HealthStats NSW

Spatial adjustment

General information

WHY USE SPATIAL ADJUSTMENT?

Many of the indicators reported on HealthStats NSW, such as disease incidence and hospital procedures, occur in less than one in 1,000 people each year in NSW. These indicators, when reported at the Local Government Area (LGA) level, will often represent a small number of cases and will therefore be statistically unstable. Reporting of raw figures for LGAs with small populations is of limited use as differences between areas or between years is often caused by random variation, making trends and comparisons hard to interpret.

By way of analogy, consider a group of cricketers where we want to know how good each one is at hitting the cricket ball over the boundary fence. If we have a cricketer who has faced 1,000 balls and has an over-the-fence rate of five in 100 we could be fairly confident that their “true rate” is around five in 100. This means that if they played another 100 games at their current skill level they would continue to have an over-the-fence rate of around five in 100.

However, if we consider another cricketer who has only faced 20 balls then the same rate only represents one ball hit over the fence. If the second cricketer played 100 games at their current skill level we would be much less confident that their rate would remain around five in 100. They may be a 10 in 100 hitter who has been unlucky, or a one in 100 hitter who has been lucky.

We can be reasonably confident that the measured level of disease or healthcare usage is a reasonable estimate of the underlying rate in an LGA like Penrith as it has a population of around 190,500 people; Penrith is like our cricketer who has faced 1,000 balls. In contrast, a low population LGA such as Walcha in the Northern Tablelands which has a population of around 3,000 is like the second cricketer who has only faced 20 balls. We will be much less certain that a disease or procedure rate represents the underlying risk of the area rather than random variation.

HOW IS SPATIAL ADJUSTMENT ACHIEVED?

HealthStats NSW performs spatial adjustment by fitting a specific type of mathematical model, referred to as a Poisson generalised additive model, to the raw data at each combination of age group and LGA.

In order to produce more accurate estimates at the LGA level we borrow information from neighbouring LGAs. The level of adjustment is based on optimising the best possible fit of the model to the raw data, while minimising the effect of the random variation associated with small population LGAs.

Considering the second cricketer who has only faced 20 balls, we don’t want a model of the cricketer’s skill to change much based on their performance, because we don’t trust that we have a good estimate of their true skill level. Similarly, if Walcha LGA has a disease rate which is very different from its neighbours we think that is likely due to random variation rather than a true higher disease rate.

This results in neighbouring LGAs affecting each other. The model is weighted by population size so that LGAs with high populations affect their neighbours more than those with low populations.

ADJUSTED ESTIMATES AND CONFIDENCE INTERVALS

Adjusted estimates are produced for each age group within each LGA, then an overall rate for each LGA is calculated from a weighted sum of the age-specific rates with weights from the ABS standard population[1] to produce a directly age-standardised rate. More information on direct standardisation can be found in the age standardisation statistical method update on the HealthStats NSW website (www.healthstats.nsw.gov.au/IndicatorGroup/publications).

Confidence intervals are produced using a method described by Dr Simon Wood.[2] Many artificial datasets consistent with the model are produced and the random variation in these datasets are used to describe the random variation in the model. This process allows us to describe the variability of complex models such as this spatial adjustment model without making oversimplifying assumptions.

Technical information

CONSTRUCTION OF MODEL

A Poisson generalised additive model with spatial correlation described by a Markov random field and a relationship with age described by a cubic spline is applied to these data. In most cases, five-year age groups are used, however for indicators which have a lot of LGA-age group combinations with zero cases, 10-year age groups are used instead.

The model is fitted in R[3] using the mgcv library,[2] and the level of smoothness is controlled by optimisation using Wahba and Craven's unbiased risk estimator.[4]

HealthStats NSW produces estimates at the LGA level as two-year rolling estimates, so, for example for hospitalisation data which is reported by financial year, the first LGA map is the average of the 2001-02 financial year and the 2002-03 financial year. The example map at Figure 1 is the average of the 2002-03 and 2003-04 financial years.

CONSTRUCTION OF POINT ESTIMATES

The model is used to generate fitted data for each combination of LGA and age group; these are then combined as a weighted sum of the rates using the ABS standard population weights for age standardisation.[1] Thus the combined rate is an adjusted directly standardised rate. This rate is multiplied by the current population to produce the spatially adjusted case count.

CONSTRUCTION OF CONFIDENCE INTERVALS

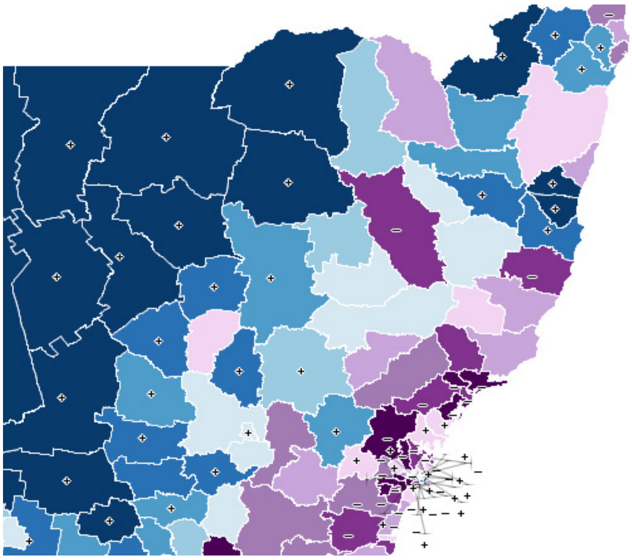
Using the simultaneous estimated uncertainty of the parameters produced by the model a large number of replicate parameter sets are simulated from the posterior distribution of the parameters. The replicate parameter sets are combined to produce estimates in the same way we constructed the point estimates.

We now have a large number of possible alternative fitted values which are in the correct scale (cases per 100,000) and are consistent with the data the model was constructed with. Therefore these data points represent a population of data points and the confidence intervals for our point estimates can be constructed from them. The lower and upper cut-offs of the 95% confidence interval are constructed as the 2.5th percentile and the 97.5th percentile respectively of the population of data points constructed above.

CALCULATING SIGNIFICANCE

An LGA indicator is considered significantly different from the state rate if the state rate lies outside the LGA's confidence interval. Significance is calculated at the 95% and 99% confidence levels. The spatial maps are annotated with plus and minus symbols (Figure 1) representing LGAs with rates which are significantly higher or lower respectively than the state average at the 95% confidence level.

Figure 1. Example map



REFERENCES

1. Australian Bureau of Statistics. Standard Population for Use in Age-Standardisation Table (Cat. no. 3101.0), 2013.
2. Wood S. Generalized Additive Models: An Introduction with R. Chapman & Hall; 2006.
3. RC Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria; 2015.
4. Craven P, Wabha G. Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. Numerische Mathematik 1978; 31: 377-404.