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
- Based on death certificate

Relative
- Compares against population mortality
Estimating Net Survival

Cause-specific

Based on death certificate

Relative

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).

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).

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).

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):

\[ d_j \sim \text{Poisson}(\mu_j) \]

\[ \log(\mu_j - d^*_j) = \log(y_j) + x\beta \]
Relative survival model

Dickman et al. (2004):

\[ d_j \sim \text{Poisson}(\mu_j) \]

\[ \log(\mu_j - d^*_j) = \log(y_j) + x\beta \]

Observed deaths
Covariate parameters
Excess deaths
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):

\[ d_{kji} \sim \text{Poisson}(\mu_{kji}) \]

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

where \( k \) = broad age groups

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

\( i = 1,2,\ldots,478 \) SLAs
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

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

\( i = 1,2,\ldots,478 \) SLAs
Bayesian relative survival model

Based on Fairley et al (2008):

\[ d_{kji} \sim \text{Poisson}(\mu_{kji}) \]

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

where \( k = \) broad age groups

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

\( i = 1,2,\ldots,478 \) SLAs

e.g. \( \sim \text{Normal}(0,1000) \)
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

e.g. \(~\text{Normal}(0,1000)\)

CAR prior
The Conditional AutoRegressive (CAR) distribution

Area full conditional distributions:

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

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

\[ n_{\delta_i} = \text{number of neighbours} \]

\[ \sigma^2 = \text{variance} \]
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

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

[Maps showing various areas with different color codes representing RER]
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
Graphs

Level of Uncertainty

Distribution of smoothed RER estimates according to:
(a) Socioeconomic status
(b) Rurality
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
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 \( k = \text{broad age groups/SES/remoteness/stage/gender} \)

\( j = 1,2,\ldots,5 \text{ follow-up years} \)

\( i = 1,2,\ldots,478 \text{ SLAs} \)
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

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


“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