Statistical methods for estimating sources of variability in count biomarkers

Kostas Tryposkiadis[1,2], Alice Sitch[1,2], Malcolm Price[1,2], Jon Deeks[1,2]

[1]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK

Correspondence: Kostas Tryposkiadis (KXT859@student.bham.ac.uk)

Background

Analysis using random effects linear models is the established method used in biological variability studies to attribute the observed variability arising from between-patient differences, within-patient differences, and measurement error. However, these models assume underlying normality, and thus may not be applicable for biomarkers based on counts.

Aim

To present methods for estimating sources of variability in count-based biomarkers and apply and compare approaches in a case study of patients with Sjogren's syndrome.

Methods

Both Poisson and negative binomial models are appropriate for analysis of count data, and methods for obtaining between and within-patient variance estimates are described in Leckie et al[1]. We analysed the biomarker data using random effects Poisson and negative binomial models, and for comparison, using a random effects linear regression model. The intra-class-correlation (ICC) was calculated as a ratio of the between-patient variance over the total variance, and was compared across the different models. The AIC and BIC criteria were used to assess each model's performance.

Data from 32 patients with Sjogren's syndrome was used as a case study, considering the focus score, calculated for each salivary gland observed in each biopsy as the number of foci over the glandular area, multiplied by 4. Between-patient and within-patient-between-gland sources of variability were estimated.

Results

The ICC estimates obtained from Poisson (0.323) and negative binomial models (0.310) were similar, and higher than the linear regression model (0.222). AIC and BIC values were similar for Poisson (AIC=463.63, BIC=469.84) and negative binomial models (AIC=465.55, BIC=474.87) and indicated both were a better fit than the linear regression model (AIC=632.69, BIC=642.01).

Conclusion

It is important to properly model the distribution of biomarkers based on count data to correctly estimate sources of variability and measurement error.

Keywords

Biomarkers, variability, count data

References

[1]Leckie et al, 2019. arXiv: 1911.06888 [stat ME].