Estimating cancer survival in small areas: possible and useful

Susanna Cramb, Kerrie Mengersen and Peter Baade

susannacramb@cancerqld.org.au

Survival

- The proportion who survive a given length of time after diagnosis


Queensland
1.73 million km2
4.4 million people (2012)


Queensland
Most dispersed population of any State/Territory in Australia

Queensland

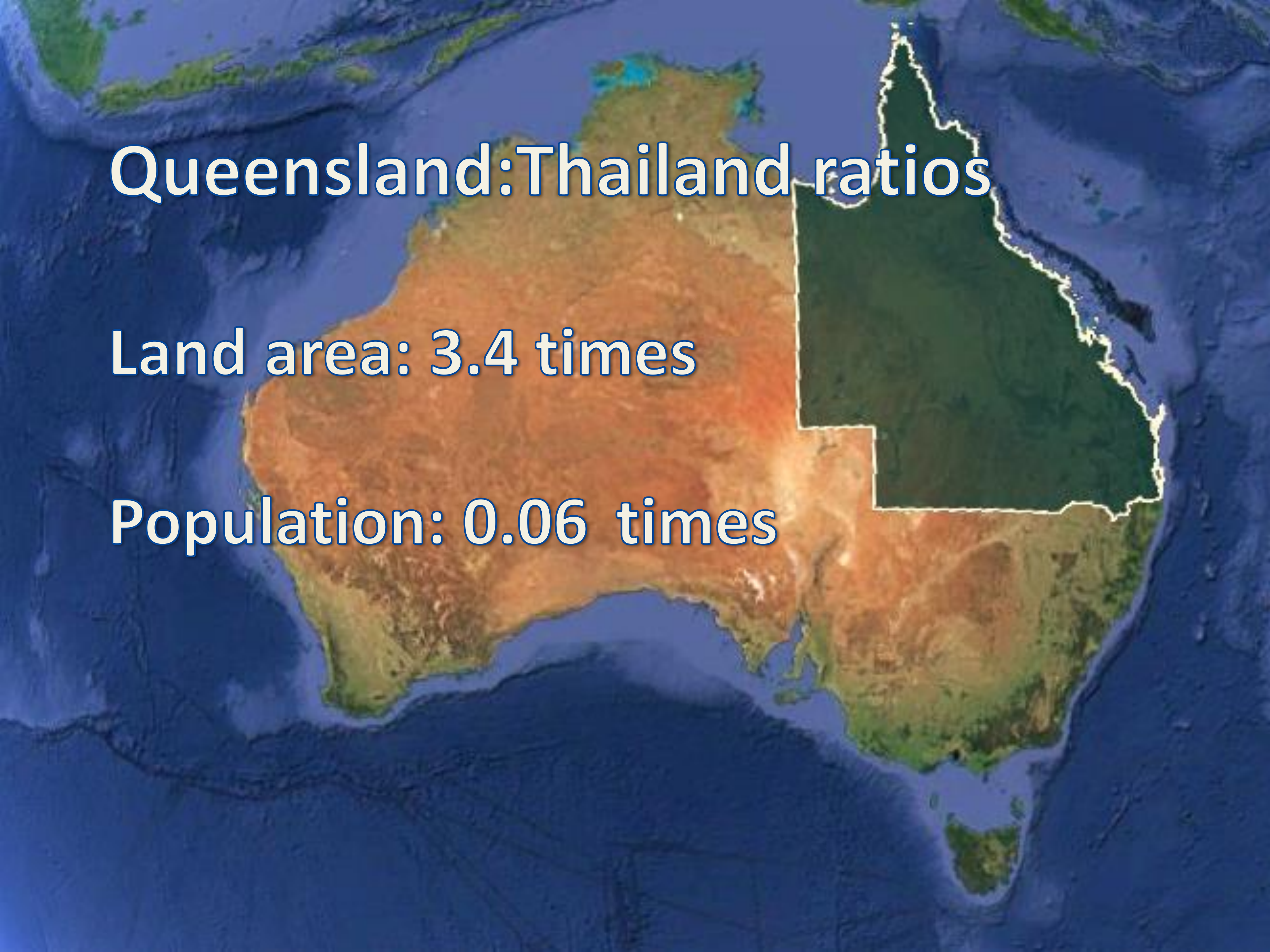
Most dispersed population of any State/Territory in Australia

Centralised health services

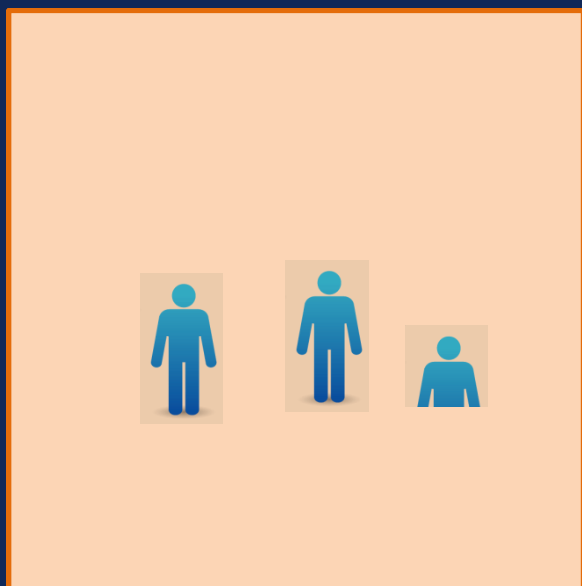
Queensland:Thailand ratios

Land area: 3.4 times

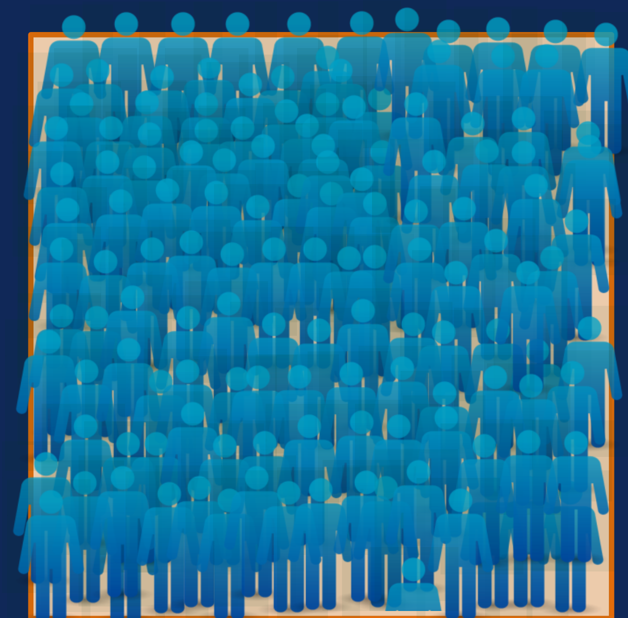
Population: 0.06 times



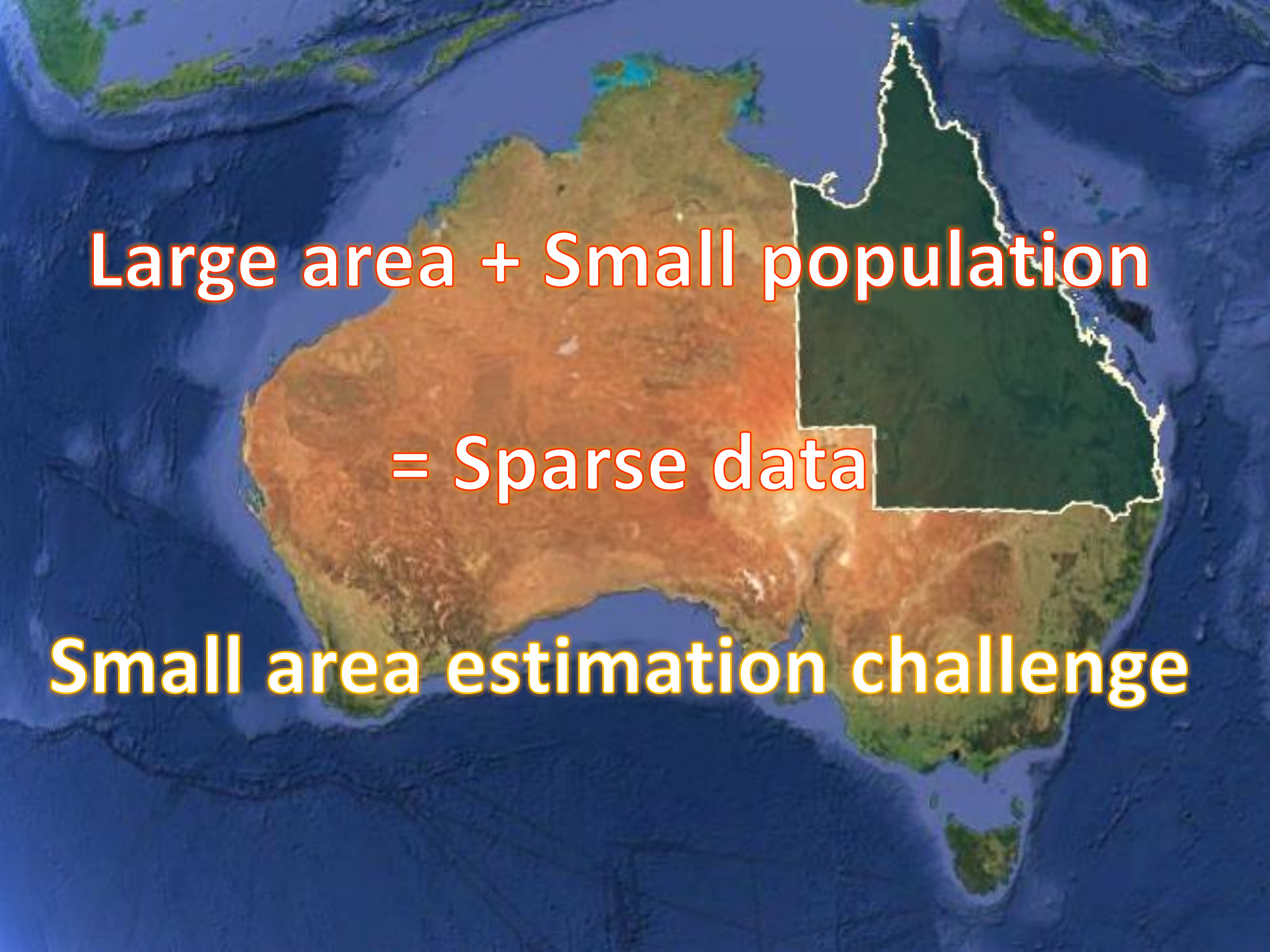
Queensland



Thailand



54 times!



Large area + Small population

= Sparse data

Small area estimation challenge

Survival

- Key measure of cancer patient care

- Allows monitoring and evaluation of health services

Estimating Net Survival

Cause-specific

Relative

Estimating Net Survival

Cause-specific

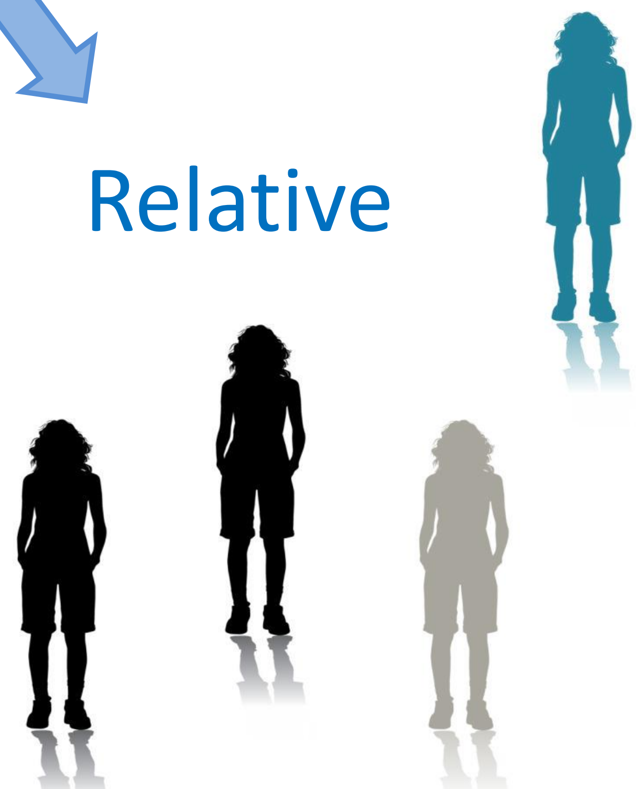
Relative

Based on death certificate

Estimating Net Survival

Cause-specific

Relative



Based on death certificate

Compares against population mortality

Estimating Net Survival

Cause-specific

Relative

Based on death certificate

Compares against population mortality

Data sources

- Cancer incidence data (contains death information)
    - Queensland Cancer Registry (population-based)
- Unit record file mortality data by age group, sex, time and area
    - Australian Bureau of Statistics
- Population data by age group, sex, time and area
    - Australian Bureau of Statistics

Queensland

478 Statistical Local Areas (SLAs) in 2006

Data preparation

1. Population mortality data

   - Create lifetables by SLA, sex and year group (e.g. 2003-2007).

_px	_lx	_Lx	_Tx	_ExpYL	_Surv
0.99907	100000	99918	7960372	79.6	1
0.99907	99907	99855	7860454	78.68	0.999074
0.99907	99815	99764	7760599	77.75	0.998148
0.99907	99722	99673	7660835	76.82	0.997223
0.99907	99630	99584	7561162	75.89	0.996299
0.99988	99538	99532	7461578	74.96	0.995375
0.99988	99526	99520	7362047	73.97	0.995259
0.99988	99514	99508	7262527	72.98	0.995143

2. Cancer incidence data

   - Calculate the person-time at risk, and the expected deaths using the lifetable data.

3. Neighbourhood adjacency matrix file

Data preparation

1. Population mortality data

   - Create lifetables by SLA, sex and year group (e.g. 2003-2007).

_px	_lx	_Lx	_Tx	_ExpYL	_Surv
0.99907	100000	99918	7960372	79.6	1
0.99907	99907	99855	7860454	78.68	0.999074
0.99907	99815	99764	7760599	77.75	0.998148
0.99907	99722	99673	7660835	76.82	0.997223
0.99907	99630	99584	7561162	75.89	0.996299
0.99988	99538	99532	7461578	74.96	0.995375
0.99988	99526	99520	7362047	73.97	0.995259
0.99988	99514	99508	7262527	72.98	0.995143

2. Cancer incidence data

   - Calculate the person-time at risk, and the expected deaths using the lifetable data.

3. Neighbourhood adjacency matrix file

Data preparation

1. Population mortality data

    - Create lifetables by SLA, sex and year group (e.g. 2003-2007).

_px	_lx	_Lx	_Tx	_ExpYL	_Surv
0.99907	100000	99918	7960372	79.6	1
0.99907	99907	99855	7860454	78.68	0.999074
0.99907	99815	99764	7760599	77.75	0.998148
0.99907	99722	99673	7660835	76.82	0.997223
0.99907	99630	99584	7561162	75.89	0.996299
0.99988	99538	99532	7461578	74.96	0.995375
0.99988	99526	99520	7362047	73.97	0.995259
0.99988	99514	99508	7262527	72.98	0.995143

2. Cancer incidence data

    - Calculate the person-time at risk, and the expected deaths using the lifetable data.

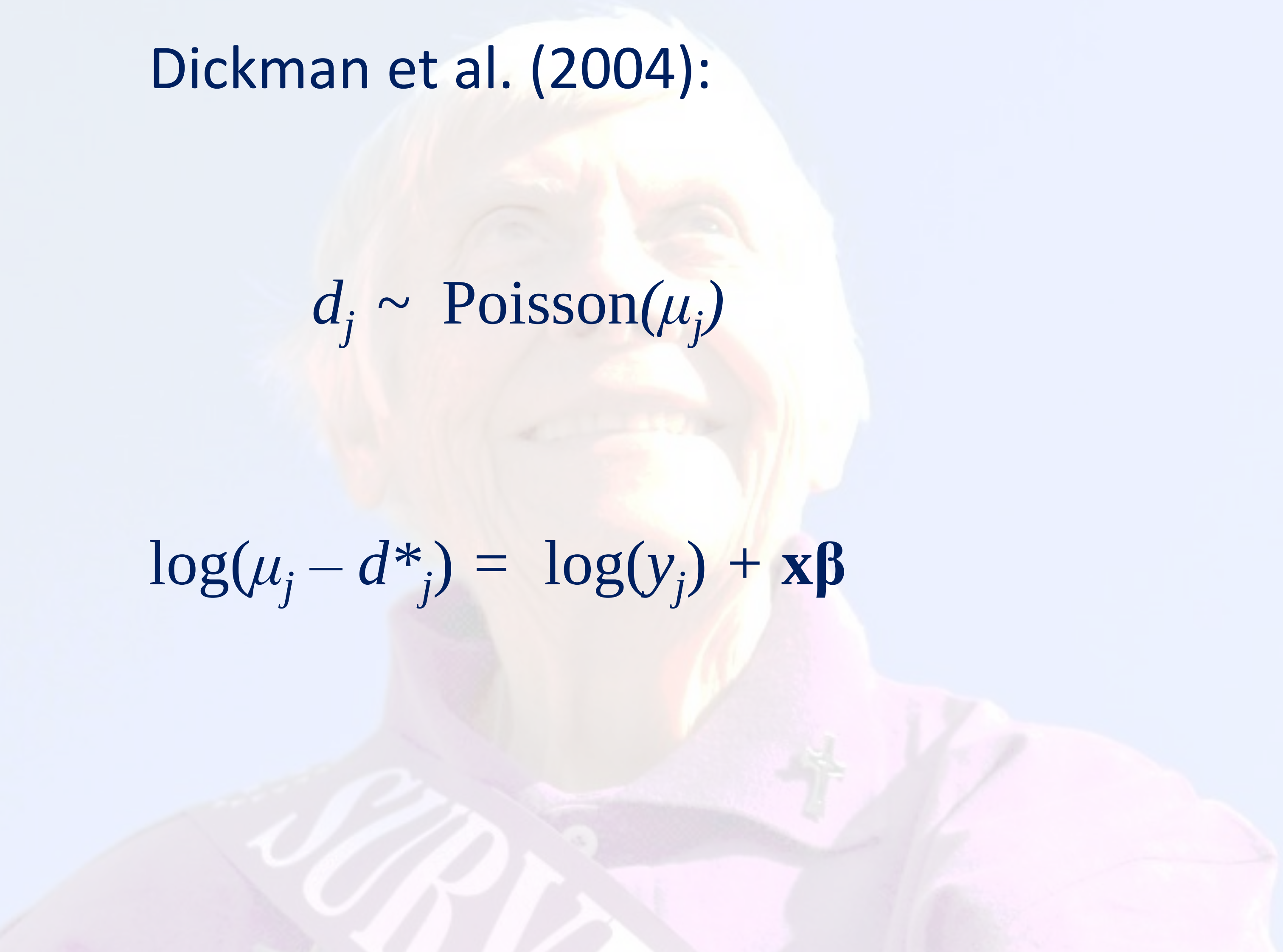
3. Neighbourhood adjacency matrix file

Relative survival model

Dickman et al. (2004):

$$d_j \sim \text{Poisson}(\mu_j)$$

$$\log(\mu_j - d^*_j) = \log(y_j) + \mathbf{x}\boldsymbol{\beta}$$



Relative survival model

Dickman et al. (2004):

Observed deaths

$$d_j \sim \text{Poisson}(\mu_j)$$

Covariate parameters

$$\log(\underbrace{\mu_j - d^*_j}_{\text{Excess deaths}}) = \log(y_j) + \mathbf{x}\boldsymbol{\beta}$$

Person-time at risk

Bayesian relative survival model

Based on Fairley *et al* (2008):

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + u_i + v_i$$

where k = broad age groups

$j = 1,2,\ldots 5$ follow-up years

$i = 1,2,\ldots 478$ SLAs

Bayesian relative survival model

Based on Fairley *et al* (2008):

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

Intercept

Unobserved and unstructured

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + u_i + v_i$$

Unobserved with spatial structure

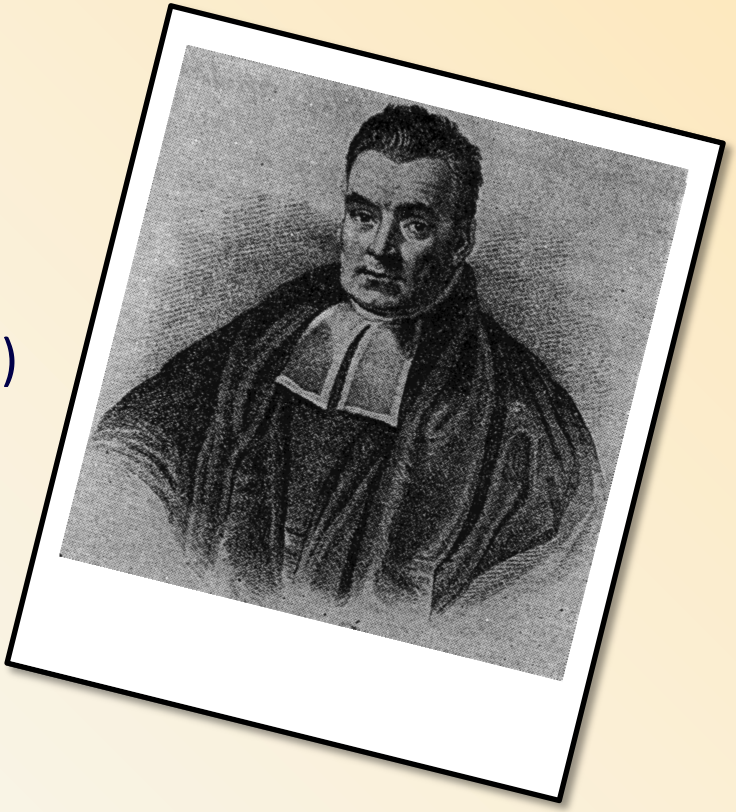
where k = broad age groups

$j = 1,2,\ldots 5$ follow-up years

$i = 1,2,\ldots 478$ SLAs

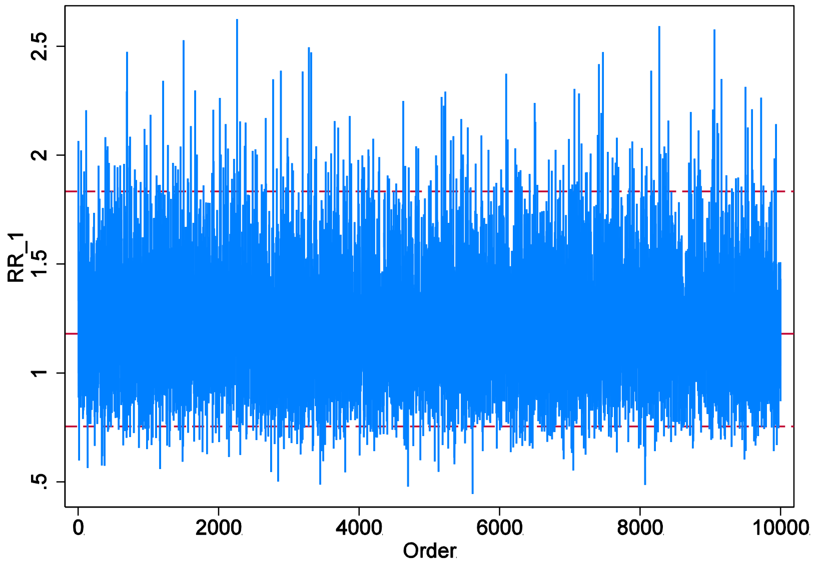
The Bayesian difference

- Parameters considered to arise from underlying distribution ("stochastic")
- Use probability distributions ("priors")
- Simplifies inclusion of spatial relationships
- Posterior distributions for output parameters
- Posterior proportional to Likelihood x Prior

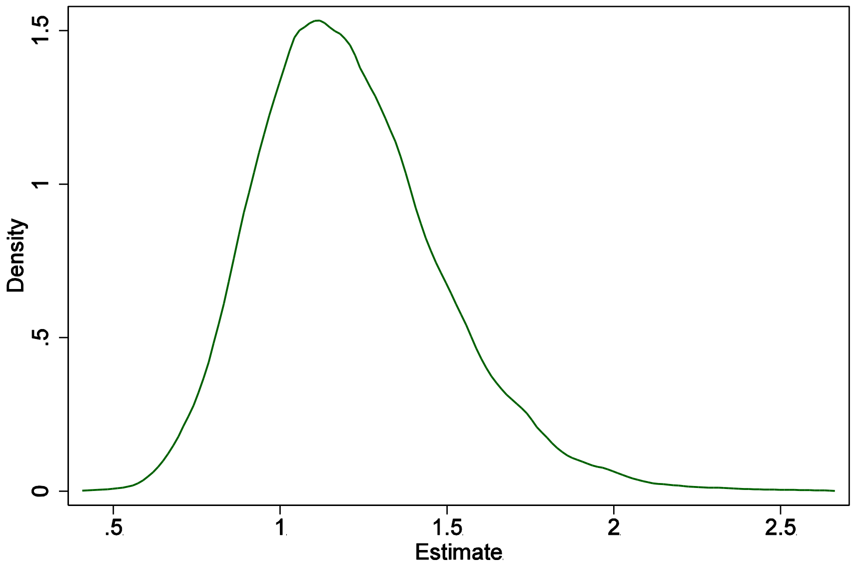


Posterior distributions

Trace plot



Density plot



Bayesian relative survival model

Based on Fairley *et al* (2008):

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + u_i + v_i$$

where k = broad age groups

$j = 1,2,\ldots5$ follow-up years

$i = 1,2,\ldots478$ SLAs

Bayesian relative survival model

Based on Fairley *et al* (2008):

e.g. ~Normal(0,1000)

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + u_i + v_i$$

where k = broad age groups

$j = 1,2,\ldots 5$ follow-up years

$i = 1,2,\ldots 478$ SLAs

Bayesian relative survival model

Based on Fairley *et al* (2008):

e.g. ~Normal(0,1000)

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + u_i + v_i$$

CAR prior

where k = broad age groups

$j = 1,2,\ldots 5$ follow-up years

$i = 1,2,\ldots 478$ SLAs

The Conditional AutoRegressive (CAR) distribution

Area full conditional distributions:

$$p(u_i | u_j, i \neq j, \sigma^2) \sim N\left(\bar{\mu}_i, \frac{\sigma^2}{n_{\delta_i}}\right)$$

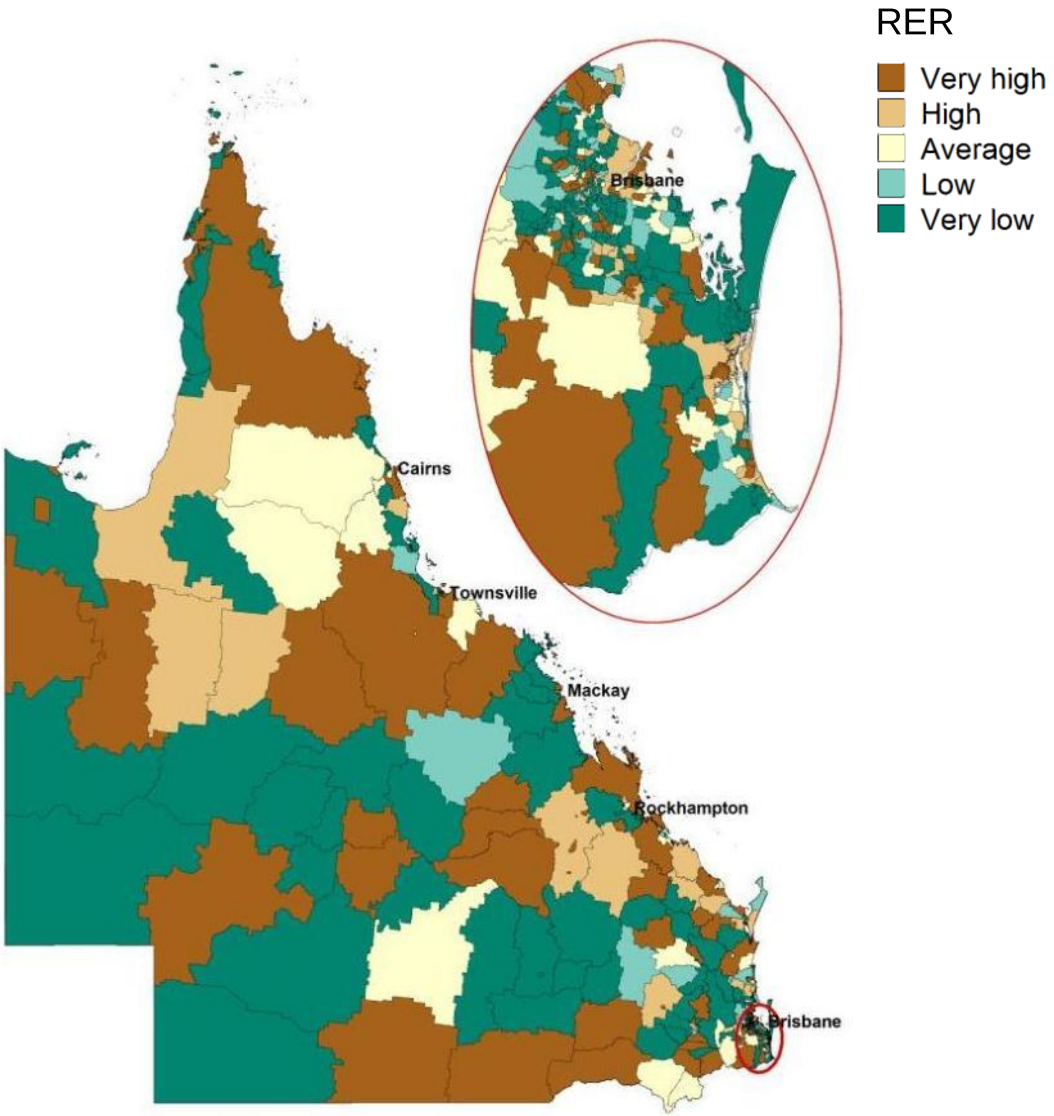
$$\bar{\mu}_i = \sum_{j \in \delta_i} \frac{u_j}{n_{\delta_i}}$$

n_{δ_i} = number of neighbours

σ^2 = variance

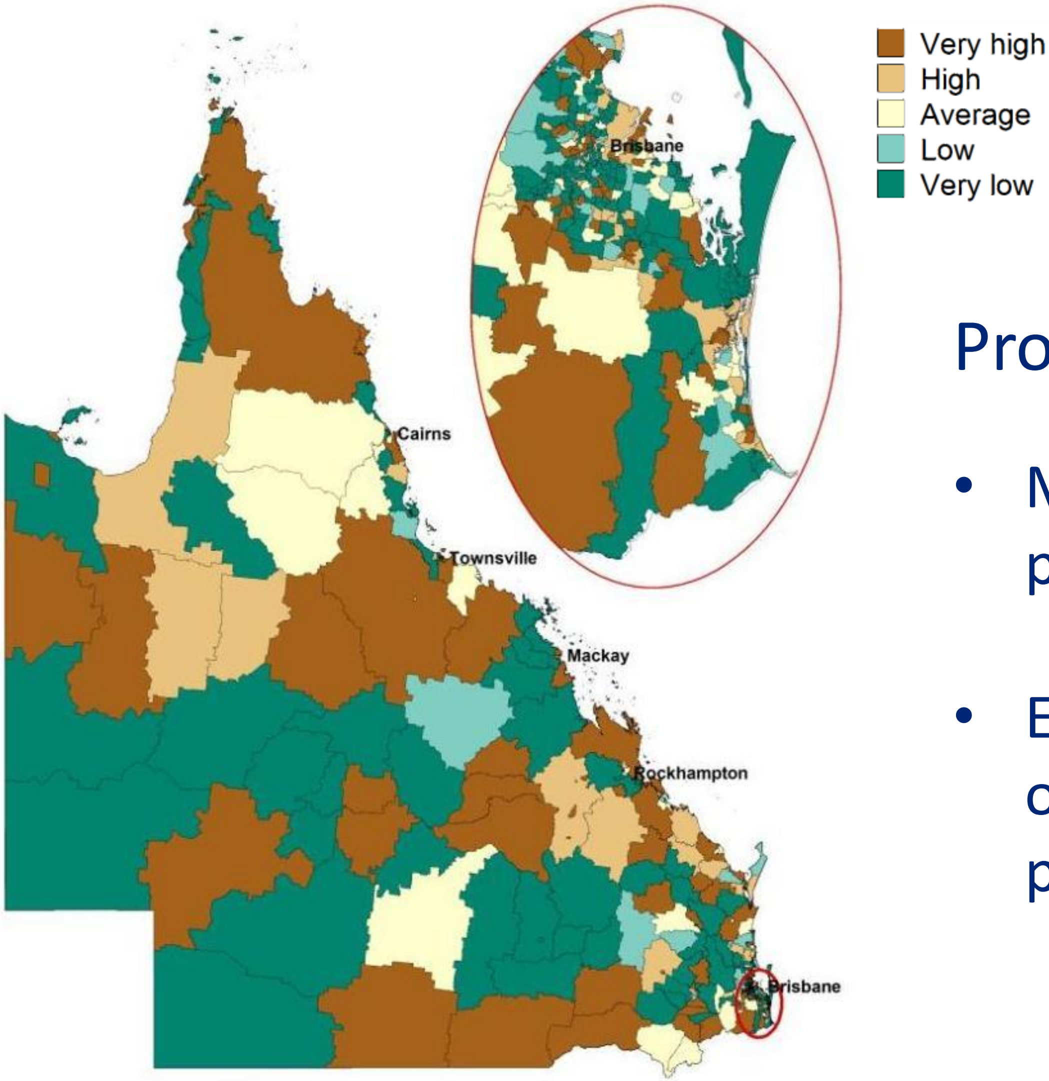
Breast cancer survival (risk of death within 5 years)

Raw estimates



Breast cancer survival (risk of death within 5 years)

Raw estimates



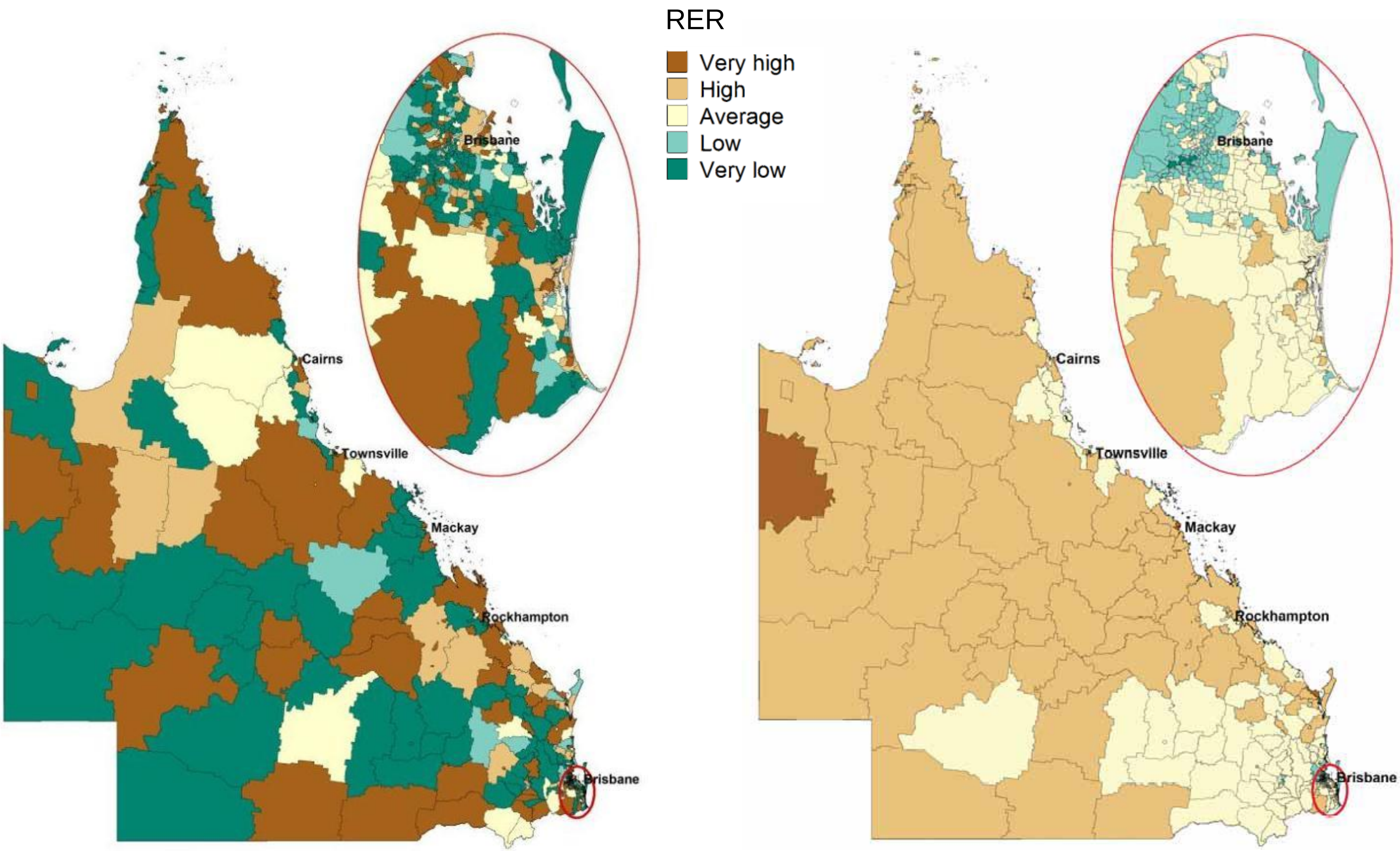
Problems

- Many large areas have small populations (and vice versa)
- Excessive random variation – obscures the true geographic pattern

Breast cancer survival (risk of death within 5 years)

Raw estimates

Smoothed estimates



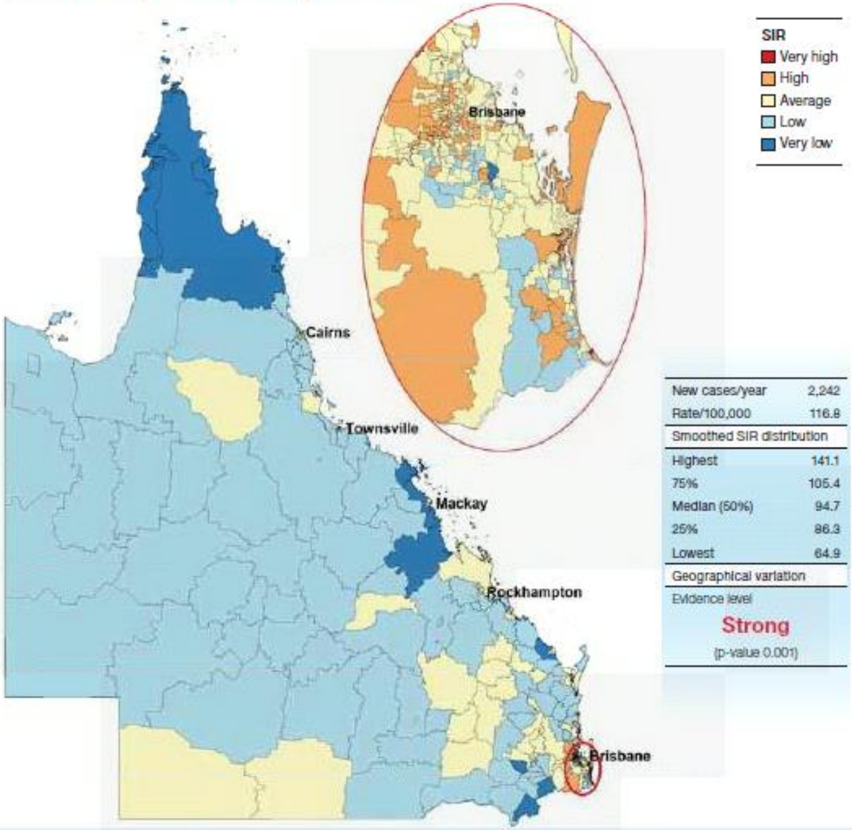
Results and Benefits

This model allows us to determine:

- Robust small area estimates with uncertainty
- Influence of important covariates
- Probabilities (e.g. probability RER > 1)
- Ranking
- Number of deaths resulting from spatial inequalities

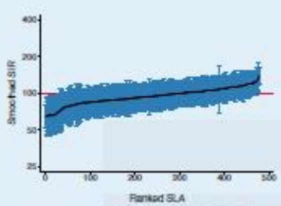
Breast cancer

Risk of diagnosis among females



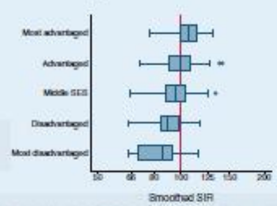
New cases/year	2,242
Rate/100,000	116.8
Smoothed SIR distribution	
Highest	141.1
75%	105.4
Median (50%)	94.7
25%	86.3
Lowest	64.9
Geographical variation	
Evidence level	
Strong	
(p-value 0.001)	

Level of Uncertainty

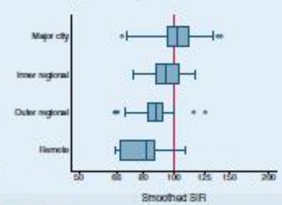


Distribution of smoothed SIR estimates according to:

(a) Socioeconomic status



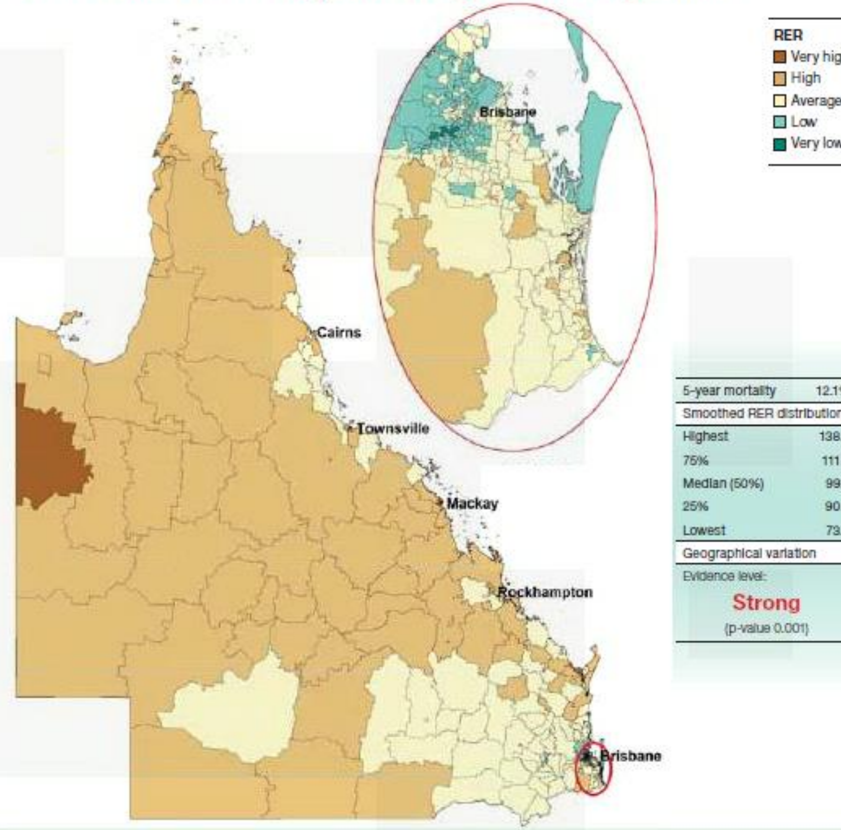
(b) Rurality



Notes: Smoothed SIR (Standardised Incidence Ratio) estimates are in comparison to the Queensland average (red line on graphs), and should not be directly compared between SLAs (Statistical Local Areas). Data are for cases diagnosed between 1998 and 2007.

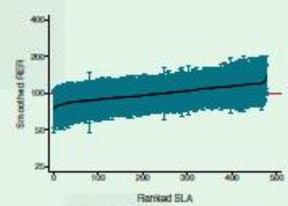
Breast cancer

Risk of death within five years of diagnosis among females



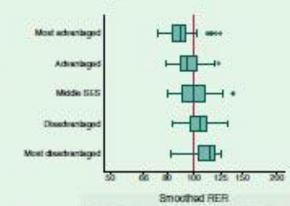
5-year mortality	12.1%
Smoothed RER distribution	
Highest	138.9
75%	111.4
Median (50%)	99.6
25%	90.6
Lowest	73.8
Geographical variation	
Evidence level:	
Strong	
(p-value 0.001)	

Level of Uncertainty

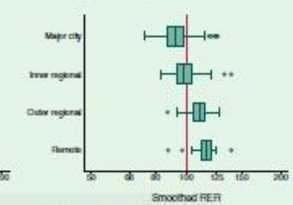


Distribution of smoothed RER estimates according to:

(a) Socioeconomic status

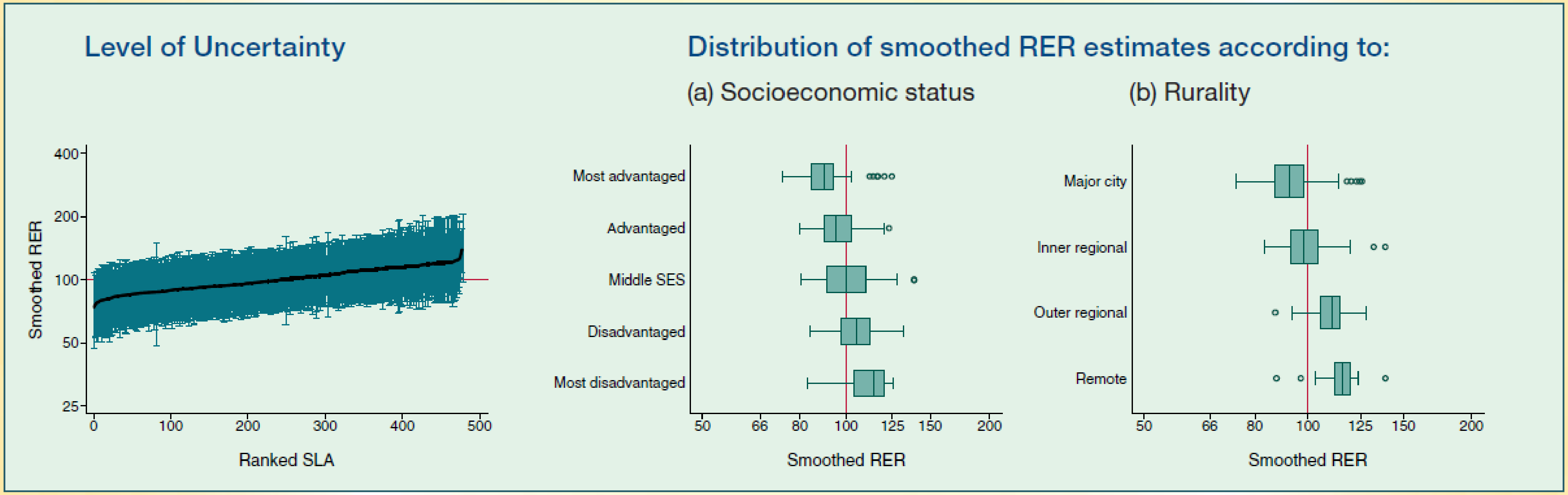


(b) Rurality



Notes: Smoothed RER (Relative Excess Risk) estimates are in comparison to the Queensland average (red line on graphs), and should not be directly compared between SLAs (Statistical Local Areas). Data are for 'at risk' cases in the period 1998 and 2007.

Graphs



Level of Uncertainty
Smoothed RER
400
200
100
50
25
0
100
200
300
400
500
Ranked SLA
Distribution of smoothed RER estimates according to:
(a) Socioeconomic status
Most advantaged
Advantaged
Middle SES
Disadvantaged
Most disadvantaged
50
66
80
100
125
150
200
Smoothed RER
(b) Rurality
Major city
Inner regional
Outer regional
Remote
50
66
80
100
125
150
200
Smoothed RER

Bayesian relative survival model

Breast and colorectal cancers

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + v_i + u_i$$

where k = broad age groups/SES/remoteness/stage/gender

$j = 1,2,\ldots5$ follow-up years

$i = 1,2,\ldots478$ SLAs

Bayesian relative survival model

Breast and colorectal cancers

$$d_{kji} \sim \text{Poisson}(\mu_{kji})$$

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \alpha_j + x\beta_k + v_i + u_i$$

where k = broad age groups/SES/remoteness/stage/gender

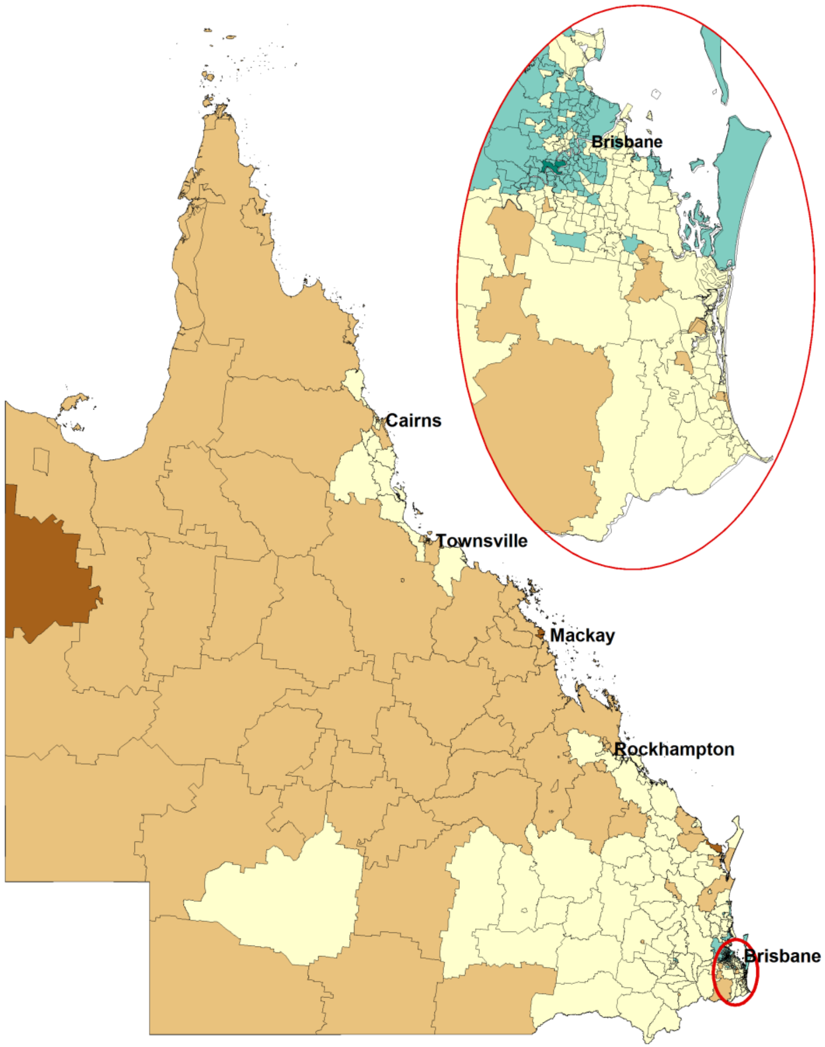
$j = 1,2,\ldots 5$ follow-up years

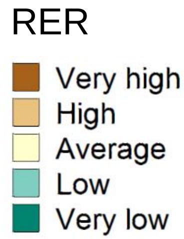
$i = 1,2,\ldots 478$ SLAs

Breast cancer survival (risk of death within 5 years)

Adjusted for age

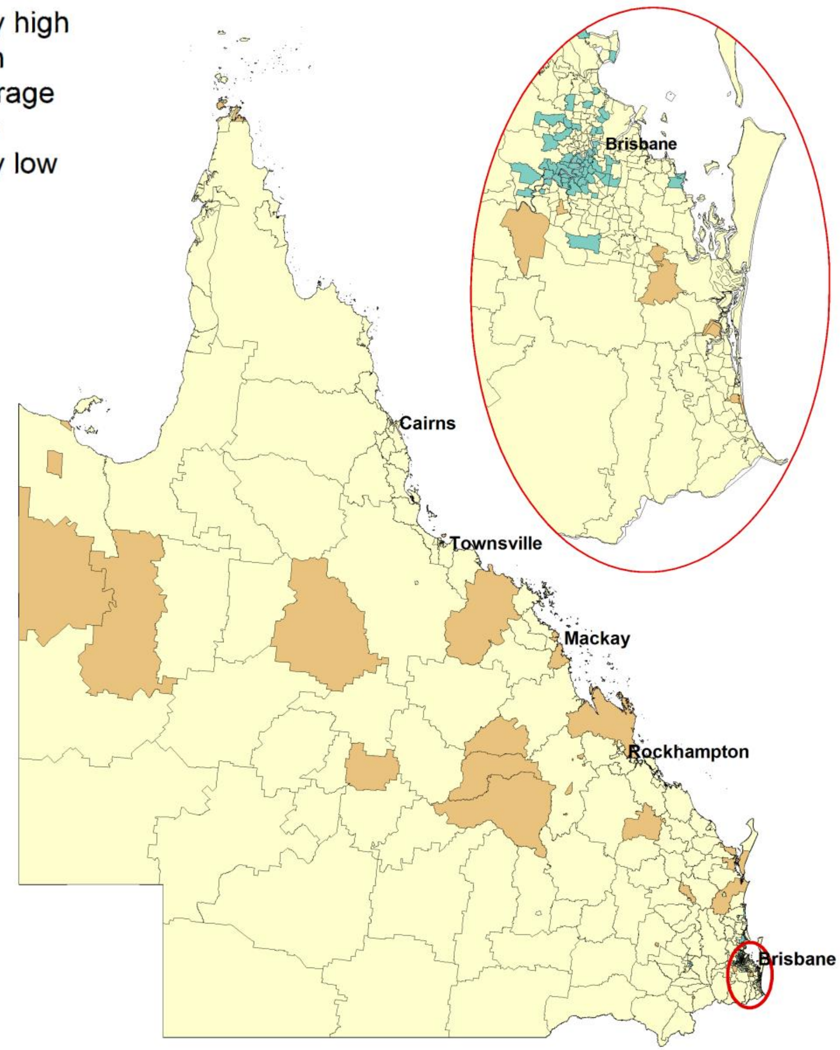
Spatial variation p-value=0.001





Adjusted for age & stage

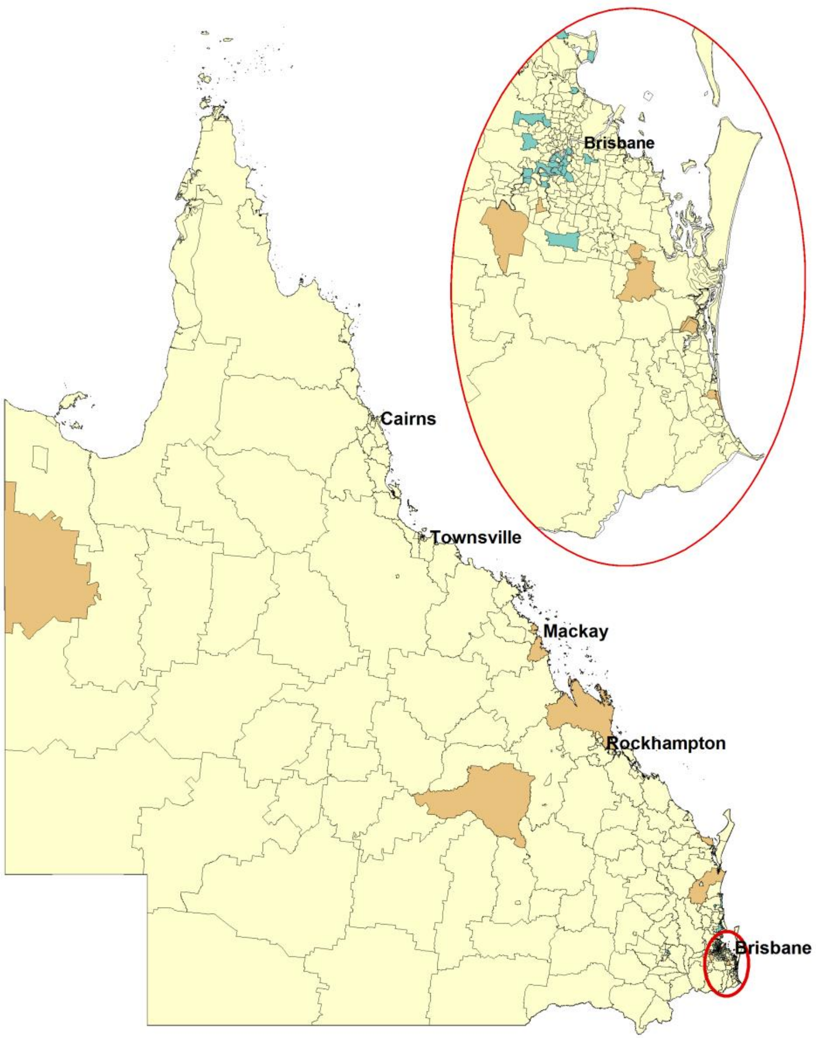
Spatial variation p-value=0.042



Breast cancer survival (risk of death within 5 years)

Adjusted for age, stage & SES

Spatial variation p-value=0.452

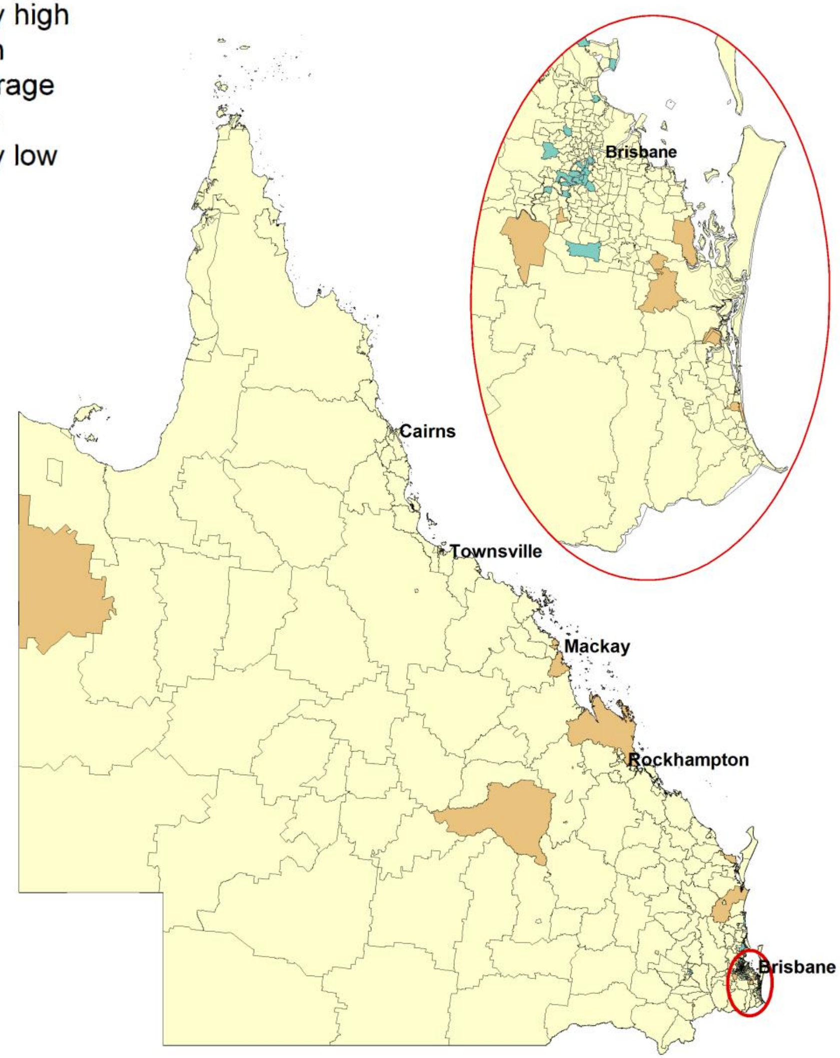


RER

- Very high
- High
- Average
- Low
- Very low

Adjusted for age , stage, SES & distance

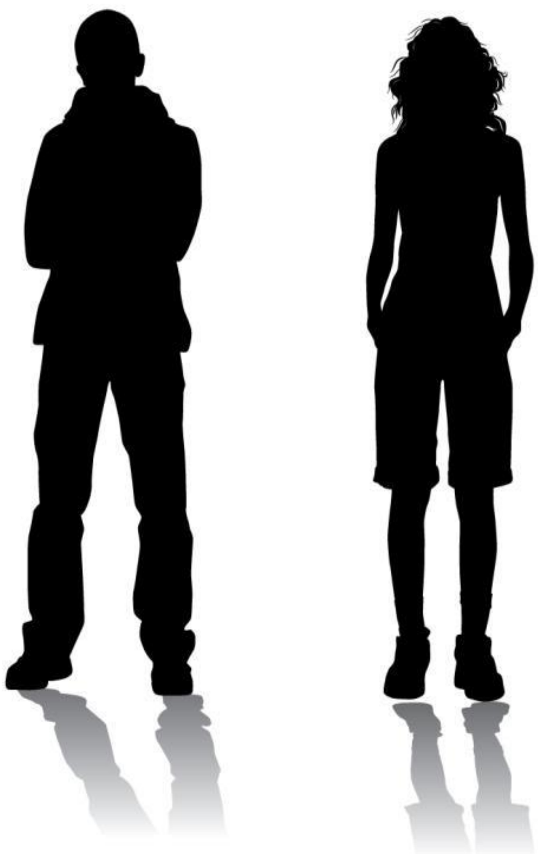
Spatial variation p-value=0.631



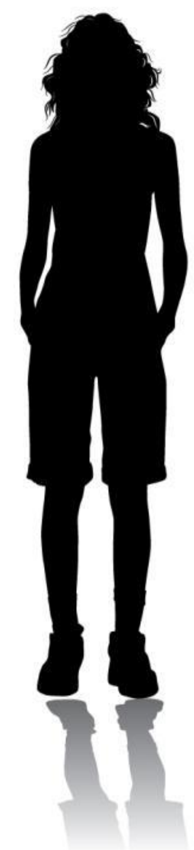
How many deaths could be prevented if no spatial inequalities?

Number of deaths within 5 years from diagnosis due to non-diagnostic spatial inequalities (1997-2008):

Colorectal cancer:



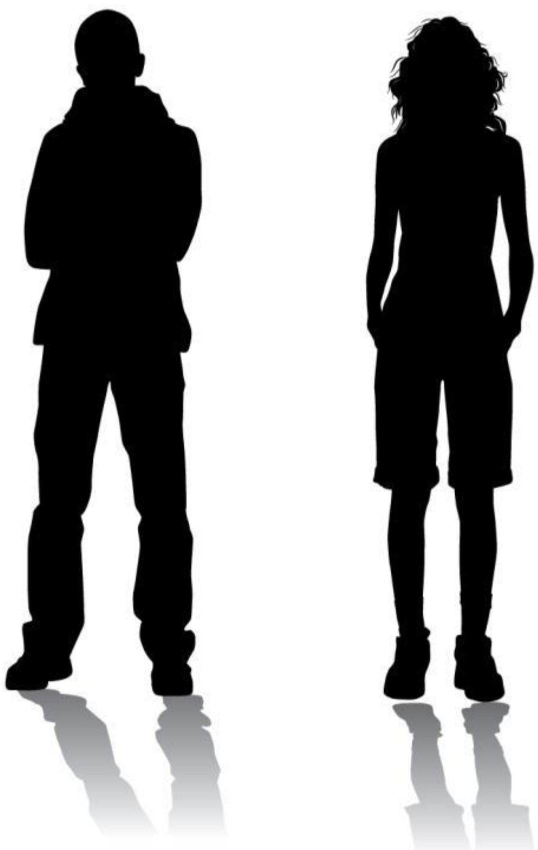
Breast cancer:



How many deaths could be prevented if no spatial inequalities?

Number of deaths within 5 years from diagnosis due to non-diagnostic spatial inequalities (1997-2008):

Colorectal cancer: 470 (7.8%)

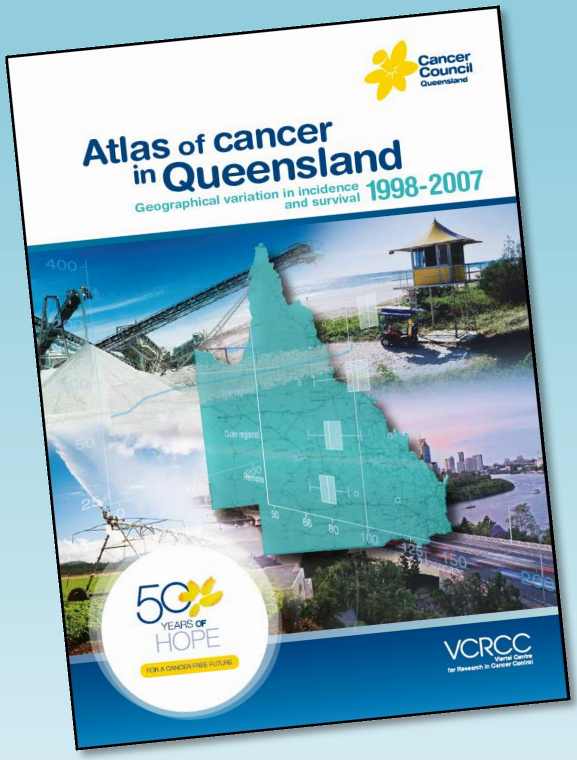


Breast cancer: 170 (7.1%)

Implementation

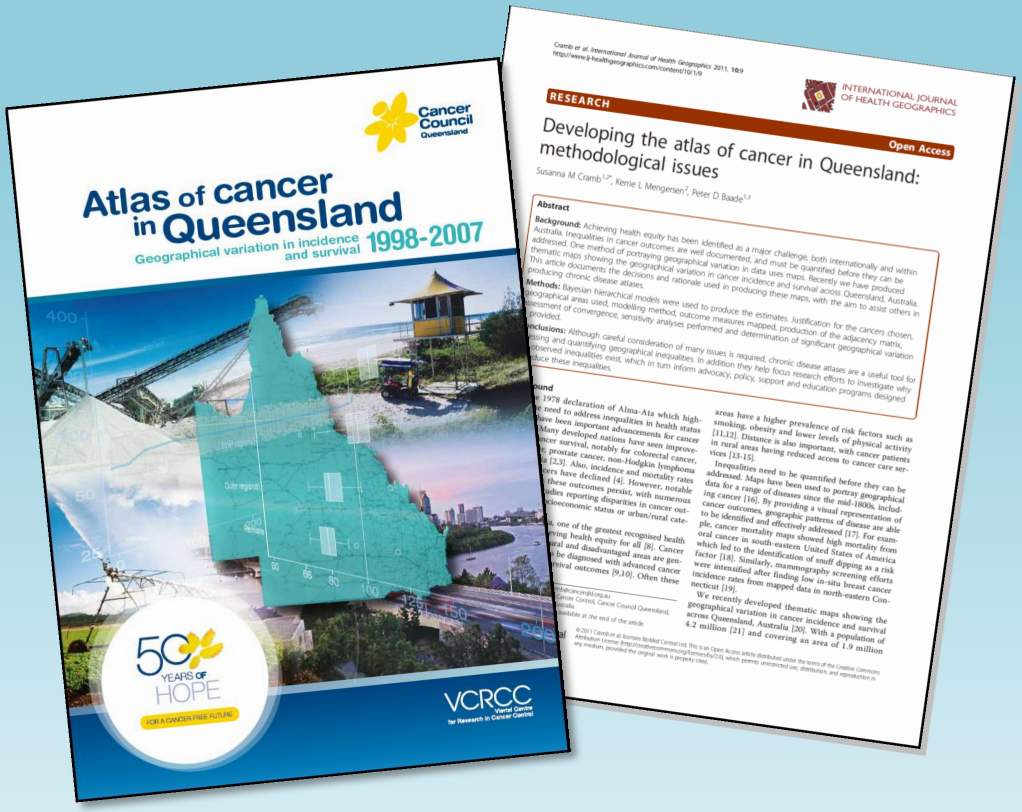
- Neighbourhood matrix created in GeoDa (https://geodacenter.asu.edu/)
- Ran in WinBUGS (Bayesian inference Using Gibbs Sampling) interfaced with Stata
    - Freely available at: www.mrc-bsu.cam.ac.uk/bugs
    - 250,000 iterations discarded, 100,000 iterations monitored (kept every 10th)
    - Time taken: 3 hours 15 minutes+
- On a dedicated server:
    - Dual CPU Quad Core Xeon E5520's: 8 Cores and 16 Threads, large 8MB Cache
    - Quick Path Interconnect: fast memory access

Cramb SM, Mengersen KL, Baade PD. 2011. The Atlas of Cancer in Queensland: Geographical variation in incidence and survival, 1998-2007. Cancer Council Queensland: Brisbane.



Cramb SM, Mengersen KL, Baade PD. 2011. The Atlas of Cancer in Queensland: Geographical variation in incidence and survival, 1998-2007. Cancer Council Queensland: Brisbane.

Cramb SM, Mengersen KL, Baade PD. 2011. Developing the atlas of cancer in Queensland: methodological issues. *Int J Health Geogr*, 10:9



Cramb SM, Mengersen KL, Baade PD. 2011. The Atlas of Cancer in Queensland: Geographical variation in incidence and survival, 1998-2007. Cancer Council Queensland: Brisbane.

Cramb SM, Mengersen KL, Baade PD. 2011. Developing the atlas of cancer in Queensland: methodological issues. *Int J Health Geogr*, 10:9

Cramb SM, Mengersen KL, Turrell G, Baade PD. 2012. Spatial inequalities in colorectal and breast cancer survival: Premature deaths and associated factors. *Health & Place*;18:1412-21.

Cramb SM, Mengersen KL, Baade PD. 2011. The Atlas of Cancer in Queensland: Geographical variation in incidence and survival, 1998-2007. Cancer Council Queensland: Brisbane.

Cramb SM, Mengersen KL, Baade PD. 2011. Developing the atlas of cancer in Queensland: methodological issues. *Int J Health Geogr*, 10:9

Cramb SM, Mengersen KL, Turrell G, Baade PD. 2012. Spatial inequalities in colorectal and breast cancer survival: Premature deaths and associated factors. *Health & Place*;18:1412-21.

Earnest A, Cramb SM, White NM. 2013. Disease mapping using Bayesian hierarchical models. In Alston CL, Mengersen KL, Pettitt AN (eds): Case Studies in Bayesian Statistical Modelling and Analysis, Wiley: Chichester.

"By increasing our understanding of the small area inequalities in cancer outcomes, this type of innovative modelling provides us with a better platform to influence government policy, monitor changes, and allocate Cancer Council Queensland resources"

~ Professor Jeff Dunn,
Cancer Council Queensland CEO

Cancer Council Queensland

QUT

crc•si