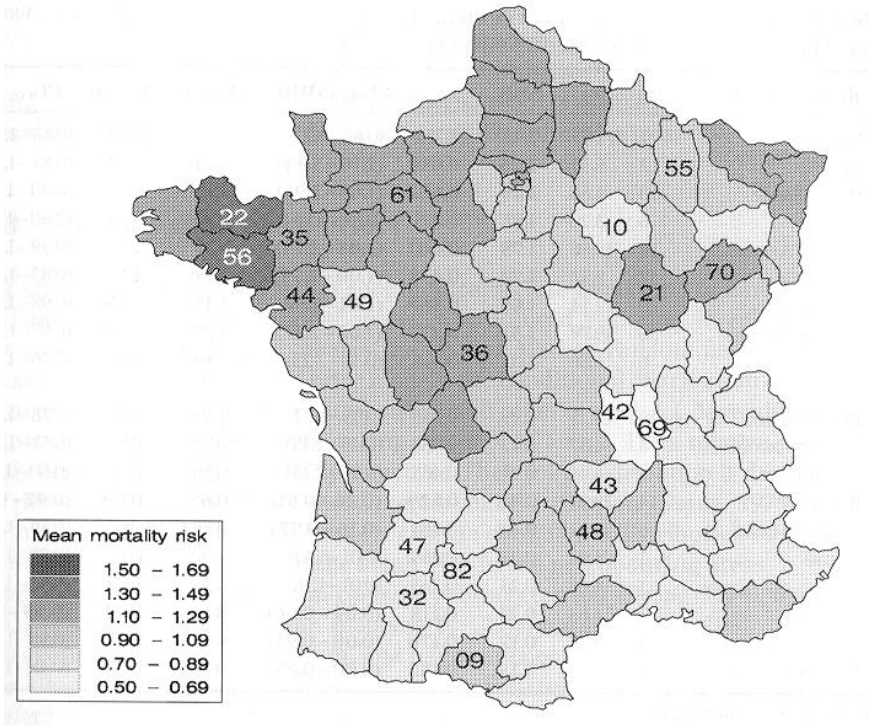
- In other words the data exhibit extra-Poisson variation or **overdispersion** (because of the within area variation the variance in disease counts is greater than that which would be expected from a Poisson distribution)
- One way to allow for **overdispersion** in the Poisson model is to take the ρ_i as **random effects**.
- Treating these parameters as random (rather than fixed) effects introduces an extra source of variability (a **latent** effect) into the model to capture the impact of unknown or unobserved confounding factors
- Essentially they allow the estimate of relative risk for each area to 'borrow strength' from data in other areas leading to a dampening or smoothing of the raw SMRs (often referred to **shrinkage**)

Testis cancer for males in France 1986-1993

SMRS

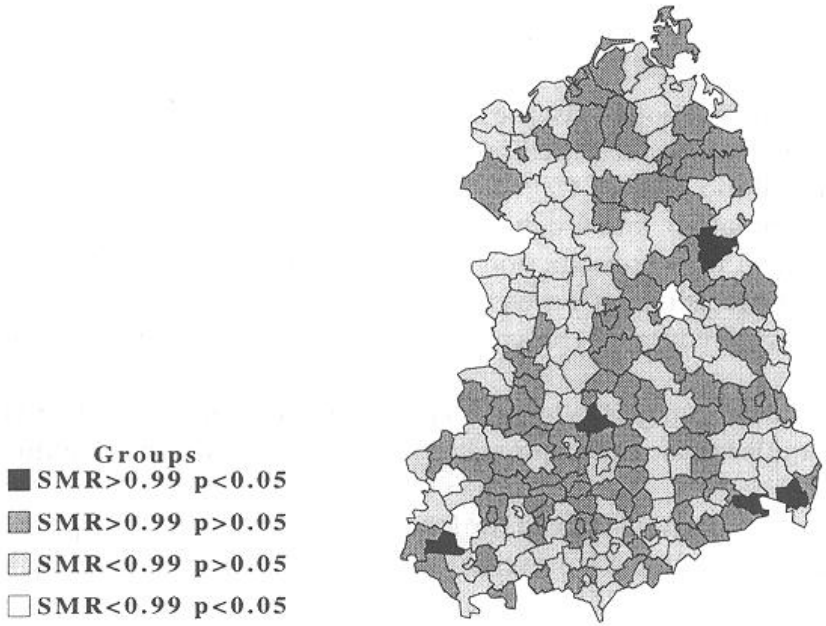


Shrunk rates

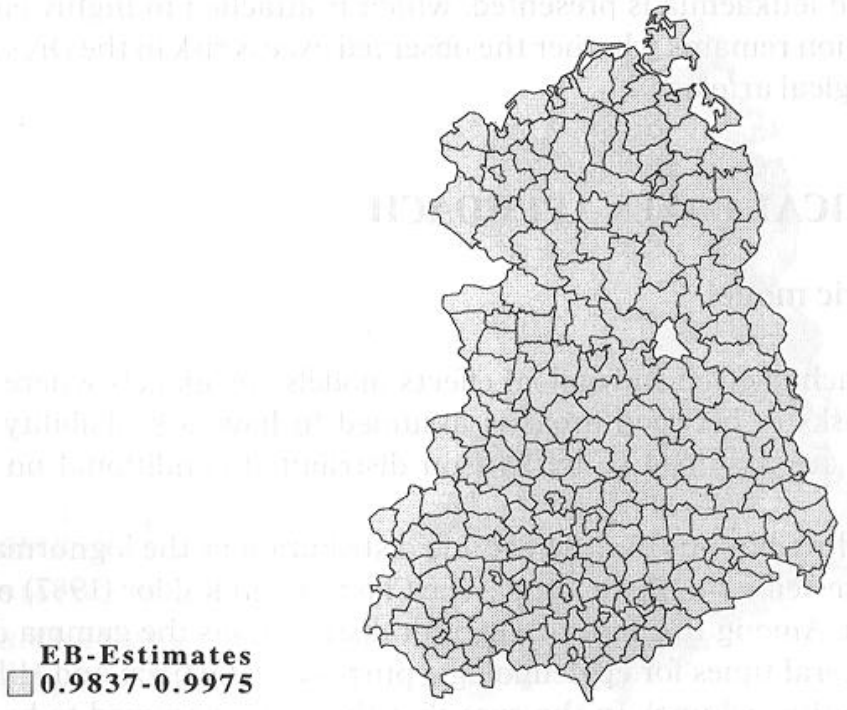


Childhood Leukaemias 1980-1989 in GDR (SMRs)

SMRS



Shrunk rates



➢ With random effects we have a Poisson **Generalised Linear Mixed Model** (**GLMM**) and one approach to fitting such a GLMM is to use a Bayesian framework. Here the simplest Bayesian model is **exchangeable** priors for $\rho_i \sim \text{Gamma}(\psi, \phi)$ (i.e. mean is $\mu_\rho = \frac{\psi}{\phi}$ and variance is $\sigma^2_\rho = \frac{\psi}{\phi^2}$)

- With random effects we have a Poisson **Generalised Linear Mixed Model** (**GLMM**) and one approach to fitting such a GLMM is to use a Bayesian framework. Here the simplest Bayesian model is **exchangeable** priors for $\rho_i \sim$ Gamma(ψ, ϕ) (i.e. mean is $\mu_\rho = \frac{\psi}{\phi}$ and variance is $\sigma^2_\rho = \frac{\psi}{\phi^2}$)
- A Gamma prior combines conveniently with a Poisson likelihood to give a Gamma posterior (it is conjugate to the Poisson) and it may be shown that: $P(\rho_i|\boldsymbol{y}) \sim$ Gamma$(\psi + y_i, \phi + e_i)$. Thus the posterior mean ($\hat{\rho}_i$) is: $\frac{(\psi+y_i)}{(\phi+e_i)}$

Poisson-Gamma Bayesian model for disease mapping

- With random effects we have a Poisson **Generalised Linear Mixed Model** (**GLMM**) and one approach to fitting such a GLMM is to use a Bayesian framework. Here the simplest Bayesian model is **exchangeable** priors for $\rho_i \sim$ Gamma(ψ, ϕ) (i.e. mean is $\mu_\rho = \frac{\psi}{\phi}$ and variance is $\sigma^2_\rho = \frac{\psi}{\phi^2}$)
- A Gamma prior combines conveniently with a Poisson likelihood to give a Gamma posterior (it is conjugate to the Poisson) and it may be shown that: $P(\rho_i|\boldsymbol{y}) \sim$ Gamma$(\psi + y_i, \phi + e_i)$. Thus the posterior mean ($\hat{\rho}_i$) is: $\frac{(\psi+y_i)}{(\phi+e_i)}$
- Therefore in areas with abundant data the posterior mean ($\hat{\rho}_i$) is $\approx \frac{y_i}{e_i}$ (i.e. the SMR) and in areas with sparse data the posterior mean ($\hat{\rho}_i$) is $\approx \frac{\psi}{\phi}$ (i.e. μ_ρ)

Poisson-Gamma Bayesian model for disease mapping

- With random effects we have a Poisson **Generalised Linear Mixed Model** (**GLMM**) and one approach to fitting such a GLMM is to use a Bayesian framework. Here the simplest Bayesian model is **exchangeable** priors for $\rho_i \sim$ Gamma(ψ, ϕ) (i.e. mean is $\mu_\rho = \frac{\psi}{\phi}$ and variance is $\sigma^2_\rho = \frac{\psi}{\phi^2}$)
- A Gamma prior combines conveniently with a Poisson likelihood to give a Gamma posterior (it is conjugate to the Poisson) and it may be shown that: $P(\rho_i|\boldsymbol{y}) \sim$ Gamma$(\psi + y_i, \phi + e_i)$. Thus the posterior mean ($\hat{\rho}_i$) is: $\frac{(\psi+y_i)}{(\phi+e_i)}$
- Therefore in areas with abundant data the posterior mean ($\hat{\rho}_i$) is $\approx \frac{y_i}{e_i}$ (i.e. the SMR) and in areas with sparse data the posterior mean ($\hat{\rho}_i$) is $\approx \frac{\psi}{\phi}$ (i.e. μ_ρ)
- So the relative risk estimates are 'shrunk' towards the global mean with the amount of shrinkage depending upon the hyperparameters ψ and ϕ (or equivalently μ_ρ and σ^2_ρ) which also have to be estimated as part of the model

- To fit this model could use **empirical Bayes** which involves obtaining point estimates for the hyperparameters $\hat{\psi}$ and $\hat{\phi}$ from global aspects of the data and then proceeding as if these quantities are known (see Clayton *al*, 1987)

Poisson-Gamma Bayesian model for disease mapping

- To fit this model could use **empirical Bayes** which involves obtaining point estimates for the hyperparameters $\hat{\psi}$ and $\hat{\phi}$ from global aspects of the data and then proceeding as if these quantities are known (see Clayton *al*, 1987)

- But better to use the full hierarchical Bayesian framework (we have the technology!!) i.e. specify a hyperprior for ψ and for ϕ and then derive a full posterior for these hyperparameters together with the relative risks $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ via use of MCMC applied to:

$$P(\boldsymbol{\rho}, \psi, \phi | \boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\rho}) P(\boldsymbol{\rho}|\psi, \phi) P(\psi) P(\phi)$$

Poisson-Gamma Bayesian model for disease mapping

- To fit this model could use **empirical Bayes** which involves obtaining point estimates for the hyperparameters $\hat{\psi}$ and $\hat{\phi}$ from global aspects of the data and then proceeding as if these quantities are known (see Clayton *al*, 1987)

- But better to use the full hierarchical Bayesian framework (we have the technology!!) i.e. specify a hyperprior for ψ and for ϕ and then derive a full posterior for these hyperparameters together with the relative risks $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)$ via use of MCMC applied to:

$$P(\boldsymbol{\rho}, \psi, \phi | \boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\rho}) P(\boldsymbol{\rho}|\psi, \phi) P(\psi) P(\phi)$$

- In practice suitable hyperpriors for ψ and ϕ would be diffuse exponential distributions

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
y[i] ~ dpois(mu[i])   # Poisson observed counts
mu[i] <- e[i]*rho[i]   # model for Poisson mean
rho[i] ~ dgamma(psi,phi)   # exchangeable prior for relative risks
}
psi ~ dexp(0.1)   # diffuse exponential hyperprior for psi
phi ~ dexp(0.1)   # diffuse exponential hyperprior for phi
mu.rho <- psi/phi   # mean of prior for relative risks
sigma.rho <- psi/pow(phi,2)   # variance of prior for relative risks
```

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
y[i] ~ dpois(mu[i])   # Poisson observed counts
mu[i] <- e[i]*rho[i]   # model for Poisson mean
rho[i] ~ dgamma(psi,phi)   # exchangeable prior for relative risks
}
psi ~ dexp(0.1)   # diffuse exponential hyperprior for psi
phi ~ dexp(0.1)   # diffuse exponential hyperprior for phi
mu.rho <- psi/phi   # mean of prior for relative risks
sigma.rho <- psi/pow(phi,2)   # variance of prior for relative risks
```

As initial values we might take $\psi = 0.1, \phi = 0.1$ and $\rho_i = 1, i = 1, \ldots, n$. ☞

Poisson-Log Normal Bayesian model for disease mapping

- A Gamma prior for ρ_i is mathematically convenient, but may be restrictive:
  - Covariate adjustment is difficult (i.e. ecological (correlation) studies)
  - Not easy to relax the independence of the ρ_i — risks in nearby areas may be spatially correlated (particularly if geographical trends or clusters in risk exist)

Poisson-Log Normal Bayesian model for disease mapping

- A Gamma prior for ρ_i is mathematically convenient, but may be restrictive:
  - Covariate adjustment is difficult (i.e. ecological (correlation) studies)
  - Not easy to relax the independence of the ρ_i — risks in nearby areas may be spatially correlated (particularly if geographical trends or clusters in risk exist)
- In practice a hierarchical Poisson-log normal formulation is more flexible i.e.:

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i) \\ \log \mu_i &= \log e_i + \log \rho_i = \log e_i + \theta_i \\ \theta_i &\sim \text{Normal}(\mu_\theta, \sigma_\theta^2) \end{aligned}$$

(so θ_i are exchangeable and relative risks are now $\rho_i = \exp(\theta_i)$)

Poisson-Log Normal Bayesian model for disease mapping

- A Gamma prior for ρ_i is mathematically convenient, but may be restrictive:
  - Covariate adjustment is difficult (i.e. ecological (correlation) studies)
  - Not easy to relax the independence of the ρ_i — risks in nearby areas may be spatially correlated (particularly if geographical trends or clusters in risk exist)
- In practice a hierarchical Poisson-log normal formulation is more flexible i.e.:

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i) \\ \log \mu_i &= \log e_i + \log \rho_i = \log e_i + \theta_i \\ \theta_i &\sim \text{Normal}(\mu_\theta, \sigma_\theta^2) \end{aligned}$$

  (so θ_i are exchangeable and relative risks are now $\rho_i = \exp(\theta_i)$)
- Typical 'non informative' hyperpriors are a diffuse Normal distribution (zero mean large variance) for μ_θ and a diffuse Gamma for the **precision** $\tau_\theta = 1/\sigma_\theta^2$.

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
y[i] ~ dpois(mu[i])   # Poisson observed counts
log(mu[i])<- log(e[i]) + theta[i]   # model Poisson mean
theta[i]~dnorm(mu.theta, tau.theta)   # exchangeable prior logRR
rho[i]<-exp(theta[i])   # modelled relative risks
}
mu.theta~dnorm(0,1.0E-6)   # normal hyperprior for mu.theta
tau.theta~dgamma(0.5, 0.0005)   # gamma hyperprior for tau.theta
sigma.theta<-sqrt(1/tau.theta)   # st dev derived from tau.theta
```

The relevant **WinBUGS** model is:

for (i in 1 : N) {

y[i] ∼ dpois(mu[i]) # Poisson observed counts

log(mu[i])← log(e[i]) + theta[i] # model Poisson mean

theta[i]∼dnorm(mu.theta, tau.theta) # exchangeable prior logRR

rho[i]←exp(theta[i]) # modelled relative risks

}

mu.theta∼dnorm(0,1.0E-6) # normal hyperprior for mu.theta

tau.theta∼dgamma(0.5, 0.0005) # gamma hyperprior for tau.theta

sigma.theta←sqrt(1/tau.theta) # st dev derived from tau.theta

As initial values we might take $\mu_\theta = 0$, $\tau_\theta = 1$ and $\theta_i = 0$, for $i = 1, \ldots, n$. ☞

- The model considered so far allows for overdispersion in the Poisson distribution of counts y_i (via the random effects) but it does not allow for explicit spatial dependence between the y_i. This may also be present (e.g. arising through lesser variability of rates in neighbouring densely populated urban areas as opposed to sparsely populated rural areas, or an infectious aetiology)

- The model considered so far allows for overdispersion in the Poisson distribution of counts y_i (via the random effects) but it does not allow for explicit spatial dependence between the y_i. This may also be present (e.g. arising through lesser variability of rates in neighbouring densely populated urban areas as opposed to sparsely populated rural areas, or an infectious aetiology)
- Can include such dependence by splitting random effect θ_i in the Poisson-log normal model into a **spatially unstructured** and a **spatially structured** term

Model extensions for spatial structure

- The model considered so far allows for overdispersion in the Poisson distribution of counts y_i (via the random effects) but it does not allow for explicit spatial dependence between the y_i. This may also be present (e.g. arising through lesser variability of rates in neighbouring densely populated urban areas as opposed to sparsely populated rural areas, or an infectious aetiology)
- Can include such dependence by splitting random effect θ_i in the Poisson-log normal model into a **spatially unstructured** and a **spatially structured** term
- θ_i is replaced by $\alpha + \phi_i + \nu_i$ where α is the mean log relative risk over all areas (i.e. our earlier μ_θ), ϕ_i a zero mean spatially unstructured (or exchangeable) log relative risk of area i compared to the map as a whole, and ν_i is corresponding spatially structured (**non-exchangeable**) random effect.
- This model is often termed a **convolution model**

➢ A typical choice for the spatially structured prior for ν_i is a conditional intrinsic Gaussian autoregressive model (**CAR**) (see Besag *et al*, 1995) i.e.:

$$\nu_i \big| \nu_{j \neq i} \sim \mathsf{N}\left(\frac{\sum_{j \neq i} w_{ij} \nu_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_\nu^2}{\sum_{j \neq i} w_{ij}} \right)$$

here w_{ij} are suitable **adjacency weights** for the areas and the new hyperparameter σ_ν controls the strength of local spatial dependence.

- A typical choice for the spatially structured prior for ν_i is a conditional intrinsic Gaussian autoregressive model (**CAR**) (see Besag *et al*, 1995) i.e.:

$$\nu_i | \nu_{j \neq i} \sim \mathrm{N}\left(\frac{\sum_{j \neq i} w_{ij} \nu_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_\nu^2}{\sum_{j \neq i} w_{ij}} \right)$$

  here w_{ij} are suitable **adjacency weights** for the areas and the new hyperparameter σ_ν controls the strength of local spatial dependence.

- Often w_{ij} are taken as simple binary values—$w_{ij} = 1$ if area i has common boundary with area j, $w_{ij} = 0$ otherwise.

- A typical choice for the spatially structured prior for ν_i is a conditional intrinsic Gaussian autoregressive model (**CAR**) (see Besag *et al*, 1995) i.e.:

$$\nu_i | \nu_{j \neq i} \sim \mathrm{N}\left(\frac{\sum_{j \neq i} w_{ij} \nu_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_\nu^2}{\sum_{j \neq i} w_{ij}} \right)$$

  here w_{ij} are suitable **adjacency weights** for the areas and the new hyperparameter σ_ν controls the strength of local spatial dependence.

- Often w_{ij} are taken as simple binary values—$w_{ij} = 1$ if area i has common boundary with area j, $w_{ij} = 0$ otherwise.

- Similar to before, the prior for ϕ_i is $\phi_i \sim N(0, \sigma_\phi^2)$. The prior for α is now taken as $\alpha \sim \mathrm{U}(-\infty, +\infty)$ to allow for the fact that the CAR is **improper** (has undefined mean) and so a 'sum to zero' constraint needs to applied to the ν_i.

➢ So the full hierarchical model is now:

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i) \\
\log \mu_i &= \log e_i + \log \rho_i = \log e_i + \alpha + \phi_i + \nu_i \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2)
\end{aligned}
$$

➢ So the full hierarchical model is now:

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i) \\ \log \mu_i &= \log e_i + \log \rho_i = \log e_i + \alpha + \phi_i + \nu_i \\ \alpha &\sim \text{U}(-\infty, +\infty) \\ \phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\ \nu_i &\sim \text{CAR}(\sigma_\nu^2) \end{aligned}$$

➢ To complete the specification diffuse gamma hyperpriors are assumed for precisions corresponding to both hyperparameters i.e. for $\tau_\phi = 1/\sigma_\phi^2$ and for $\tau_\nu = 1/\sigma_\nu^2$

Spatially structured Poisson-Log Normal model: Larynx Cancer in Mersey & W Lancashire

➢ The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
 y[i]   ~ dpois(mu[i])                          # Poisson likelihood for observed counts
 log(mu[i]) <- log(e[i])+alpha+phi[i]+nu[i]   # model for Poisson mean
 phi[i] ~ dnorm(0, tau.phi)                     # normal prior for spatially unstructured effects
 rho[i] <- exp(alpha+phi[i]+nu[i])              # R Risks compared to reference rate
 rholocal[i] <- exp(phi[i]+nu[i])               # R Risks compared to overall risk in study area
 Phigh[i] <- step(rholocal[i] - 1.5)            # Prob that local rholocal[i] > 1.5 (note how easy this is!)
}
nu[1:N] ~ car.normal(adj[],weights[],num[],tau.nu) #CAR prior for spatially structured effects
alpha   ~ dflat()                  # uniform prior for mean log relative risk
tau.phi   ~ dgamma(0.5, 0.0005)    # diffuse gamma hyperprior for tau.phi
tau.nu   ~  dgamma(0.5, 0.0005)    # diffuse gamma hyperprior for tau.nu
sigma.phi <- sqrt(1/tau.phi)       # st dev of prior for spatially unstructured effects
sigma.nu <- sqrt(1/tau.nu)         # st dev of prior for spatially structured effects
```

Spatially structured Poisson-Log Normal model: Larynx Cancer in Mersey & W Lancashire

➢ The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
 y[i]   ~ dpois(mu[i])                          # Poisson likelihood for observed counts
 log(mu[i]) <- log(e[i])+alpha+phi[i]+nu[i]     # model for Poisson mean
 phi[i] ~ dnorm(0, tau.phi)                     # normal prior for spatially unstructured effects
 rho[i] <- exp(alpha+phi[i]+nu[i])              # R Risks compared to reference rate
 rholocal[i] <- exp(phi[i]+nu[i])               # R Risks compared to overall risk in study area
 Phigh[i] <- step(rholocal[i] - 1.5)            # Prob that local rholocal[i] > 1.5 (note how easy this is!)
}
nu[1:N] ~ car.normal(adj[],weights[],num[],tau.nu) #CAR prior for spatially structured effects
alpha   ~ dflat()                  # uniform prior for mean log relative risk
tau.phi   ~ dgamma(0.5, 0.0005)    # diffuse gamma hyperprior for tau.phi
tau.nu   ~  dgamma(0.5, 0.0005)    # diffuse gamma hyperprior for tau.nu
sigma.phi <- sqrt(1/tau.phi)       # st dev of prior for spatially unstructured effects
sigma.nu <- sqrt(1/tau.nu)         # st dev of prior for spatially structured effects
```

➢ Initial values take: $\alpha = 0$, $\tau_\phi = \tau_\nu = 1$, and $\phi_i = \nu_i = 0$, $i = 1, \ldots, n$. ☞

Ecological (correlation) studies

Ecological (correlation) studies

- Refers to investigations where the focus is on examining associations between disease incidence and risk factors measured on groups (we have already mentioned the so-called **ecological fallacy**)

Ecological (correlation) studies

- Refers to investigations where the focus is on examining associations between disease incidence and risk factors measured on groups (we have already mentioned the so-called **ecological fallacy**)
- As usual there are various approaches (see references). But we focus on extensions to the Bayesian hierarchical models we employed in disease mapping

Ecological (correlation) studies

- Refers to investigations where the focus is on examining associations between disease incidence and risk factors measured on groups (we have already mentioned the so-called **ecological fallacy**)
- As usual there are various approaches (see references). But we focus on extensions to the Bayesian hierarchical models we employed in disease mapping
- Following the basic model development we then consider a number of further issues concerned with such models and their interpretation, for example the handling of censored values and missing values, predictive distributions and correction for specification bias and measurement error

- We use a straightforward extension of the disease mapping model discussed earlier to include p covariates $(x_{i1} \ldots, x_{ip})$ measured in each area i i.e. $y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$ with:

$$\log \mu_i = \log e_i + \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \phi_i + \nu_i$$

note overall relative risks are now $\rho_i = \exp(\alpha + \sum_j \beta_j x_{ij} + \phi_i + \nu_i)$ and $\exp(\alpha + \phi_i + \nu_i)$ is the **residual relative risk** after 'correcting' for the covariates.

- We use a straightforward extension of the disease mapping model discussed earlier to include p covariates $(x_{i1} \ldots, x_{ip})$ measured in each area i i.e. $y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$ with:

$$\log \mu_i = \log e_i + \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \phi_i + \nu_i$$

note overall relative risks are now $\rho_i = \exp(\alpha + \sum_j \beta_j x_{ij} + \phi_i + \nu_i)$ and $\exp(\alpha + \phi_i + \nu_i)$ is the **residual relative risk** after 'correcting' for the covariates.

- Priors and hyperpriors relating to ϕ_i, ν_i and α are as before. Non-informative Normal priors (zero mean large variances) are adopted for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$.

- We use a straightforward extension of the disease mapping model discussed earlier to include p covariates $(x_{i1} \ldots, x_{ip})$ measured in each area i i.e. $y_i \sim \text{Poisson}(\mu_i) = \text{Poisson}(e_i \rho_i)$ with:

$$\log \mu_i = \log e_i + \alpha + \sum_{j=1}^{p} \beta_j x_{ij} + \phi_i + \nu_i$$

  note overall relative risks are now $\rho_i = \exp(\alpha + \sum_j \beta_j x_{ij} + \phi_i + \nu_i)$ and $\exp(\alpha + \phi_i + \nu_i)$ is the **residual relative risk** after 'correcting' for the covariates.

- Priors and hyperpriors relating to ϕ_i, ν_i and α are as before. Non-informative Normal priors (zero mean large variances) are adopted for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$.

- Then use MCMC to obtain samples from $P(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\nu}, \tau_\phi, \tau_\nu | \boldsymbol{y})$ where hyperparameters $\tau_\phi = 1/\sigma_\phi^2$ and $\tau_\nu = 1/\sigma_\nu^2$ again refer to the precisions of the priors for spatially unstructured and spatially structured random effects ϕ_i and ν_i.

Prostate cancer mortality in Spanish provinces

- As an example consider a study on relationship of prostate cancer mortality in Spanish provinces to nitrate concentrations in drinking water

Prostate cancer mortality in Spanish provinces

- As an example consider a study on relationship of prostate cancer mortality in Spanish provinces to nitrate concentrations in drinking water
- model used was: $\log \mu_i = \log p_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \phi_i + \nu_i$ where in area i: p_i is population, x_{i1} is proportion of population over 40 and x_{i2} is nitrate concentration in drinking water. Note that here the direct standardisation term $\log(e_i)$ has been dropped in favour of an **indirect standardisation**—i.e. relevant age/sex specific population measures are included amongst the covariates

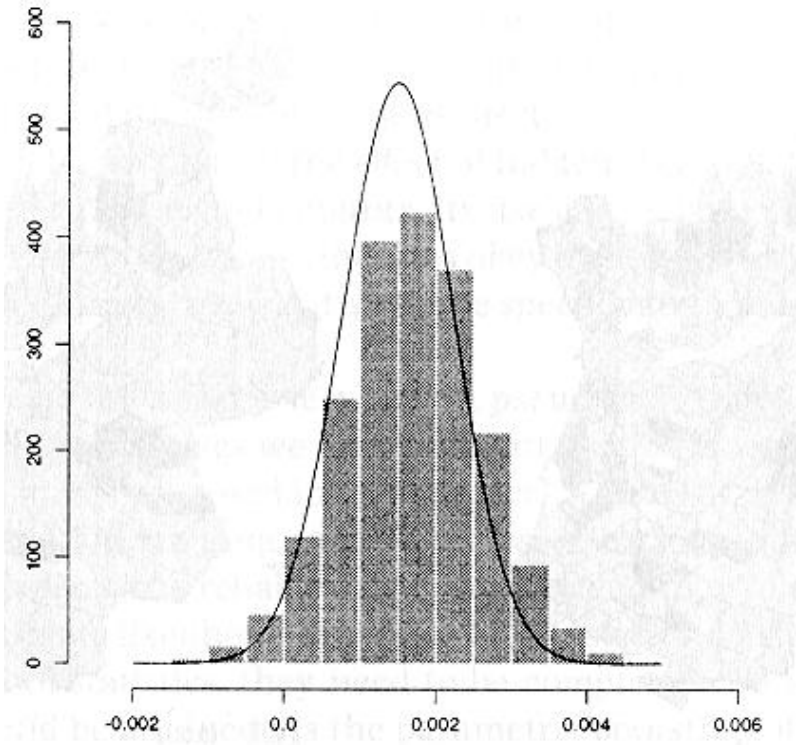
Prostate cancer mortality in Spanish provinces

- As an example consider a study on relationship of prostate cancer mortality in Spanish provinces to nitrate concentrations in drinking water
- model used was: $\log \mu_i = \log p_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \phi_i + \nu_i$ where in area i: p_i is population, x_{i1} is proportion of population over 40 and x_{i2} is nitrate concentration in drinking water. Note that here the direct standardisation term $\log(e_i)$ has been dropped in favour of an **indirect standardisation**—i.e. relevant age/sex specific population measures are included amongst the covariates
- Results showed the posterior **credible interval** for β_2 did not contain zero in the absence of the ν_i term, but when this term is present in the model then the β_2 posterior credible interval did contain zero
- There is therefore no clear evidence of the nitrate effect, but it cannot be entirely ruled out.

Prostate cancer mortality in Spanish provinces

- As an example consider a study on relationship of prostate cancer mortality in Spanish provinces to nitrate concentrations in drinking water
- model used was: $\log \mu_i = \log p_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \phi_i + \nu_i$ where in area i: p_i is population, x_{i1} is proportion of population over 40 and x_{i2} is nitrate concentration in drinking water. Note that here the direct standardisation term $\log(e_i)$ has been dropped in favour of an **indirect standardisation**—i.e. relevant age/sex specific population measures are included amongst the covariates
- Results showed the posterior **credible interval** for β_2 did not contain zero in the absence of the ν_i term, but when this term is present in the model then the β_2 posterior credible interval did contain zero
- There is therefore no clear evidence of the nitrate effect, but it cannot be entirely ruled out. Remember —*absence of evidence is not evidence of absence*

Prostate cancer mortality in Spanish provinces Posterior distribution of nitrate coefficient β_2



- The preceding model allows for differences in areas through a combination of unstructured and spatially structured random effects, but the nature of the relationships between relative risk and the ecological covariates is assumed homogeneous over the study region—there is no local variation in β

- The preceding model allows for differences in areas through a combination of unstructured and spatially structured random effects, but the nature of the relationships between relative risk and the ecological covariates is assumed homogeneous over the study region—there is no local variation in $\boldsymbol{\beta}$
- An alternative perspective on spatial heterogeneity is to allow non constant covariate coefficients over the study region i.e. allow $\boldsymbol{\beta}$ to be area specific. In the case of a single covariate, x_i, a suitable model might be:

$$\log \mu_i = \log e_i + \alpha + \beta_i x_i + \phi_i + \nu_i$$

with exchangeable priors $\beta_i \sim \text{Normal}(\mu_\beta, \sigma^2_\beta)$, $i = 1, \ldots, n$ and other priors as before. Here μ_β represents the average relationship with x_i over the region

- The preceding model allows for differences in areas through a combination of unstructured and spatially structured random effects, but the nature of the relationships between relative risk and the ecological covariates is assumed homogeneous over the study region—there is no local variation in $\boldsymbol{\beta}$
- An alternative perspective on spatial heterogeneity is to allow non constant covariate coefficients over the study region i.e. allow $\boldsymbol{\beta}$ to be area specific. In the case of a single covariate, x_i, a suitable model might be:

$$\log \mu_i = \log e_i + \alpha + \beta_i x_i + \phi_i + \nu_i$$

  with exchangeable priors $\beta_i \sim$ Normal$(\mu_\beta, \sigma^2_\beta)$, $i = 1, \ldots, n$ and other priors as before. Here μ_β represents the average relationship with x_i over the region
- Note **identifiability** may be a problem in such a model (inability to uniquely distinguish between certain parameters because an exactly identical set of outcomes can arise from more than one set of parameter values). Some parameter constraints may be needed.

➢ If the area specific relationships in the previous model are expected to be differentiated in a spatially distinct pattern (i.e. similar relationships are spatially clustered) then we can use a model such as:

$$\log \mu_i = \log e_i + \alpha + \beta x_i + \beta_i x_i + \phi_i + \nu_i$$

with the β_i assumed to be spatially dependent i.e. non-exchangeable priors $\beta_i \sim \text{CAR}(\sigma^2_\beta), i = 1, \ldots, n$.

- If the area specific relationships in the previous model are expected to be differentiated in a spatially distinct pattern (i.e. similar relationships are spatially clustered) then we can use a model such as:

$$\log \mu_i = \log e_i + \alpha + \beta x_i + \beta_i x_i + \phi_i + \nu_i$$

with the β_i assumed to be spatially dependent i.e. non-exchangeable priors $\beta_i \sim \text{CAR}(\sigma^2_\beta), i = 1, \ldots, n$.

- The βx_i term (with prior as $\beta \sim \text{U}(-\infty, +\infty)$) is included in the model to represent the overall global relationship since the CAR is improper and a sum to zero constraint will need to be imposed on β_i. The β_i therefore now represent deviations from the overall relationship.

- Clearly identifiability problems are compounded with such a model and further issues arise relating to potential confounding between the spatially dependent area-specific coefficients β_i and the spatially dependent random effects ν_i.

- Clearly identifiability problems are compounded with such a model and further issues arise relating to potential confounding between the spatially dependent area-specific coefficients β_i and the spatially dependent random effects ν_i.
- For these reasons alternative less direct formulations have been suggested which avoid the CAR and instead incorporate a multivariate set of underlying unstructured random effects which induce spatial dependence in the β_i and the ν_i by being linked to them via scaled adjacency weighting systems (see Congdon, 2003; Leyland *et al*, 2000).
- Such an approach is particularly useful when models involving area-specific coefficients for more than one explanatory variable need to be considered, since the CAR formulation is difficult to extend to this case.

- The **spatial expansion model** and **geographically weighted regression** (GWR) represent examples of an entirely different (non Bayesian) approach to estimating area specific covariate coefficients
- Rather than use a single model, such approaches instead reuse the data n times, with the ith regression being considered to be 'centred' on the ith area.

- The **spatial expansion model** and **geographically weighted regression** (GWR) represent examples of an entirely different (non Bayesian) approach to estimating area specific covariate coefficients
- Rather than use a single model, such approaches instead reuse the data n times, with the ith regression being considered to be 'centred' on the ith area.
- For example, GWR essentially consists of performing n weighted regressions, with the ith of these being 'centred' on the ith area and using weights on data points inversely proportional to their distance from i (see Brunsdon *et al*, 1998). For more details of the spatial expansion model (see Casetti, 1992)

- The **spatial expansion model** and **geographically weighted regression** (GWR) represent examples of an entirely different (non Bayesian) approach to estimating area specific covariate coefficients
- Rather than use a single model, such approaches instead reuse the data n times, with the ith regression being considered to be 'centred' on the ith area.
- For example, GWR essentially consists of performing n weighted regressions, with the ith of these being 'centred' on the ith area and using weights on data points inversely proportional to their distance from i (see Brunsdon *et al*, 1998). For more details of the spatial expansion model (see Casetti, 1992)
- Note that the terminology 'geographically weighted regression' is sometimes now used to refer generally to any spatial regression model with area-specific covariate coefficients and not just the Brunsdon method from which the name originated

Leprosy surveillance in Olinda (Brazil) 1991-1995

- As a more extended example of the use of Bayesian ecological models let us consider application of the basic model (with non area specific covariate coefficients) to the data on leprosy incidence from Olinda in Brazil. This example will also allow us to explore how our previous Bayesian models can be extended to handle censored (and missing) data values

Leprosy surveillance in Olinda (Brazil) 1991-1995

- As a more extended example of the use of Bayesian ecological models let us consider application of the basic model (with non area specific covariate coefficients) to the data on leprosy incidence from Olinda in Brazil. This example will also allow us to explore how our previous Bayesian models can be extended to handle censored (and missing) data values
- There is *a priori* reason to believe that numbers of leprosy cases will be higher in the poorer and more socially deprived areas. To allow for this we include a single covariate x_i in the model — the proportion of heads of household with monthly income below one minimum legal wage (approximately US$80).

Leprosy surveillance in Olinda (Brazil) 1991-1995

- As a more extended example of the use of Bayesian ecological models let us consider application of the basic model (with non area specific covariate coefficients) to the data on leprosy incidence from Olinda in Brazil. This example will also allow us to explore how our previous Bayesian models can be extended to handle censored (and missing) data values
- There is *a priori* reason to believe that numbers of leprosy cases will be higher in the poorer and more socially deprived areas. To allow for this we include a single covariate x_i in the model — the proportion of heads of household with monthly income below one minimum legal wage (approximately US$80).
- Expected cases e_i in each area are derived from the population at risk and the global leprosy detection rate over the whole study area

Leprosy surveillance in Olinda (Brazil) 1991-1995

- As a more extended example of the use of Bayesian ecological models let us consider application of the basic model (with non area specific covariate coefficients) to the data on leprosy incidence from Olinda in Brazil. This example will also allow us to explore how our previous Bayesian models can be extended to handle censored (and missing) data values
- There is *a priori* reason to believe that numbers of leprosy cases will be higher in the poorer and more socially deprived areas. To allow for this we include a single covariate x_i in the model — the proportion of heads of household with monthly income below one minimum legal wage (approximately US$80).
- Expected cases e_i in each area are derived from the population at risk and the global leprosy detection rate over the whole study area
- In the CAR definition, w_{ij} are taken as the standard binary adjacency weights.

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
   y[i] ~ dpois(mu[i])   # Poisson counts
   log(mu[i]) <- log(e[i]) + alpha + beta * x[i] + phi[i] + nu[i]   # model for mean
   phi[i] ~ dnorm(0.0, tau.phi)  # prior for phi
   rho[i] <- exp(alpha+beta*x[i]+phi[i]+nu[i])   # Leprosy relative risks
}
nu[1:N] ~ car.normal(adj[], weights[], num[], tau.nu)   # CAR prior for nu
alpha ~ dflat()   # prior for alpha
beta ~ dnorm(0.0, 1.0E-5)   # prior for beta
tau.phi ~ dgamma(1.0E-3, 1.0E-3)   # hyperprior for tau.phi
tau.nu ~ dgamma(1.0E-3, 1.0E-3)   # hyperprior for tau.nu
sigma.phi <- 1 / sqrt(tau.phi)   # st dev of prior for unstructured rand effects
sigma.nu <- 1 / sqrt(tau.nu)   # st dev of prior for structured rand effects
```

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
   y[i] ~ dpois(mu[i])   # Poisson counts
   log(mu[i]) <- log(e[i]) + alpha + beta * x[i] + phi[i] + nu[i]   # model for mean
   phi[i] ~ dnorm(0.0, tau.phi)  # prior for phi
   rho[i] <- exp(alpha+beta*x[i]+phi[i]+nu[i])   # Leprosy relative risks
}
nu[1:N] ~ car.normal(adj[], weights[], num[], tau.nu)   # CAR prior for nu
alpha ~ dflat()   # prior for alpha
beta ~ dnorm(0.0, 1.0E-5)   # prior for beta
tau.phi ~ dgamma(1.0E-3, 1.0E-3)   # hyperprior for tau.phi
tau.nu ~ dgamma(1.0E-3, 1.0E-3)   # hyperprior for tau.nu
sigma.phi <- 1 / sqrt(tau.phi)   # st dev of prior for unstructured rand effects
sigma.nu <- 1 / sqrt(tau.nu)   # st dev of prior for structured rand effects
```

As initial values we take $\alpha = \beta = 0$, $\tau_\phi = \tau_\nu = 1$, and $\phi_i = \nu_i = 0$, $i = 1, \ldots, n$.

☞

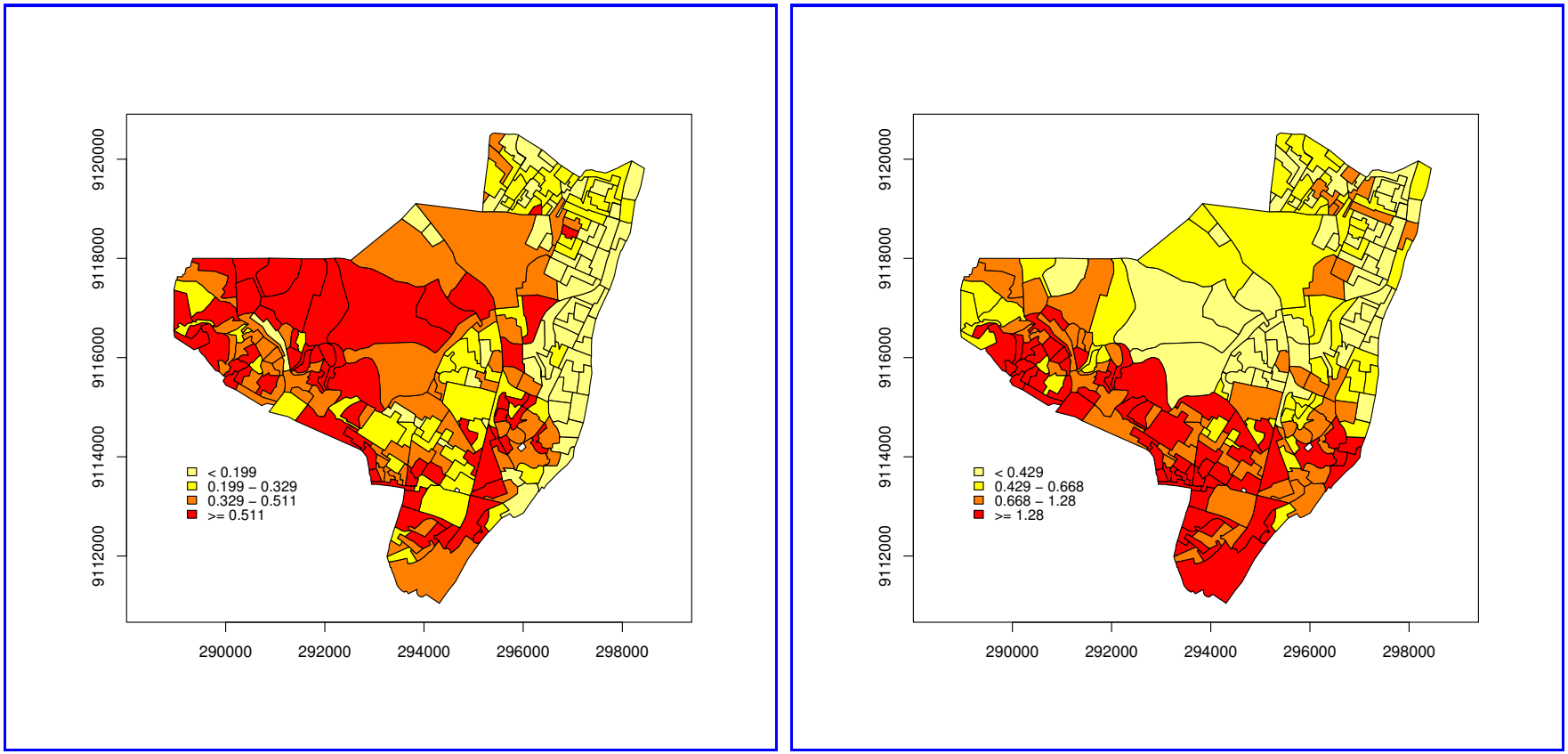
Leprosy surveillance in Olinda 1991-1995

➢ MCMC (10,000 samples with 'burn in' of 5000 and thinning of 10) provides following posterior mean estimates for a selection of the parameters

Model	$\hat{\alpha}$		$\hat{\beta}$	
	mean	95% cred int	mean	95% cred int
1991-1995 std.	-0.5	(-0.6, -0.2)	0.4	(0.1, 1.2)

Model	$\hat{\sigma}_\phi$		$\hat{\sigma}_\nu$	
	mean	sd	mean	sd
1991-1995 std.	0.4	0.1	1.0	0.2

Olinda deprivation (left) and leprosy relative risk estimated from 'standard' model (right)



Leprosy surveillance in Olinda 1991-1995

- Could be some problems with this analysis
  - A number of areas exhibit contrasting and counter intuitive extremes of high deprivation scores combined with low relative risk of leprosy.

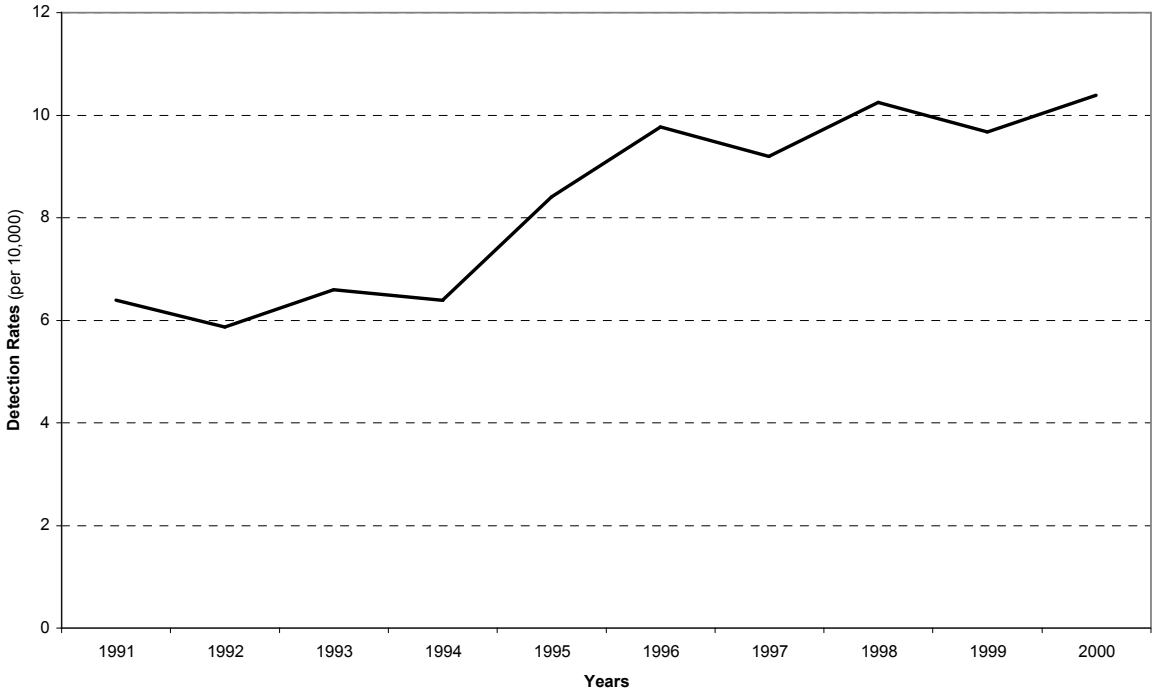
- Could be some problems with this analysis
  - A number of areas exhibit contrasting and counter intuitive extremes of high deprivation scores combined with low relative risk of leprosy.
  - The estimate of β is not convincingly different from zero, a result which is surprising given strong *a priori* reasons for the belief that leprosy rates will be higher in the more socio-economically deprived areas.

Leprosy surveillance in Olinda 1991-1995

- Could be some problems with this analysis
  - A number of areas exhibit contrasting and counter intuitive extremes of high deprivation scores combined with low relative risk of leprosy.
  - The estimate of β is not convincingly different from zero, a result which is surprising given strong *a priori* reasons for the belief that leprosy rates will be higher in the more socio-economically deprived areas.
- These observations suggest some differences in the quality of data from area to area. It could be that there is significant under-detection of cases in the poorer areas during the period 1991-1995.

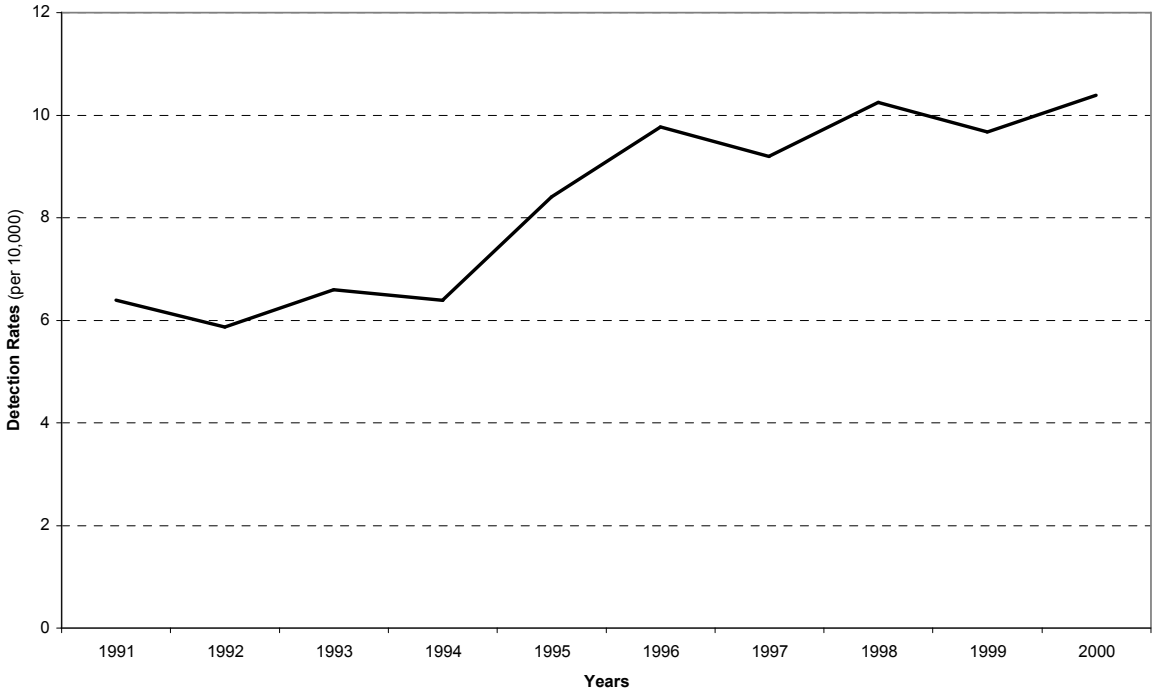
Leprosy Detection Rates between 1991 and 2000

This suspicion is confirmed by also looking at more recent detection rates in the period 1996-2000.



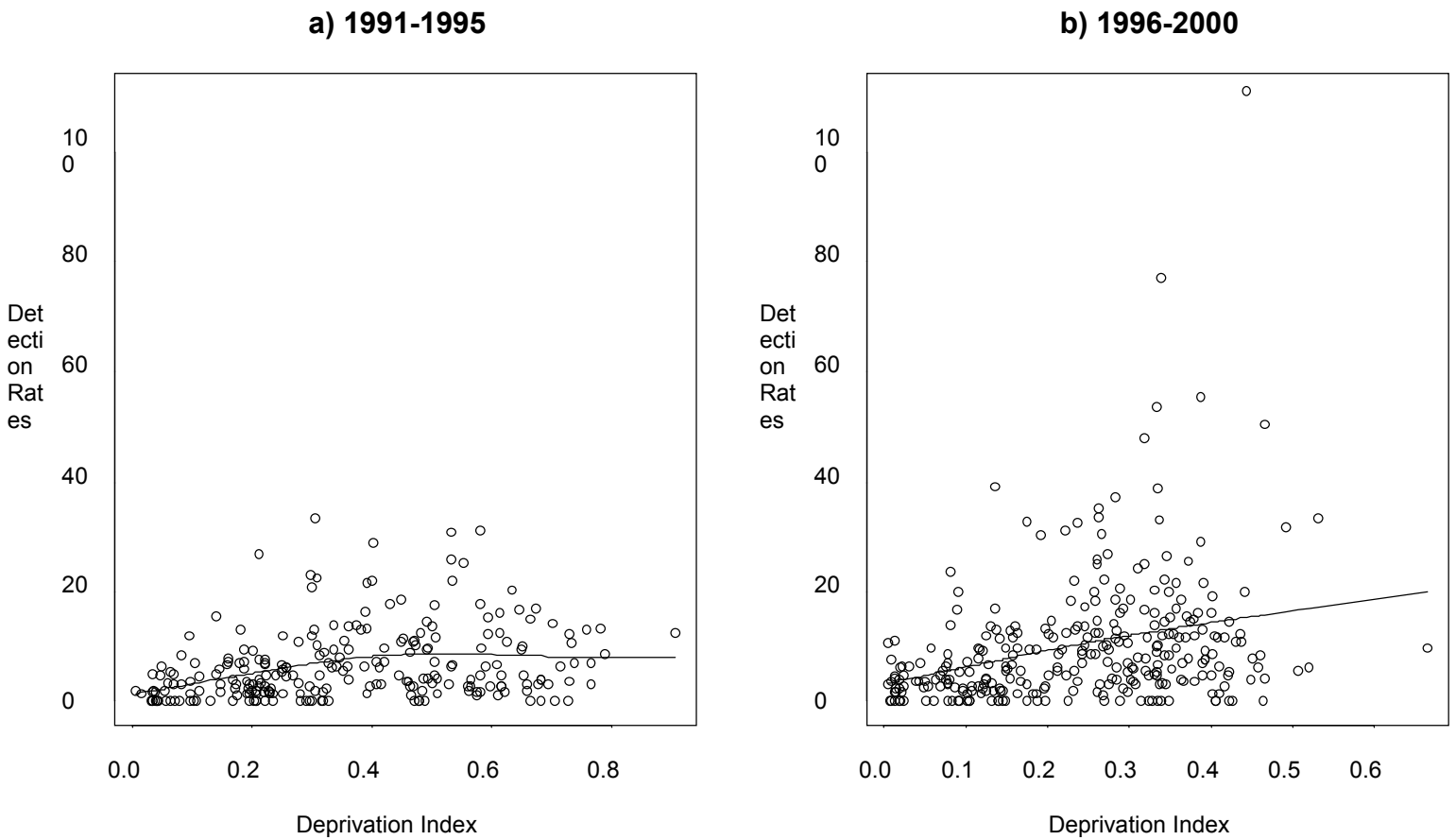
Leprosy Detection Rates between 1991 and 2000

This suspicion is confirmed by also looking at more recent detection rates in the period 1996-2000.



There was a significant extension of the coverage and efficacy of the control programme in 1995 and so the subsequent period should more accurately reflect the true picture regarding numbers of cases

Leprosy detection rates versus deprivation index in the two periods with superimposed non parametric smoothed line



Censored model for Leprosy in Olinda (Brazil) 1991-1995

- One way to handle possible under-detection is to treat number of cases in the 1991-1995 data as **censored** in certain areas and use the corresponding observed counts as lower bounds for the true disease counts

Censored model for Leprosy in Olinda (Brazil) 1991-1995

- One way to handle possible under-detection is to treat number of cases in the 1991-1995 data as **censored** in certain areas and use the corresponding observed counts as lower bounds for the true disease counts
- The dividing line between reliable and unreliable disease counts might perhaps best be left to experience with the surveillance system and local researchers suggest that the number of leprosy cases in the 1991-1995 period should be treated as suspect where over 60% of population receive an income of less than one minimum wage (consistent with the observed “flattening” of increase in log relative risk with deprivation score which is observed in the 1991-1995 period)

Censored model for Leprosy in Olinda (Brazil) 1991-1995

- One way to handle possible under-detection is to treat number of cases in the 1991-1995 data as **censored** in certain areas and use the corresponding observed counts as lower bounds for the true disease counts
- The dividing line between reliable and unreliable disease counts might perhaps best be left to experience with the surveillance system and local researchers suggest that the number of leprosy cases in the 1991-1995 period should be treated as suspect where over 60% of population receive an income of less than one minimum wage (consistent with the observed “flattening” of increase in log relative risk with deprivation score which is observed in the 1991-1995 period)
- Some 16% of the areas in the study region fall into the suspect category under this assumption. and some of the poorest of these contain examples of ‘favelas’

Censored model for Leprosy in Olinda 1991-1995

- Using this 60% cut-off as a working assumption (could obviously experiment with alternatives) we then have a need for a model that can incorporate censoring and this provides an example of how relatively straightforward it is to handle censored values in the Bayesian framework more generally

Censored model for Leprosy in Olinda 1991-1995

- Using this 60% cut-off as a working assumption (could obviously experiment with alternatives) we then have a need for a model that can incorporate censoring and this provides an example of how relatively straightforward it is to handle censored values in the Bayesian framework more generally
- The basic model for the leprosy counts remains the same, but now data is: $(\boldsymbol{y}, \boldsymbol{y}^*) = (y_1, \ldots, y_m, y^*_{m+1}, \ldots, y^*_n)$, where y^*_i refer to the censored data values in the areas where the deprivation indicator exceeds 60% (for convenience we reorder the data by deprivation score).
- So now likelihood is: $P(y_1, \ldots, y_m|\boldsymbol{\theta})P(Y_{m+1} \geq y^*_{m+1}, \ldots, Y_n \geq y^*_n|\boldsymbol{\theta})$ rather than simply $P(y_1, \ldots, y_n|\boldsymbol{\theta})$ as before

Censored model for Leprosy in Olinda 1991-1995

- Using this 60% cut-off as a working assumption (could obviously experiment with alternatives) we then have a need for a model that can incorporate censoring and this provides an example of how relatively straightforward it is to handle censored values in the Bayesian framework more generally
- The basic model for the leprosy counts remains the same, but now data is: $(\boldsymbol{y}, \boldsymbol{y}^*) = (y_1, \ldots, y_m, y^*_{m+1}, \ldots, y^*_n)$, where y^*_i refer to the censored data values in the areas where the deprivation indicator exceeds 60% (for convenience we reorder the data by deprivation score).
- So now likelihood is: $P(y_1, \ldots, y_m|\boldsymbol{\theta})P(Y_{m+1} \geq y^*_{m+1}, \ldots, Y_n \geq y^*_n|\boldsymbol{\theta})$ rather than simply $P(y_1, \ldots, y_n|\boldsymbol{\theta})$ as before
- MCMC then provides posterior: $P(\boldsymbol{\theta}, y_{m+1}, \ldots, y_n|\boldsymbol{y}, \boldsymbol{y}^*)$ i.e. the joint distribution of the parameter set in the model $\boldsymbol{\theta}$ together with estimates for the $n - m$ censored values given the m exactly observed data values $\boldsymbol{y}$ and the $n - m$ censoring points $\boldsymbol{y}^*$

The relevant **WinBUGS** model for the censored case is:

```
for (i in 1 : N) {
    y[i] ~ dpois(mu[i])I(cens[i],)
    phi[1] ~ dnorm(0.0, tau.phi)
    log(mu[i]) <- log(e[i]) + alpha + beta * x[i] + phi[i] + nu[i]
    rho[i] <- exp(alpha+beta*x[i]+phi[i]+nu[i])
}
etc ... as before for other distributions
```

The relevant **WinBUGS** model for the censored case is:

for (i in 1 : N) {

y[i] $\sim$ dpois(mu[i])**I(cens[i],)**

phi[1] $\sim$ dnorm(0.0, tau.phi)

log(mu[i]) $\leftarrow$ log(e[i]) + alpha + beta * x[i] + phi[i] + nu[i]

rho[i] $\leftarrow$ exp(alpha+beta*x[i]+phi[i]+nu[i])

}

etc ... as before for other distributions

y_i now contains missing values for censored observations (i.e. where $x_i \geq 0.6$) whereas 'cens[i]' is set to zero for real observations and to the counts observed for the censored observations.

The relevant **WinBUGS** model for the censored case is:

for (i in 1 : N) {

y[i] $\sim$ dpois(mu[i])**I(cens[i],)**

phi[1] $\sim$ dnorm(0.0, tau.phi)

log(mu[i]) $\leftarrow$ log(e[i]) + alpha + beta * x[i] + phi[i] + nu[i]

rho[i] $\leftarrow$ exp(alpha+beta*x[i]+phi[i]+nu[i])

}

etc ... as before for other distributions

y_i now contains missing values for censored observations (i.e. where $x_i \geq 0.6$) whereas 'cens[i]' is set to zero for real observations and to the counts observed for the censored observations.

Initial values are as before and in addition censored values of y_i are initialised to the observed counts at the censored observations (or just above)

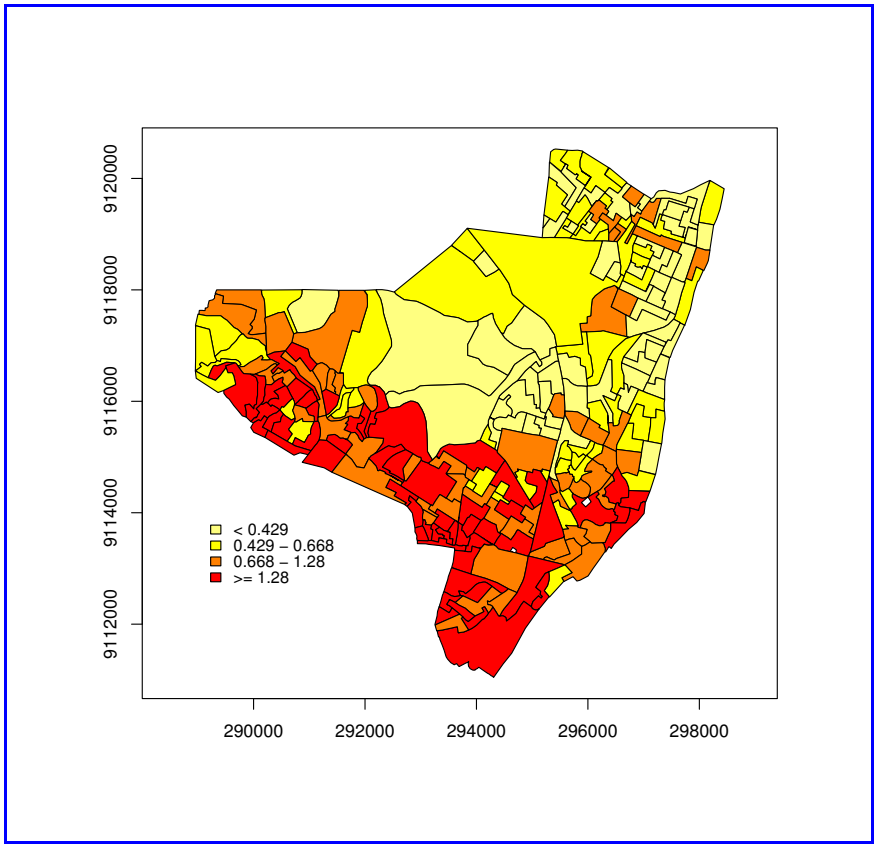
☞

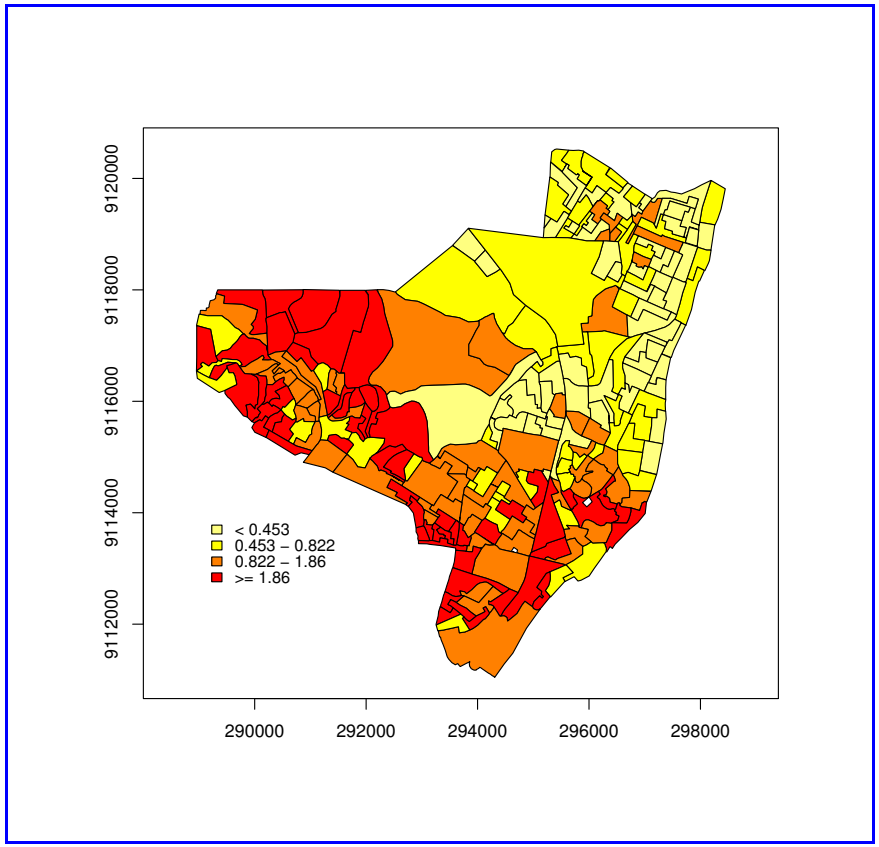
Results for standard model versus censoring for leprosy in Olinda

Model	$\hat{\alpha}$		$\hat{\beta}$	
	mean	90% cred int	mean	90% cred int
1991-1995 std.	-0.5	(-0.6, -0.2)	0.4	(-0.1, 1.2)
1991-1995 cens.	-0.9	(-1.2, -0.6)	1.9	(1.1, 2.7)

Model	$\hat{\sigma}_\phi$		$\hat{\sigma}_\nu$	
	mean	sd	mean	sd
1991-1995 std.	0.4	0.1	1.0	0.2
1991-1995 cens.	0.5	0.1	0.9	0.2

Modelled leprosy relative risks standard (left) and censored (right)





Leprosy surveillance in Olinda 1991-1995

- Treatment of the suspected under-detections via censoring would appear to have been relatively successful in producing more realistic estimates of true cases in the poorer areas. The estimated total of 1991-1995 cases is now 1590, as opposed to 1135 observed and predicted from non-censored model— more similar to the 1,766 cases actually detected in 1996-2000.

Leprosy surveillance in Olinda 1991-1995

- Treatment of the suspected under-detections via censoring would appear to have been relatively successful in producing more realistic estimates of true cases in the poorer areas. The estimated total of 1991-1995 cases is now 1590, as opposed to 1135 observed and predicted from non-censored model— more similar to the 1,766 cases actually detected in 1996-2000.
- Model can be used to estimate number of under-detections in each area. If such under-detection estimates had been available in 1995 then improved surveillance could have been targeted in areas where particularly bad under-detection had occurred with knowledge of the suspected numbers of missed detections in those areas

Leprosy surveillance in Olinda 1991-1995

- Treatment of the suspected under-detections via censoring would appear to have been relatively successful in producing more realistic estimates of true cases in the poorer areas. The estimated total of 1991-1995 cases is now 1590, as opposed to 1135 observed and predicted from non-censored model— more similar to the 1,766 cases actually detected in 1996-2000.
- Model can be used to estimate number of under-detections in each area. If such under-detection estimates had been available in 1995 then improved surveillance could have been targeted in areas where particularly bad under-detection had occurred with knowledge of the suspected numbers of missed detections in those areas
- Example illustrates how the statistical modelling of disease rates can directly lead to the identification of valuable public health responsive action. Application discussed concerns leprosy control, but the methods may equally well be applied in surveillance of other diseases where under-reporting of cases is a potential problem.

Handling missing data values

- Also worth noting at this point that missing data values (as opposed to censored values) are also very simply handled in the Bayesian framework

Handling missing data values

- Also worth noting at this point that missing data values (as opposed to censored values) are also very simply handled in the Bayesian framework
- The data vector is then $(\boldsymbol{y}, \boldsymbol{y}^{(*)}) = (y_1, \ldots, y_m, y_{m+1}^{(*)}, \ldots, y_n^{(*)})$, where the y_i refer to actual data values and the $y_i^{(*)}$ refer to missing data values (for convenience we assume the data are ordered accordingly)

Handling missing data values

- Also worth noting at this point that missing data values (as opposed to censored values) are also very simply handled in the Bayesian framework
- The data vector is then $(\boldsymbol{y}, \boldsymbol{y}^{(*)}) = (y_1, \ldots, y_m, y_{m+1}^{(*)}, \ldots, y_n^{(*)})$, where the y_i refer to actual data values and the $y_i^{(*)}$ refer to missing data values (for convenience we assume the data are ordered accordingly)
- The model (i.e. likelihood, priors, hyperpriors) remains the same but now MCMC provides samples from $p(\boldsymbol{\theta}, \boldsymbol{y}^{(*)}|\boldsymbol{y})$ – the joint posterior distribution of the set of real parameters in the model $\boldsymbol{\theta}$ together with the $n - m$ missing values $\boldsymbol{y}^{(*)}$, given the m actual observed data values $\boldsymbol{y}$

Handling missing data values

- Also worth noting at this point that missing data values (as opposed to censored values) are also very simply handled in the Bayesian framework
- The data vector is then $(\boldsymbol{y}, \boldsymbol{y}^{(*)}) = (y_1, \ldots, y_m, y_{m+1}^{(*)}, \ldots, y_n^{(*)})$, where the y_i refer to actual data values and the $y_i^{(*)}$ refer to missing data values (for convenience we assume the data are ordered accordingly)
- The model (i.e. likelihood, priors, hyperpriors) remains the same but now MCMC provides samples from $p(\boldsymbol{\theta}, \boldsymbol{y}^{(*)}|\boldsymbol{y})$ – the joint posterior distribution of the set of real parameters in the model $\boldsymbol{\theta}$ together with the $n - m$ missing values $\boldsymbol{y}^{(*)}$, given the m actual observed data values $\boldsymbol{y}$
- Point estimates, standard errors etc. for any particular missing values are then obtained from the marginal posterior distribution for this quantity, in exactly the same way as they would be for any other parameter of the model

- A further example of using an ecological model is provided by returning to the larynx cancer data and recalling that we have a three level indicator for the prevalence of smoking in each of these districts (1='low', 2='moderate', 3= 'high'). We now incorporate this categorical factor into the earlier spatially structured Poisson-log normal model.

- A further example of using an ecological model is provided by returning to the larynx cancer data and recalling that we have a three level indicator for the prevalence of smoking in each of these districts (1='low', 2='moderate', 3= 'high'). We now incorporate this categorical factor into the earlier spatially structured Poisson-log normal model.
- The resulting **WinBUGS** model can also be extended to predict the excess number of cases associated with smoking in any particular area and the probability that reducing smoking levels to 1 in that area will lead to reduction of more than 15 cases. This requires the use of the idea of a Bayesian **predictive distribution**.

- Suppose that the original data consists of observations $\boldsymbol{y}$ associated with p covariates $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ in a Bayesian model that involves a set of parameters $\boldsymbol{\theta}$
- Further suppose that we wish to predict the response y^* at a new set of covariate values $(x_1^*, \ldots, x_p^*)$. Then the relevant **predictive distribution** is defined as:

$$P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{y}, \boldsymbol{X}] = \int_{\boldsymbol{\theta}} P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{\theta}] P[\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}] \, d\boldsymbol{\theta}$$

- Suppose that the original data consists of observations $\boldsymbol{y}$ associated with p covariates $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ in a Bayesian model that involves a set of parameters $\boldsymbol{\theta}$
- Further suppose that we wish to predict the response y^* at a new set of covariate values $(x_1^*, \ldots, x_p^*)$. Then the relevant **predictive distribution** is defined as:

$$P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{y}, \boldsymbol{X}] = \int_{\boldsymbol{\theta}} P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{\theta}] P[\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}] \, d\boldsymbol{\theta}$$

- i.e. the predictive distribution averages over the uncertainty in the parameter values as reflected by the posterior distribution

- Suppose that the original data consists of observations $\boldsymbol{y}$ associated with p covariates $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ in a Bayesian model that involves a set of parameters $\boldsymbol{\theta}$
- Further suppose that we wish to predict the response y^* at a new set of covariate values $(x_1^*, \ldots, x_p^*)$. Then the relevant **predictive distribution** is defined as:

$$P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{y}, \boldsymbol{X}] = \int_{\boldsymbol{\theta}} P[y^*|(x_1^*, \ldots, x_p^*), \boldsymbol{\theta}] P[\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}] \, d\boldsymbol{\theta}$$

- i.e. the predictive distribution averages over the uncertainty in the parameter values as reflected by the posterior distribution
- In fact we have already used this idea in predicting the values of censored in the Olinda example and in our discussion of handling missing data values

The relevant **WinBUGS** model (focussing on area 53 for predictive purposes) is:

```
for (i in 1 : N) {
y[i]  ~ dpois(mu[i])                        # Poisson likelihood for observed counts
log(mu[i]) <- log(e[i])+alpha+beta[smoke[i]]+phi[i]+nu[i] # model for Poisson mean
   phi[i] ~ dnorm(0, tau.phi)               # normal prior for spatially unstructured effects
   rho[i] <- exp(alpha+beta[smoke[i]]+phi[i]+nu[i])  # RRs compared to reference rate
   rholocaladj[i] <- exp(phi[i]+nu[i])      # RRs compared to overall risk in study area
                                            # after adjusting for smoking
}
nu[1:N] ~ car.normal(adj[],weights[],num[],tau.nu) # CAR prior for spatially structured effects
alpha  ~ dflat()              # locally uniform prior for mean log relative risk
beta[1] <- 0                  # set level 1 of smoking to be the reference category
beta[2] ~ dnorm(0, 0.0001)    # diffuse normal prior for beta[2]
beta[3] ~ dnorm(0, 0.0001)    # diffuse normal prior for beta[3]
tau.phi  ~ dgamma(0.5, 0.0005)    # diffuse gamma hyperprior for tau.phi
tau.nu  ~ dgamma(0.5, 0.0005)     # diffuse gamma hyperprior for tau.nu
sigma.phi <- sqrt(1/tau.phi)      # st dev of prior for spatially unstructured effects
sigma.nu <- sqrt(1/tau.nu)        # st dev of prior for spatially structured effects
mu.pred53 <- exp(e[53]+alpha+beta[1]+phi[53]+nu[53])  # predict mean in 53 with smoking level 1
y.pred53 ~ dpois(mu.pred53)   # predict individual value in 53 with smoking level 1
y.diff53 <- y[53] - y.pred53  # predict reduction in cases in 53 if no smoking
P.diff53 <- step(y.diff53-15) # predict probability reduction > 15 cases
```

Further topics in ecological studies

- It is often acknowledged that case-control studies are the ‘gold standard’ in studying the relationship between disease and risk factors. But at the same time it is admitted that these usually require the collection of new data, they are expensive and time-consuming and there are problems of selection and other biases

Further topics in ecological studies

- It is often acknowledged that case-control studies are the ‘gold standard’ in studying the relationship between disease and risk factors. But at the same time it is admitted that these usually require the collection of new data, they are expensive and time-consuming and there are problems of selection and other biases
- So aggregate level (‘ecological’) studies with suitable models should not be dismissed:

Further topics in ecological studies

- It is often acknowledged that case-control studies are the ‘gold standard’ in studying the relationship between disease and risk factors. But at the same time it is admitted that these usually require the collection of new data, they are expensive and time-consuming and there are problems of selection and other biases
- So aggregate level (‘ecological’) studies with suitable models should not be dismissed:
  - data involved are cheap and widely available
  - range of exposure to risk factors in populations concerned is potentially larger than in studies on individuals
  - exposure measurement errors are typically dampened by averaging over areas

But one should always appreciate the potential problems and biases associated with aggregate level studies:

But one should always appreciate the potential problems and biases associated with aggregate level studies:

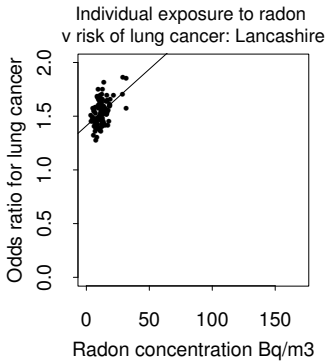
- Problems of **spatial scale**—typically the health, exposure and population data are obtained from different sources and this can lead to problems of imprecise geographical matching and data aggregation. The choice of aggregation unit needs to trade off between data precision, the ability to detect localised patterns of risk and the scale over which an environmental risk factor may be expected to operate.

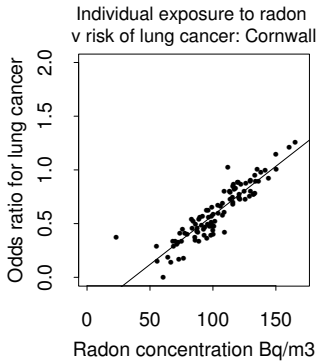
But one should always appreciate the potential problems and biases associated with aggregate level studies:

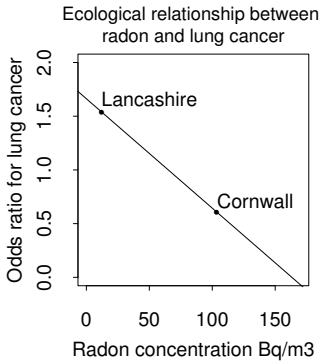
- Problems of **spatial scale**—typically the health, exposure and population data are obtained from different sources and this can lead to problems of imprecise geographical matching and data aggregation. The choice of aggregation unit needs to trade off between data precision, the ability to detect localised patterns of risk and the scale over which an environmental risk factor may be expected to operate.

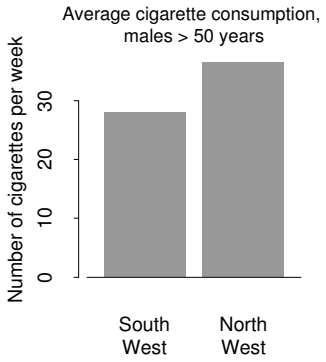
- Problems of **confounding**—an omitted variable which is related to both the disease and to some of the included risk factors. E.g. area-level socio-deprivation is strongly correlated with many diseases, but it also coincides with such things as industrial sites, busy roads and smoking.

Hypothetical result of not accounting for regional smoking differences in studying relationship of lung cancer to indoor radon exposure at an aggregate level









- Problems of **specification bias**—the difference between individual and group level relationships between disease incidence and risk caused by non-constant exposure to risk within the group

- Problems of **specification bias**—the difference between individual and group level relationships between disease incidence and risk caused by non-constant exposure to risk within the group
  - For example, suppose that we are considering a single risk factor and that at the *individual level* the relative risk of contracting a disease given a level of exposure x is $\exp(\alpha + \beta x)$ (i.e. a log-linear relationship as in the ecological models we have considered)

- Problems of **specification bias**—the difference between individual and group level relationships between disease incidence and risk caused by non-constant exposure to risk within the group
  - For example, suppose that we are considering a single risk factor and that at the *individual level* the relative risk of contracting a disease given a level of exposure x is $\exp(\alpha + \beta x)$ (i.e. a log-linear relationship as in the ecological models we have considered)
  - Then the relationship between group relative risk and mean exposure (μ_x) at an *area-level* will *not* be $\exp(\alpha + \beta \mu_x)$ unless the exposure of all individuals in the area is the same (i.e. all have exposure μ_x)

- Problems of **specification bias**—the difference between individual and group level relationships between disease incidence and risk caused by non-constant exposure to risk within the group
  - For example, suppose that we are considering a single risk factor and that at the *individual level* the relative risk of contracting a disease given a level of exposure x is $\exp(\alpha + \beta x)$ (i.e. a log-linear relationship as in the ecological models we have considered)
  - Then the relationship between group relative risk and mean exposure (μ_x) at an *area-level* will *not* be $\exp(\alpha + \beta \mu_x)$ unless the exposure of all individuals in the area is the same (i.e. all have exposure μ_x)
  - Instead this relationship will be a weighted average of the function $\exp(\alpha + \beta x)$ over values of x with the weights reflecting the probabilities of individuals within the region receiving exposure levels x

Specification bias in ecological studies

- A simple case is when the within area probability distribution of individual levels of exposure is $N(\mu_x, \sigma_x^2)$
- Then it may be shown that the area-level relationship is actually $\exp(\alpha + \beta\mu_x + \beta^2 \frac{\sigma_x^2}{2})$.

Specification bias in ecological studies

- A simple case is when the within area probability distribution of individual levels of exposure is $N(\mu_x, \sigma_x^2)$
- Then it may be shown that the area-level relationship is actually $\exp(\alpha + \beta\mu_x + \beta^2 \frac{\sigma_x^2}{2})$.
- The key general point is that to adjust for specification bias, we need information on the within-area distribution of exposure — say from a small random sample of individuals within each area.

Specification bias in ecological studies

- A simple case is when the within area probability distribution of individual levels of exposure is $N(\mu_x, \sigma_x^2)$
- Then it may be shown that the area-level relationship is actually $\exp(\alpha + \beta\mu_x + \beta^2 \frac{\sigma_x^2}{2})$.
- The key general point is that to adjust for specification bias, we need information on the within-area distribution of exposure — say from a small random sample of individuals within each area.
- For two or more exposures we would need information on the *joint* exposure distribution within areas.

E.g. For a single covariate and given a sub-sample of the exposures of M individuals in each of the N areas a relevant **WinBUGS** model might be something like:

```
for (i in 1 : N) {
  y[i] ~ dpois(mu[i])   # observed counts
  log(mu[i])<-log(e[i])+alpha+beta*mu.x[i]+pow(beta,2)*sigmasq.x[i]/2   # mean model
  for (j in 1 : M) {
      x[i,j] ~ dnorm(mu.x[i],tau.x[i])   # exposure sub-sample
  }
  mu.x[i] ~ dnorm(0, 1.0E-6)   # mean area-level exposure
  tau.x[i] ~ dgamma(.01,.01)   # precision area-level exposure
  sigmasq.x[i] <- 1/tau.x[i]   # area-level exposure variance
}
alpha ~ dnorm(0, 1.0E-6)   # prior for alpha
beta ~ dnorm(0, 1.0E-6)   # prior for beta
```

where, for simplicity of presentation we have ignored the random effect terms that would usually be additionally included

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.
- The latter involves several possible sources of error including:
  - equating environmental (external) exposure with biological (internal dose)

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.
- The latter involves several possible sources of error including:
  - equating environmental (external) exposure with biological (internal dose)
  - equating current exposure with past exposure

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.
- The latter involves several possible sources of error including:
  - equating environmental (external) exposure with biological (internal dose)
  - equating current exposure with past exposure
  - equating modelled estimates with true exposure

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.
- The latter involves several possible sources of error including:
  - equating environmental (external) exposure with biological (internal dose)
  - equating current exposure with past exposure
  - equating modelled estimates with true exposure
  - equating average exposure for an area with individual exposure

Other issues in ecological studies

- Finally, we should comment on problems of **data errors**—these could arise in recording of health events, in demographic variables or from measurement errors related to exposure.
- The latter involves several possible sources of error including:
  - equating environmental (external) exposure with biological (internal dose)
  - equating current exposure with past exposure
  - equating modelled estimates with true exposure
  - equating average exposure for an area with individual exposure
- *Some* of these exposure measurement problems may be addressed by various forms of **errors-in-variables** modelling.

Errors-in-variables modelling

- One of the simplest forms of errors-in-variables modelling concerns **classical measurement error** where observed values of exposure are assumed to vary about the true measurement

Errors-in-variables modelling

- One of the simplest forms of errors-in-variables modelling concerns **classical measurement error** where observed values of exposure are assumed to vary about the true measurement
- If present and not allowed for then such measurement error can result in **attenuation** effects when estimating model parameters

Errors-in-variables modelling

- One of the simplest forms of errors-in-variables modelling concerns **classical measurement error** where observed values of exposure are assumed to vary about the true measurement
- If present and not allowed for then such measurement error can result in **attenuation** effects when estimating model parameters
- Commonly such attenuation leads to covariate coefficient estimates being biased (usually towards the null) and sampling error in the response being overestimated.

➢ For **continuous exposures** classical measurement error is often described by the **reliability coefficient**:

$$\rho = \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{err}}$$

where σ^2_{true} is the variance of the true exposure and σ^2_{err} reflects the variance of measurement errors.

- For **continuous exposures** classical measurement error is often described by the **reliability coefficient**:

$$\rho = \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{err}}$$

where σ^2_{true} is the variance of the true exposure and σ^2_{err} reflects the variance of measurement errors.

- The average size of errors for **categorical exposures** can be described by a matrix of misclassification probabilities p_{jk}, where p_{jk} is the conditional probability that a subject is classified as level k given that they are truly exposed to level j

- For **continuous exposures** classical measurement error is often described by the **reliability coefficient**:

$$\rho = \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{err}}$$

where σ^2_{true} is the variance of the true exposure and σ^2_{err} reflects the variance of measurement errors.

- The average size of errors for **categorical exposures** can be described by a matrix of misclassification probabilities p_{jk}, where p_{jk} is the conditional probability that a subject is classified as level k given that they are truly exposed to level j

- Given information on these quantities the ecological models that we have described can be adjusted to allow for measurement errors in the explanatory variables

Adjusting Larynx Cancer risk for air pollution

- In a previous model we adjusted the risk of larynx cancer according to a three level smoking factor. We now include as an additional covariate a measure of air pollution—the annual mean levels of particulates in each area estimated from a dispersion model based on traffic flow

Adjusting Larynx Cancer risk for air pollution

- In a previous model we adjusted the risk of larynx cancer according to a three level smoking factor. We now include as an additional covariate a measure of air pollution—the annual mean levels of particulates in each area estimated from a dispersion model based on traffic flow
- From previous validation studies the reliability coefficient for these air pollution estimates is:

$$\rho = \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{err}} = 0.71$$

Adjusting Larynx Cancer risk for air pollution

- In a previous model we adjusted the risk of larynx cancer according to a three level smoking factor. We now include as an additional covariate a measure of air pollution—the annual mean levels of particulates in each area estimated from a dispersion model based on traffic flow
- From previous validation studies the reliability coefficient for these air pollution estimates is:

$$\rho = \frac{\sigma^2_{true}}{\sigma^2_{true} + \sigma^2_{err}} = 0.71$$

- We consider a **WinBUGS** model that includes the air pollution covariate and at the same time allows for errors in observed values of this covariate in accord with the above reliability coefficient

Adjusting larynx cancer risk for air pollution & measurement error

The relevant **WinBUGS** model is:

```
for (i in 1 : N) {
 y[i]  ~ dpois(mu[i])                    # Poisson likelihood for observed counts
 log(mu[i]) <- log(e[i])+alpha+beta1[smoke[i]]+beta2*truepoll[i]+phi[i]+nu[i] #model for mean
 phi[i] ~ dnorm(0, tau.phi)              # prior for unstructured random effects
 truepoll[i] ~ dnorm(mu.true,tau.true) # distribution of true exposure
 poll[i] ~ dnorm(truepoll[i],tau.err)   # distribution of measurement error
 rholocaladj[i] <- exp(phi[i]+nu[i])    # R risks compared to overall risk in study area after
}                                        # adjusting for smoking and air pollution
nu[1:N] ~ car.normal(adj[],weights[],num[],tau.nu) # CAR prior for structured random effects
alpha  ~ dflat()                 # uniform prior for alpha
beta1[1] <- 0                    # set beta1[1] as the reference smoking level
beta1[2] ~ dnorm(0, 0.0001)      # diffuse normal prior for beta1[2]
beta1[3] ~ dnorm(0, 0.0001)      # diffuse normal prior for beta1[3]
beta2    ~ dnorm(0, 0.0001)      # diffuse normal prior for beta2
tau.phi  ~ dgamma(0.5, 0.0005)   # hyperprior for tau.phi
tau.nu   ~ dgamma(0.5, 0.0005)   # hyperprior for tau.phi
sigma.phi <- sqrt(1/tau.phi)     # st dev of unstructured rand effects
sigma.nu <- sqrt(1/tau.nu)       # st dev of structured rand effects
mu.true  ~ dnorm(0, .00001)      # diffuse normal hyperprior for mu.true
tau.true  ~ dgamma(0.5, 0.0005) # diffuse gamma hyperprior for tau.true
sigmasq.true <- 1/tau.true       # variance of true measurements
rho <- 0.71                      # reliability coefficient
sigmasq.err <-  sigmasq.true*(1-rho)/rho # variance of measurement error
tau.err <- 1/sigmasq.err         # precision of measurement error
```

Spatio-Temporal Models

- So far we have only considered models involving spatial outcomes, but obviously the evolution over time of such outcomes may also be of considerable interest
- There exist a broad class of models that may be used in modelling (and perhaps forecasting) spatio-temporal disease incidence by area

- So far we have only considered models involving spatial outcomes, but obviously the evolution over time of such outcomes may also be of considerable interest
- There exist a broad class of models that may be used in modelling (and perhaps forecasting) spatio-temporal disease incidence by area
- We focus here only on illustrating the **potential** for spatio-temporal modelling of small area disease rates, restricting our discussion to fairly simple extensions to the Bayesian ecological models that we have used in the purely spatial context

- So far we have only considered models involving spatial outcomes, but obviously the evolution over time of such outcomes may also be of considerable interest
- There exist a broad class of models that may be used in modelling (and perhaps forecasting) spatio-temporal disease incidence by area
- We focus here only on illustrating the **potential** for spatio-temporal modelling of small area disease rates, restricting our discussion to fairly simple extensions to the Bayesian ecological models that we have used in the purely spatial context
- In particular we do not explore in any detail the various alternative formulations of space-time interaction in such models — this is a substantive topic and we can only touch upon the issues here (for more details see Knorr-Held and Besag, 1998)

- The simplest temporal extension of the purely spatial Bayesian disease mapping model discussed earlier is to include a temporally unstructured time effect into the model.

➢ The simplest temporal extension of the purely spatial Bayesian disease mapping model discussed earlier is to include a temporally unstructured time effect into the model. Given disease counts y_{it} (in areas i and time periods t) and corresponding expected numbers of cases e_{it} (derived from suitable reference rates) the model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_{it}\rho_{it}) \\
\log \mu_{it} &= \log e_{it} + \log \rho_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \quad \text{(as a baseline to avoid identifiability problems)} \\
\delta_t &\sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

➢ The simplest temporal extension of the purely spatial Bayesian disease mapping model discussed earlier is to include a temporally unstructured time effect into the model. Given disease counts y_{it} (in areas i and time periods t) and corresponding expected numbers of cases e_{it} (derived from suitable reference rates) the model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_{it}\rho_{it}) \\
\log \mu_{it} &= \log e_{it} + \log \rho_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \quad \text{(as a baseline to avoid identifiability problems)} \\
\delta_t &\sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

so relative risks are $\rho_{it} = \exp(\alpha + \phi_i + \nu_i + \delta_t)$ with $\rho_{i1} = \exp(\alpha + \phi_i + \nu_i)$.

- The simplest temporal extension of the purely spatial Bayesian disease mapping model discussed earlier is to include a temporally unstructured time effect into the model. Given disease counts y_{it} (in areas i and time periods t) and corresponding expected numbers of cases e_{it} (derived from suitable reference rates) the model is:

$$\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_{it}\rho_{it}) \\
\log \mu_{it} &= \log e_{it} + \log \rho_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \quad \text{(as a baseline to avoid identifiability problems)} \\
\delta_t &\sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T
\end{aligned}$$

so relative risks are $\rho_{it} = \exp(\alpha + \phi_i + \nu_i + \delta_t)$ with $\rho_{i1} = \exp(\alpha + \phi_i + \nu_i)$.

- To complete the specification diffuse gamma hyperpriors are assumed for precisions corresponding to all hyperparameters i.e. for $\tau_\phi = 1/\sigma_\phi^2$, $\tau_\nu = 1/\sigma_\nu^2$ and $\tau_\delta = 1/\sigma_\delta^2$

- The previous model imposes no structure on the temporal effects and it may be that *temporally persistent* differences in the outcome are important i.e. the time effects should be temporally structured (smoothed)

- The previous model imposes no structure on the temporal effects and it may be that *temporally persistent* differences in the outcome are important i.e. the time effects should be temporally structured (smoothed)
- This may be expressed by introducing a temporally auto-correlated effect so that:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t$$

with for example $\omega_t \sim \text{Normal}(\omega_{t-1}, \sigma^2_\omega) \qquad t = 2, \ldots, T$ and $\omega_1 \sim \text{Normal}(0, \sigma^2_{\omega_1})$.
All other priors are as before

- The previous model imposes no structure on the temporal effects and it may be that *temporally persistent* differences in the outcome are important i.e. the time effects should be temporally structured (smoothed)
- This may be expressed by introducing a temporally auto-correlated effect so that:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t$$

with for example $\omega_t \sim \text{Normal}(\omega_{t-1}, \sigma^2_\omega) \qquad t = 2, \ldots, T$ and $\omega_1 \sim \text{Normal}(0, \sigma^2_{\omega_1})$. All other priors are as before

- Various alternative specifications to the above simple **random walk** for the temporally auto-correlated component of this model are possible. For example a second order auto-regression may be preferred if one is interested in *predicting* future disease rates.

- The previous model imposes no structure on the temporal effects and it may be that *temporally persistent* differences in the outcome are important i.e. the time effects should be temporally structured (smoothed)
- This may be expressed by introducing a temporally auto-correlated effect so that:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t$$

with for example $\omega_t \sim \text{Normal}(\omega_{t-1}, \sigma^2_\omega) \qquad t = 2, \ldots, T$ and $\omega_1 \sim \text{Normal}(0, \sigma^2_{\omega_1})$. All other priors are as before

- Various alternative specifications to the above simple **random walk** for the temporally auto-correlated component of this model are possible. For example a second order auto-regression may be preferred if one is interested in *predicting* future disease rates.
- Note that identifiability problems arise with these kinds of formulations and will need to be addressed by imposing constraints on some parameters.

- Where interest focusses on modelling trends in the relative risk relative to the reference levels, then one might impose a stronger parametric structure on the temporal effects

- Where interest focusses on modelling trends in the relative risk relative to the reference levels, then one might impose a stronger parametric structure on the temporal effects
- E.g. a linear trend (identical across all areas) would correspond to a model:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma t$$

with $\gamma \sim$ Normal$(0, \sigma_\gamma^2)$ and all other priors as before.

- Where interest focusses on modelling trends in the relative risk relative to the reference levels, then one might impose a stronger parametric structure on the temporal effects
- E.g. a linear trend (identical across all areas) would correspond to a model:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma t$$

with $\gamma \sim \text{Normal}(0, \sigma_\gamma^2)$ and all other priors as before.

- To allow for differentiated trends between areas, e.g. with some falling more or some less than the global trend one could specify an area specific growth rate via:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma_i t$$

with exchangeable priors $\gamma_i \sim \text{Normal}(\mu_\gamma, \sigma_\gamma^2)$, where μ_γ is the overall average growth rate.

- Where interest focusses on modelling trends in the relative risk relative to the reference levels, then one might impose a stronger parametric structure on the temporal effects
- E.g. a linear trend (identical across all areas) would correspond to a model:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma t$$

with $\gamma \sim \text{Normal}(0, \sigma_\gamma^2)$ and all other priors as before.

- To allow for differentiated trends between areas, e.g. with some falling more or some less than the global trend one could specify an area specific growth rate via:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma_i t$$

with exchangeable priors $\gamma_i \sim \text{Normal}(\mu_\gamma, \sigma_\gamma^2)$, where μ_γ is the overall average growth rate.

- identifiability remains an issue and some parameter constraints may need to be imposed

- If such trends are expected to be differentiated in a spatially distinct pattern (i.e. similar falls or rises are spatially clustered) then the γ_i might be assumed to be spatially dependent

- If such trends are expected to be differentiated in a spatially distinct pattern (i.e. similar falls or rises are spatially clustered) then the γ_i might be assumed to be spatially dependent
- For example we could take $\gamma_i \sim \text{CAR}(\sigma_\gamma^2)$ in which case we would also need a γt term in the model to represent the global trend with $\gamma \sim \text{U}(-\infty, +\infty)$ since the CAR is improper and a sum to zero constraint on γ_i will need to be imposed.

- If such trends are expected to be differentiated in a spatially distinct pattern (i.e. similar falls or rises are spatially clustered) then the γ_i might be assumed to be spatially dependent
- For example we could take $\gamma_i \sim \text{CAR}(\sigma^2_\gamma)$ in which case we would also need a γt term in the model to represent the global trend with $\gamma \sim \text{U}(-\infty, +\infty)$ since the CAR is improper and a sum to zero constraint on γ_i will need to be imposed.
- so overall the model that results is:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma t + \gamma_i t$$

with γ_i now representing deviations from the overall γ.

- If such trends are expected to be differentiated in a spatially distinct pattern (i.e. similar falls or rises are spatially clustered) then the γ_i might be assumed to be spatially dependent
- For example we could take $\gamma_i \sim \text{CAR}(\sigma_\gamma^2)$ in which case we would also need a γt term in the model to represent the global trend with $\gamma \sim \text{U}(-\infty, +\infty)$ since the CAR is improper and a sum to zero constraint on γ_i will need to be imposed.
- so overall the model that results is:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \gamma t + \gamma_i t$$

with γ_i now representing deviations from the overall γ.

- Again identifiability is an issue and some parameter constraints may need to be imposed

- Models with area specific growth rates are not *separable* in space and time — they allow for spatio-temporal interactions i.e. there can be some shuffling of spatial relativities in the relative risks over time

- Models with area specific growth rates are not *separable* in space and time — they allow for spatio-temporal interactions i.e. there can be some shuffling of spatial relativities in the relative risks over time
- However, the form of those models imposes a restricted structure on this space-time interaction and a more flexible class of models is obtained by adding an interaction term to the model discussed earlier which involved a temporally auto-correlated main effect so that this now becomes:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t + \psi_{it}$$

- Models with area specific growth rates are not *separable* in space and time — they allow for spatio-temporal interactions i.e. there can be some shuffling of spatial relativities in the relative risks over time
- However, the form of those models imposes a restricted structure on this space-time interaction and a more flexible class of models is obtained by adding an interaction term to the model discussed earlier which involved a temporally auto-correlated main effect so that this now becomes:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t + \psi_{it}$$

- The modelling options for spatial and temporal structure in the area-time interactions ψ_{it} are very wide, since autocorrelations over areas may be combined in various ways with those over time.

- Models with area specific growth rates are not *separable* in space and time — they allow for spatio-temporal interactions i.e. there can be some shuffling of spatial relativities in the relative risks over time
- However, the form of those models imposes a restricted structure on this space-time interaction and a more flexible class of models is obtained by adding an interaction term to the model discussed earlier which involved a temporally auto-correlated main effect so that this now becomes:

$$\log \mu_{it} = \log e_{it} + \alpha + \phi_i + \nu_i + \delta_t + \omega_t + \psi_{it}$$

- The modelling options for spatial and temporal structure in the area-time interactions ψ_{it} are very wide, since autocorrelations over areas may be combined in various ways with those over time.
- Knorr-Held (2000) discusses four types of interaction schemes, ranging from independence of all interactions to complete space/time dependence in the interactions

- Finally note that variability in relative risks over both space and time may be caused by changing impacts of social and other risk variables.

- Finally note that variability in relative risks over both space and time may be caused by changing impacts of social and other risk variables.
- All the space time models discussed can be extended to include ecological covariates $(x_{it1} \ldots, x_{itp})$ relating to areas, to time periods or to both

- Finally note that variability in relative risks over both space and time may be caused by changing impacts of social and other risk variables.
- All the space time models discussed can be extended to include ecological covariates $(x_{it1} \ldots, x_{itp})$ relating to areas, to time periods or to both
- A very wide range of formulations is possible, depending upon whether covariate measures are available only at each time point (spatially constant), or only for each area (constant in time), or for each space-time combination

- Finally note that variability in relative risks over both space and time may be caused by changing impacts of social and other risk variables.
- All the space time models discussed can be extended to include ecological covariates $(x_{it1} \ldots, x_{itp})$ relating to areas, to time periods or to both
- A very wide range of formulations is possible, depending upon whether covariate measures are available only at each time point (spatially constant), or only for each area (constant in time), or for each space-time combination
- Associated covariate model coefficients can likewise be modelled as globally constant, varying only over time, varying only over space or varying over both time and space.

- For example trends in the impact of a single time-specific predictor (x_{it} might be modelled via: $\log \mu_{it} = \log e_{it} + \alpha + \beta_t x_{it} + \phi_i + \nu_i$

  with β_t taken as either temporally unstructured or structured (e.g. by a random walk)

- For example trends in the impact of a single time-specific predictor (x_{it} might be modelled via: $\log \mu_{it} = \log e_{it} + \alpha + \beta_t x_{it} + \phi_i + \nu_i$

  with β_t taken as either temporally unstructured or structured (e.g. by a random walk)

- Whereas a model such as: $\log \mu_{it} = \log e_{it} + \alpha + \beta_i x_{it} + \phi_i + \nu_i + \delta_t$

  with spatially unstructured or structured β_i, would allow one to model differences in the importance of the explanatory variable between areas

- For example trends in the impact of a single time-specific predictor (x_{it} might be modelled via: $\log \mu_{it} = \log e_{it} + \alpha + \beta_t x_{it} + \phi_i + \nu_i$

  with β_t taken as either temporally unstructured or structured (e.g. by a random walk)

- Whereas a model such as: $\log \mu_{it} = \log e_{it} + \alpha + \beta_i x_{it} + \phi_i + \nu_i + \delta_t$

  with spatially unstructured or structured β_i, would allow one to model differences in the importance of the explanatory variable between areas

- Note that in general identifiability problems will need to be addressed in such models.

- For example trends in the impact of a single time-specific predictor (x_{it} might be modelled via: $\log \mu_{it} = \log e_{it} + \alpha + \beta_t x_{it} + \phi_i + \nu_i$

  with β_t taken as either temporally unstructured or structured (e.g. by a random walk)

- Whereas a model such as: $\log \mu_{it} = \log e_{it} + \alpha + \beta_i x_{it} + \phi_i + \nu_i + \delta_t$

  with spatially unstructured or structured β_i, would allow one to model differences in the importance of the explanatory variable between areas

- Note that in general identifiability problems will need to be addressed in such models.

- Also note that models with covariate coefficients with are both temporally and spatially varying may need to use the specialised methods referred to earlier in relation to varying covariate coefficients in purely spatial ecological models.

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an example of the use of spatio-temporal models we consider the data comprising diagnosed cases of Leptospirosis by year for the period 1997-2002 (total of 367 cases) in 157 districts of the city of Rio de Janeiro.

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an example of the use of spatio-temporal models we consider the data comprising diagnosed cases of Leptospirosis by year for the period 1997-2002 (total of 367 cases) in 157 districts of the city of Rio de Janeiro.
- Annual area expected values e_i are taken as constant over time and based on the district populations in the 2001 census, using as a reference rate the overall disease incidence for the six years as a proportion of 6 $\times$ the total 2001 population in the study region

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an example of the use of spatio-temporal models we consider the data comprising diagnosed cases of Leptospirosis by year for the period 1997-2002 (total of 367 cases) in 157 districts of the city of Rio de Janeiro.
- Annual area expected values e_i are taken as constant over time and based on the district populations in the 2001 census, using as a reference rate the overall disease incidence for the six years as a proportion of 6 $\times$ the total 2001 population in the study region
- We also include two area specific deprivation covariates from the 2001 census—x_{i1} (proportion of families with income of less than one minimum wage) and x_{i2} (proportion of the population living in favelas)

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an example of the use of spatio-temporal models we consider the data comprising diagnosed cases of Leptospirosis by year for the period 1997-2002 (total of 367 cases) in 157 districts of the city of Rio de Janeiro.
- Annual area expected values e_i are taken as constant over time and based on the district populations in the 2001 census, using as a reference rate the overall disease incidence for the six years as a proportion of 6 $\times$ the total 2001 population in the study region
- We also include two area specific deprivation covariates from the 2001 census—x_{i1} (proportion of families with income of less than one minimum wage) and x_{i2} (proportion of the population living in favelas)
- A further area and time specific covariate x_{it3} is maximum annual rainfall in the years 1997-2002 interpolated to districts from observations recorded at 32 weather stations dotted across the city. This provides some indication of the risk of floods in each district in the year in question.

Overall the model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_i \rho_{it}) \\
\log \mu_{it} &= \log e_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{it3} + \phi_i + \nu_i + \delta_t + \omega_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\beta_1 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_2 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_3 &\sim \text{Normal}(0, 1.0E-5) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \text{ and } \delta_t \sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T \\
\omega_1 &\sim \text{Normal}(0, \sigma_{\omega_1}^2) \text{ and } \omega_t \sim \text{Normal}(\omega_{t-1}, \sigma_\omega^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

Overall the model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_i \rho_{it}) \\
\log \mu_{it} &= \log e_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{it3} + \phi_i + \nu_i + \delta_t + \omega_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\beta_1 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_2 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_3 &\sim \text{Normal}(0, 1.0E-5) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \text{ and } \delta_t \sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T \\
\omega_1 &\sim \text{Normal}(0, \sigma_{\omega_1}^2) \text{ and } \omega_t \sim \text{Normal}(\omega_{t-1}, \sigma_\omega^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

Diffuse Gamma hyperpriors are assumed for precisions relating to all hyperparameters.

The relevant **WinBUGS** model is:

```
for (i in 1 : regions) {
  for (t in 1 : time) {
   cases[i,t] ~ dpois(mu[i,t])
   log(mu[i,t]) <- log(e[i]) + alpha + beta1*x1[i]+beta2*x2[i]+beta3*x3[i,t]+phi[i]+nu[i]+delta[t]+omega[t]
   rho[i,t]<-exp(alpha + beta1*x1[i]+beta2*x2[i]+beta3*x3[i,t]+phi[i]+nu[i]+delta[t]+omega[t]) # RR
   rhoadj[i,t]<-exp(phi[i]+nu[i]+delta[t]+omega[t]) # RR adjusted for covariates
  }
  phi[i] ~ dnorm(0,tau.phi)
  rhoiadj[i]<-exp(phi[i]+nu[i]) # RR adjusted for covariates averaged over all years
}
nu[1:regions] ~ car.normal(adj[], weights[], num[], tau.nu)
delta[1]<-0
omega[1] ~ dnorm(0, tau.omega1)
rhotadj[1]<-exp(omega[t]) # RR adjusted for covariates in year 1 averaged over all districts
for (t in 2 :time) {
  delta[t] ~ dnorm(0,tau.delta)
  omega[t]~dnorm(omega[t-1],tau.omega)
  rhotadj[t]<-exp(delta[t]+omega[t]) # RR adjusted for covariates in years 2-6 averaged over all districts
}
alpha  ~ dflat()
beta1  ~ dnorm(0.0, 1.0E-5)
beta2  ~ dnorm(0.0, 1.0E-5)
beta3  ~ dnorm(0.0, 1.0E-5)
tau.phi  ~ dgamma(0.1,0.1)
tau.nu ~ dgamma(0.1,0.1)
tau.delta  ~ dgamma(0.1,0.1)
tau.omega1 ~ dgamma(0.1,0.1)
tau.omega ~ dgamma(0.1,0.1)
```

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an alternative we could attempt to model the monthly variation averaged over all years 1997-2002 (total of 367 cases) in the 157 districts.

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an alternative we could attempt to model the monthly variation averaged over all years 1997-2002 (total of 367 cases) in the 157 districts.
- The monthly area expected values e_i for the six years are taken as constant over all months and are based on the district populations in the 2001 census, using as a reference rate $\frac{1}{12}$ of the overall disease incidence for the six years as a proportion of the total 2001 population in the study region

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an alternative we could attempt to model the monthly variation averaged over all years 1997-2002 (total of 367 cases) in the 157 districts.
- The monthly area expected values e_i for the six years are taken as constant over all months and are based on the district populations in the 2001 census, using as a reference rate $\frac{1}{12}$ of the overall disease incidence for the six years as a proportion of the total 2001 population in the study region
- As before we include two area specific deprivation covariates from the 2001 census—x_{i1} (proportion of families with income of less than one minimum wage) and x_{i2} (proportion of the population living in favelas)

Leptospirosis incidence in Rio de Janeiro 1997-2002

- As an alternative we could attempt to model the monthly variation averaged over all years 1997-2002 (total of 367 cases) in the 157 districts.
- The monthly area expected values e_i for the six years are taken as constant over all months and are based on the district populations in the 2001 census, using as a reference rate $\frac{1}{12}$ of the overall disease incidence for the six years as a proportion of the total 2001 population in the study region
- As before we include two area specific deprivation covariates from the 2001 census—x_{i1} (proportion of families with income of less than one minimum wage) and x_{i2} (proportion of the population living in favelas)
- We then include an area specific covariate x_{i3} as the average maximum annual rainfall over the years 1997-2002 interpolated to districts from observations recorded at 32 weather stations dotted across the city. This provides some indication of the average risk of floods in each district over all months in question.

Overall the monthly model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_i \rho_{it}) \\
\log \mu_{it} &= \log e_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \phi_i + \nu_i + \delta_t + \omega_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\beta_1 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_2 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_3 &\sim \text{Normal}(0, 1.0E-5) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \text{ and } \delta_t \sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T \\
\omega_1 &\sim \text{Normal}(0, \sigma_{\omega_1}^2) \text{ and } \omega_t \sim \text{Normal}(\omega_{t-1}, \sigma_\omega^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

Overall the monthly model is:

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(\mu_{it}) = \text{Poisson}(e_i \rho_{it}) \\
\log \mu_{it} &= \log e_i + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \phi_i + \nu_i + \delta_t + \omega_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\beta_1 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_2 &\sim \text{Normal}(0, 1.0E-5) \\
\beta_3 &\sim \text{Normal}(0, 1.0E-5) \\
\phi_i &\sim \text{Normal}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\delta_1 &= 0 \text{ and } \delta_t \sim \text{Normal}(0, \sigma_\delta^2) \qquad t = 2, \ldots, T \\
\omega_1 &\sim \text{Normal}(0, \sigma_{\omega_1}^2) \text{ and } \omega_t \sim \text{Normal}(\omega_{t-1}, \sigma_\omega^2) \qquad t = 2, \ldots, T
\end{aligned}
$$

Diffuse Gamma hyperpriors are assumed for precisions relating to all hyperparameters.

The relevant **WinBUGS** model is:

```
for (i in 1 : regions) {
  for (t in 1 : time) {
   cases[i,t] ~ dpois(mu[i,t])
   log(mu[i,t]) <- log(e[i]) + alpha + beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+phi[i]+nu[i]+delta[t]+omega[t]
   rho[i,t]<-exp(alpha + beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+phi[i]+nu[i]+delta[t]+omega[t]) # RR
   rhoadj[i,t]<-exp(phi[i]+nu[i]+delta[t]+omega[t]) # Adjusted RR
  }
  phi[i] ~ dnorm(0,tau.phi)
  rhoiadj[i]<-exp(phi[i]+nu[i]) # Adjusted RR over all months/years
}
nu[1:regions] ~ car.normal(adj[], weights[], num[], tau.nu)
delta[1]<-0
omega[1] ~ dnorm(0, tau.omega1)
rhotadj[1]<-exp(omega[t]) # Adjusted RR in month 1 over all districts/years
for (t in 2 :time) {
  delta[t] ~ dnorm(0,tau.delta)
  omega[t]~dnorm(omega[t-1],tau.omega)
  rhotadj[t]<-exp(delta[t]+omega[t]) # Adjusted RR in months 2-12 over all districts/years
}
alpha  ~ dflat()
beta1  ~ dnorm(0.0, 1.0E-5)
beta2  ~ dnorm(0.0, 1.0E-5)
beta3  ~ dnorm(0.0, 1.0E-5)
tau.phi  ~ dgamma(0.1,0.1)
tau.nu ~ dgamma(0.1,0.1)
tau.delta  ~ dgamma(0.1,0.1)
tau.omega1 ~ dgamma(0.1,0.1)
tau.omega ~ dgamma(0.1,0.1)
```

Concluding remarks

Concluding remarks

There are a number of issues which I have not have time to cover or comment on in this course and perhaps I should at least list one or two of these (what you do *not say* may be just as important as what you do)

There are a number of issues which I have not have time to cover or comment on in this course and perhaps I should at least list one or two of these (what you do *not say* may be just as important as what you do)

- Some doubts have been expressed about the ability of the kind of disease mapping models I have discussed to identify sharp discontinuities in disease maps (e.g. a low morbidity area surrounded by high morbidity areas). Various forms of discrete **mixture models** have been suggested as an alternative in which areas are probabilistically allocated to clusters (e.g. see Schalttmann *et al*, 1996, 1993; Lawson and Clarke, 2002)

There are a number of issues which I have not have time to cover or comment on in this course and perhaps I should at least list one or two of these (what you do *not say* may be just as important as what you do)

- Some doubts have been expressed about the ability of the kind of disease mapping models I have discussed to identify sharp discontinuities in disease maps (e.g. a low morbidity area surrounded by high morbidity areas). Various forms of discrete **mixture models** have been suggested as an alternative in which areas are probabilistically allocated to clusters (e.g. see Schalttmann *et al*, 1996, 1993; Lawson and Clarke, 2002)
- We have mostly been concerned with models in which spatially structured components have been formulated through a CAR. Alternative formulations of spatial correlation structure are possible which focus on direct parametric modelling of the variance/covariance matrix (e.g. see Leyland *et al*, 2000)

Concluding remarks

- We have focussed on Bayesian models - there are a range of alternatives which do not use a Bayesian framework (e.g. Prentice *et al*, 1995; Yasui *et al*, 1997)

Concluding remarks

- We have focussed on Bayesian models - there are a range of alternatives which do not use a Bayesian framework (e.g. Prentice *et al*, 1995; Yasui *et al*, 1997)
- A further approach has been the use of Geostatistical models (e.g. see Webster *et al*, 1994; Diggle *et al*, 1998)

Concluding remarks

- We have focussed on Bayesian models - there are a range of alternatives which do not use a Bayesian framework (e.g. Prentice *et al*, 1995; Yasui *et al*, 1997)
- A further approach has been the use of Geostatistical models (e.g. see Webster *et al*, 1994; Diggle *et al*, 1998)
- I said at the outset that I was not going to discuss methods explicitly designed to detect disease clustering, either in space or in space and time, or at focussed or unfocussed locations. There is a substantial literature on this important subject and I have included a special section of references for those who wish to follow it up

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - **Disease mapping**
  - **Ecological (correlation) studies**
  - **Further topics in ecological studies**
  - **Space-time models**
  - **Concluding remarks**

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - Disease mapping
  - Ecological (correlation) studies
  - Further topics in ecological studies
  - Space-time models
  - Concluding remarks

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - **Disease mapping**
  - Ecological (correlation) studies
  - Further topics in ecological studies
  - Space-time models
  - Concluding remarks

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - **Disease mapping**
  - **Ecological (correlation) studies**
  - Further topics in ecological studies
  - Space-time models
  - Concluding remarks

Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - **Disease mapping**
  - **Ecological (correlation) studies**
  - **Further topics in ecological studies**
  - Space-time models
  - Concluding remarks

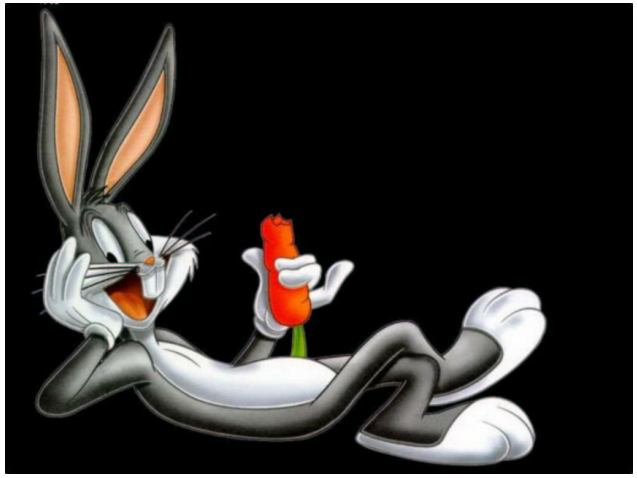
Structure of the remainder of the course

- Will attempt to briefly review **selected** topics under following headings:
  - **Preliminaries — a tour of statistical/Bayesian modelling**
  - **Disease mapping**
  - **Ecological (correlation) studies**
  - **Further topics in ecological studies**
  - **Space-time models**
  - Concluding remarks

Contact Details

Professor Trevor C Bailey
School of Engineering, Computing and Mathematics
Harrison Building
Exeter University
Exeter
EX4 4QE
U.K.

Tel: +44 (0)1392725223
Email: T.C.Bailey@ex.ac.uk
WWW: http://www.secam.ex.ac.uk/



... *THAT'S ALL FOLKS!!* ...