Projecting the future burden of lung cancer using Bayesian age-period-cohort models

Jessica Cameron1,2, Cynthia Lu1 and Peter Baade1,2
1. Viertel Cancer Research Centre, Cancer Council Queensland
2. School of Mathematics and Statistics, Queensland University of Technology
jessicacameron@cancerqld.org.au

Aim: Project the future burden of cancer

Motivation: Identify opportunities for targeted campaigns to reduce the burden of cancer

Guide resource management

Background

Age-period-cohort (APC) models can be used to project the incidence of cancer, using age, period and birth cohort as surrogate measures of latent processes that affect cancer risks. APC models were fitted to incident count data for seven different cancers. This poster focuses on methods and results for lung cancer.

Original model

The number of cases, \( y_{ij} \), for age group \( i \) and period \( j \) was modelled using the population at risk, \( R_{ij} \), and incidence rate, \( \lambda_{ij} \):

\[
\frac{y_{ij}}{R_{ij}} = \text{Poisson}(\lambda_{ij}R_{ij})
\]

where \( \lambda_{ij} \) is an overdispersion parameter. Birth cohort, \( c \), is a linear combination of age group and period, \( b = \frac{age - max age group index + 1}{age group number} \) so that the maximum age group index and \( F \) is the number of periods (for example, years) per age group.

To ensure identifiability, the effects were constrained to sum to zero, \( \sum_{i} \beta_i = 0 \), \( \sum_{j} \gamma_j = 0 \), and \( \sum_{c} \zeta_c = 0 \). Second-order random walks (RW2) priors were applied to the age, period and cohort effects to account for the autocorrelation between consecutive age or period groups. The intercept and overdispersion effects were given normal priors, the latter having a mean of zero.

Alternative models

Previously, Raftery and colleagues3 found that projections were strengthened by stratifying by region. Papola et al.4 included sex-specific intercepts and age and period effects, with spatially-structured and structured random effects and a space-time interaction. Finally, Smith and Wakefield proposed including slopes, a sex, a period, and age group period.

In the absence of pooled data, we attempted to add strength to the models by stratifying by sex. Using the code of Papola et al.,3 we devised model B with a sex-specific, \( \alpha_{ij} \), intercept:

\[
\lambda_{ij} = \alpha_{ij} + \beta_i + \gamma_j + \tau_c + \epsilon_{ij}
\]

Additionally interaction terms between sex and period and cohort slopes were fitted with sum-to-zero constraints for the period and cohort random effects:

\[
\lambda_{ij} = \alpha_{ij} + \beta_i + \gamma_j + \tau_c + \epsilon_{ij} + \epsilon_{ij} + \epsilon_{ij}
\]

Model C was fitted with and without the cohort slope and interaction term \( \tau_{icj} \). Unless otherwise specified, the results below show the model where \( \tau_{icj} \neq 0 \) (B) and where \( \tau_{icj} = 0 \) (C).

Methods

Data

Queensland Cancer Registry data were used from 2002 to 2016, with counts of diagnosis aggregated into five-year age groups for males, females and person-years. The models were fitted to the first 20 years of data (1992 – 2012) and the predictive accuracy was assessed using data and projections for the subsequent 15 years (2002 – 2016).

Results

Raw data

- Different trends were observed in the raw incidence counts for lung cancer between sexes.
- The slope versus birth cohort decreased with age group. The slopes decreased more rapidly among males than females.
- These plots suggested that sex-specific effects were required for lung cancer.
- Raw, sex-specific count data for lung cancer by age group or period

Model A (F / M)

Model B

Model C (PH)

Age-standardised incidence rates

- Observed age-standardised incidence rates (per 100,000) have sex-specific trends.
- Male lung cancer rates were decreasing.
- Female lung cancer rates were increasing.
- Models A, B and C (PH) appear to fit the data well.
- However, (A) and (B) do not allow comparisons between effects for males and females.
- Model B is exhibited:
  - Systematic error
  - “Near” estimates, suggesting over-fitting
  - A step increase in the width of the credible bands where predictions commence
- An observed in the raw data, sex-specific effects, additional to the intercept, are required.
- Model C (PH) did not have systematic error and poor diagnostics indicating the cohort slope and interaction were also necessary.

Predictions commenced:

Diagnostics

- Greater predictive errors (CRPS and ZTE) were observed for females (A/F) and B.
- Lower errors were for models A(B), A(C) and C(PH).
- Weak evidence of poor calibration was observed for Model C(PH). Other models appear to be well-calibrated.
- Positive relative implied predictions of Model C(PH) may be under-dispersed compared with data.
- Coverage probabilities were good for all models.

Relative risk

- Relative risks for males vs females can be calculated using Models B and C.
- At all ages, there have a much higher (95% CI 1.0-1.2) risk of lung cancer in males.
- The relative risk is constant with period and birth cohort.
- There is weak evidence of an age difference in the relative risk.

Conclusions

Bayesian age-period-cohort models were successfully fitted to cancer incidence data.
- Predictive errors were small, especially projecting 15 years.
- The predictions in the data was modeled well when \( \epsilon_{ij} \) was included.
- The most appropriate models were:
  - A(+) and A(+) the standard APC model, fitted to males and females separately.
  - CPH including interactions with sex and the slopes of both period and cohort.
  - Sex must be modeled separately or interaction terms must be used.
- Including sex in the model has practical advantages:
  - Relative risk of males and females can be informed.
  - Models of redhead incidence, for example, but vary wide 95% credible intervals unless the model included both sexes, modeled separately.

References