

Bayesian partition models for exposure data

Øivind Skare¹, Jon Wakefield² and Paolo Boffetta³

¹ Division of Epidemiology, Norwegian Institute of Public Health, Norway

² Department of Biostatistics, University of Washington, USA

³ Lifestyle, Environment and Cancer Group, International Agency for Research on Cancer, France

It is now becoming increasingly common to measure multiple “exposures”, be they occupational, dietary, environmental, or genetic, in epidemiological studies investigating disease-exposure associations. The large number of regression coefficients that need to be simultaneously estimated has led to the need for new analysis methods, and a number of hierarchical models have been described that assume that the collection of coefficients arise from a distribution, or set of distributions indexed by covariates. Here we propose a new model that assumes that the covariates that define the groups may be surrogates for an underlying true exposure; in this way the partitioning is given more flexibility, and outlying effects within groups can be placed in more appropriate groups. As an alternative we consider a model that automatically partitions the coefficient space, to provide clues to aetiology. The approach is demonstrated first with simulated data, and then with a case-control study investigating the association between lung cancer and 75 exposures.