We describe the Bayesian methods employed for the accurate analysis of data from genomic microarrays designed for *Streptococcus pneumoniae* molecular serotyping.

Streptococcus pneumoniae is a major cause of mortality worldwide. Infected individuals commonly carry more than one *Streptococcus pneumoniae* serotype. Monitoring multiple carriage is essential for surveillance of pre- and post-vaccination populations, enabling the epidemiology of disease association, vaccine introduction and serotype replacement to be investigated. A novel genomic microarray capable of measuring multiple serotype carriage in clinical samples has been developed. To date, around 3,000 samples have been analyzed from numerous studies worldwide, and the method has demonstrated an enhanced ability to detect multiple serotype carriage and determine the relative abundance of serotypes present. The method has also excelled in independent tests of currently available multiple carriage monitoring methods. Additional features on the array enable it to assess genetic relatedness of samples, monitor antibiotic resistance and detect co-colonizing pathogens.

Here we describe the Bayesian solutions we have developed for the data analysis problems presented by these genomic arrays. The technique presents a number of analysis challenges. Noise is a problem, particularly at low relative serotype abundance, as is cross-hybridisation. 91 serotypes of *Streptococcus pneumoniae* have to be distinguished using probes for 432 capsular genes, each serotype containing a small subset of these 432 genes. However many of the serotypes have similar combinations of capsular genes, making it difficult to distinguish the identity of a serotype in a sample, particularly in the case of multiple carriage. In addition some pairs of serotypes have identical complements of capsular genes necessitating extra probes on the microarray which must be incorporated into the analysis. And in cases of multiple carriage, a measure of the relative abundance of the different serotypes in the sample was required. The Bayesian approach we adopted enabled the development of a flexible and expandable statistical model, which produces a robust and highly accurate analysis of the data.