Methods

Full details of the data sources and statistical analyses are described in Appendix B. A summary of the methodology is provided below.

Data sources

De-identified data on all cancers diagnosed among people living in Queensland during 1996 to 2007 were obtained from the Queensland Cancer Registry (QCR). The QCR is a population-based cancer registry that maintains a register of all cancers (excluding basal and squamous cell carcinomas) diagnosed among Queensland residents since 1982. Ethical approval to conduct this study was obtained from the Queensland Health Human Research Ethics Committee. Approval to extract the data was obtained from Queensland Health.

Population estimates[14,15] and general population mortality data[16] were obtained from the Australian Bureau of Statistics.

Geographical areas

In 2006, there were 478 SLAs in Queensland defined by the ASGC.[17] Incident cancer cases were assigned to an SLA based on place of residence at diagnosis. To account for changes in SLA boundaries over time, the SLA definitions for people diagnosed in other years were adjusted to the 2006 ASGC definition using suburb and postcode at diagnosis. Boundary adjustments could only be made from 1996 onwards due to major differences in the SLA definitions prior to that date. This adjustment of SLA definitions was conducted within the Queensland Cancer Registry before the data were extracted for analysis. SLAs were also grouped into broad categories of rurality (using the ARIA+ classification[18]) and area-level socioeconomic status (using the IRSAD[19]) (Appendix D).

Statistical analysis

When examining cancer data by small geographical areas, crude estimates tend to be unreliable and fluctuate widely due to the few cases observed among a small population. Since neighbouring SLAs are likely to have similar characteristics, statistical methods that “borrow strength” from the data in these neighbouring SLAs have been shown to produce more reliable estimates than those methods relying solely on the data within a specific SLA. One such method is Bayesian hierarchical modelling.

The effect of using Bayesian hierarchical models is to “smooth” the estimate of incidence or survival for a particular SLA towards the State average and the average of the surrounding (or neighbourhood) areas. For some areas, even though the crude estimate might be higher than the Queensland average, the impact of the neighbouring areas may mean that the smoothed estimate is lower than the State average, and vice versa. Generally the “smoothing” effect is more pronounced when there are a smaller number of cases in a particular geographical area.

The statistical evidence for spatial variation was assessed using Tango’s Maximised Excess Events Test (MEET).[20] A low p-value (< 0.05) from this test suggests that the observed geographical differences are likely to be real. Higher p-values (≥ 0.05) suggest that chance is more likely to be a plausible explanation for any apparent variation. The statistical evidence for spatial variation was categorised into “Strong” ($p < 0.01$), “Moderate” ($0.01 \leq p < 0.05$), “Weak” ($0.05 \leq p < 0.10$) and “None” ($p \geq 0.10$).

Incidence

Incidence refers to the number of new cancer cases diagnosed within a certain time period. All primary invasive cancers diagnosed in the 10-year period between 1998 and 2007 were included. Since variation between geographical areas may be simply due to differences in the age distribution of the population, incidence rates were standardised by age and sex. Due to the small number of cancer cases in some geographical areas indirect standardisation was used.

Indirectly standardised incidence ratios (SIR) were calculated for each SLA by dividing the observed number of cancer cases by the expected number and multiplying the result by 100, where the expected number of cases was calculated by applying the age- and sex-specific incidence rates for total Queensland to the corresponding components of the SLA population.

Smoothed SIR estimates were then generated by entering the components of the ‘crude’ SIR (i.e. observed and expected cases) into a specific type of Bayesian model known as the Besag, York and

Mollié (BYM) model.[21] This model is currently the standard Bayesian model for disease mapping research studies.[22] For this analysis we have not incorporated a time component into the model to see if the geographical variation has changed over time, however this is an avenue for future investigation.

When there was strong or moderate evidence for spatial variation in cancer incidence, the combined observed and expected counts from the Bayesian model were used to calculate the overall risk of being diagnosed with cancer by broad rurality and area-level socioeconomic categories.

Survival

Relative survival compares the survival of cancer patients against a comparable group from the general population, taking into account age, sex and year of diagnosis. Relative survival is the preferred measure of estimating survival from population-based Cancer Registry data as it removes the impact of any inaccuracies inherent in cause of death coding while still providing an estimate of the mortality burden caused by the specific cancer.[23]

Cancer patients were considered "at risk" of death if they were diagnosed with cancer between 1996 and 2007, and were a prevalent case (that is, were alive after being diagnosed with cancer) for at least some time between 1 January 1998 to 31 December 2007. These "at risk" patients were included in the relative survival calculations, with survival calculated up to five years after diagnosis using the period method.

As is standard practice, reporting of survival information is expressed in terms of Relative Excess Risk of death. Areas with lower survival are those that have higher excess risk of death, and those areas that have higher survival will have lower excess deaths.

The "five-year mortality" is the complement of survival (i.e. one minus relative survival), and is expressed as a percentage. This represents the percentage of patients who died within five years after diagnosis in the hypothetical situation where the cancer of interest is the only possible cause of death.

The Bayesian model used for this part of the analysis was based on the relative survival model recommended by Dickman et al,[23] including additional random effects to account for differences in the geographical areas.[24] The model assumes constant hazards within each follow-up time (years) and was adjusted for age group. Survival estimates from this model were presented in terms of Relative Excess Risk (RER), which reflect the ratio of the smoothed estimate of excess deaths in a specific SLA to the Queensland average.

For cancers with strong or moderate evidence of spatial variation in cancer survival, the overall excess risk of death by broad rurality and area-level socioeconomic categories was calculated using the observed and expected number of deaths within five years of diagnosis from the Bayesian model, as well as an estimate of the number of excess deaths that could be attributable to geographic location.