Hey everyone, my name is Eric Ellwanger – a fellow classmate in Stat 578 this semester. I’m going to be presenting on the paper that I was assigned.

The paper that I read was “Towards measles elimination in Italy: Monitoring herd immunity by Bayesian mixture modeling of serological data”. The paper was published in the publication Epidemics – Volume 4, Issue 3 in August 2012. The paper can be found starting on page 124 through page 131. The authors were Emanuele Del Fava, Ziv Sckedy, Angela Bechini, Paolo Bonanni, and Piero Manfredi.

In 2003 Italy embarked on a National Measles Elimination Plan to get in-line with vaccination levels proposed by the World Health Organization. This involved increasing the percentage of the children getting a first dose vaccination, starting a campaign to provide a booster (or second) dose, and implementing the National Measles Elimination Plan targeting children under 15 years old. In order to assess the effectiveness of the program, data was collected before the plan started, and than again 2-3 years after the program started. The level of antibodies to measles present in the subjects blood measured in log10 mIU/ml, as well as the subjects age were collected.

Here is a quick snapshot of the data for both periods. All data comes from the Tuscany region, in 2003 data is gathered from Florence, Sienna, and Lucca, where in 2006 all data came from Florence. Vaccination rates from all of the areas are very similar, so are considered exchangeable for this paper. Again, this shows what data was collected. It should be noted that prior infection or vaccination of the subjects was not collected, so there is no way of knowing if antibodies present are due to vaccination or natural infection.

The authors wanted to accurately assess the impact of the vaccination program using the post-vaccination data. Typical analysis uses classical cut-off methods suggested by the assay manufacturers to classify a subject as susceptible or immune. These assay manufacturers target test specificity which measures the True Negative rate (a measure of susceptibility) at the cost of test sensitivity. This is thought to classify a fraction of the immune population as susceptible. Using a Bayesian normal-mixture model allows for more than 2 classifications of immunity/susceptibility. To properly model this there are 2 assumptions that the authors make: 1) That immunity is lifelong and 2) The mortality rate from vaccination is negligible – both seem reasonable assumptions.

The authors proposed two different normal mixture models – an age-independent model and an age-dependent model. I will be discussing the age dependent model. As you look at this model you can see that the data is modeled after the sum of j normal components (or submodels). The pi\_j’s are the mixture probabilities that a subject aged a\_i belongs to the category (or subpopulation) j. The Y\_i’s are the antibody counts (or the data collected) and is assumed to be normally distributed in each component with each component having it’s own mean mu and variance sigma squared. Just to note that this model suggests that the means and variances are independent of age.

Here is a mathematical description of this model. Here we replace the mixture probability with a binary latent classification random variable T\_j(a\_i), which represents which subpopulation that a subject aged a\_i belongs to. The T\_i(a\_i’s) are have a categorical distribution based on the mixture weights that a subject age a\_i belongs to the jth subpopulation. Because a subject can only belong to 1 subpopulation, the sum of the T\_j(a\_i)’s must be equal to 1. The other data, the Y\_i’s, or the antibody levels in a subjects blood are assumed to be distributed normally within each jth subpopulation with it’s own mean mu and variance sigma squared. Where I is inclusive of all of the samples 1 through n. To minimize the effects of the priors, the authors chose non-informative, flat priors. The mean mu is chosen to have a uniform prior in the range of the minimum age to the maximum age. The one expectation is that the means are in order from low to high. This is to insure that low antibody levels are in the first 1st group and as the antibody levels increase they are in successive jth groups. The variance sigma squared is chosen to have a relatively flat prior from the inverse-gamma distribution with alpha and beta = .01. Lastly the mixture probabilities pi(a) are chosen to have a flat prior with a deereeyclay distribution with alphas equalt to 1. The deereeclay distribution is can be thought of as a multivariate version of the beta distribution.

This Directed Acyclic Graph shows this model where the N plate represents all of the data 1 to n, and the K plate represents the different categories or subpopulations of the chosen model.

This is the JAGS model that the authors used for the age dependent model. As you can see it shows the data Y\_i’s and the latent classification variable T\_i’s on the N plate. The mean mu’s and each T\_j(a\_i) are on the K plate. And while not explicitly shown to have K components there are actually K sigma squares which happen to be all equal.

Here are the details on the MCMC details that I could gather. The authors used JAGS software version 3.1. They modeled each of the model parameters to find it’s posterior mean and 95% credible interval, which were used to help select the number of components for the model. The authors ran several different age dependent mixture models with the number of components from 3-6 (they chose not to include 2 components due to the fact that it didn’t fit the data very well). The used a form of deviance measurement (Penalized expected deviance) to analyze the different models. PED is similar to DIC because of the use of deviance, but the penalty is usually around 2 times higher than DIC. Using this measure, the authors found that 3 components fit the data best in the 2003 period and 4 components for the 2005-2006 period. I did reach out to Dr. Del Fava to try to get the rjags code to determine number of chains, what if any convergence tests were done, and if any model checks such as normality and outliers was done. After several emails back and forth, Dr Del Fava chose not to share his code with me.

The process that the authors used after determining the number of components to use for the models, was to run the models to get the approximate means of the different mixtures to help classify the different components. They used these means as well as previous knowledge (such as the classical cutoffs) to classify the different components as susceptible or immune. They classified the lower antibody group as susceptible and the higher antibody groups as immune. There were cases in both periods that fell between the two classical cutoffs that thy labeled as weakly immune. Once they had the different component groups labeled, they assigned each subject to a particular age-dependent group. Having all of that information they were able to make inferences about the age-dependent probabilities about being susceptible to the measles.

What were the results? In both periods, the researcher found that the highest antibody levels were found in children ages under 5 and people older than 25. There was a generally lower antibody count in subjects ages 10-20 years. As you probably expect there was a noticeable lag (approximately 5 years) between the 2 periods (which happened to be about 3 years apart).

This graph shows that fairly well. Where the black line represents the data from the 2003 samples and the red line represents the samples from the 2005-2006 period.

As mentioned earlier the authors found 3 distinct clusters in the 2003 data and 4 distinct clusters in the 2005-2006 data.

You can see here in both periods that group 2 falls within these 2 classical cutoff values. In non-Bayesian analysis these subjects would fall in the tails of either the susceptible or immune categories. In this Bayesian analysis because this is a distinct group, the authors are able to distinguish and label this group as ‘low repsonders’

Another result the authors found was, as expected (because of the test specificity), the number of susceptible subjects is greater in the classical cutoff methods. But surprisingly that did not hold in the 2005-2006 period – possibly due to vaccination wearing off over time??

You can see that in this graph, where the graph shows the proportion of subjects that are susceptible (seronegative) by age with the black dots representing the convential cutoff method and the red dots representing the Bayesian normal-mixture model.

This last graph summarizes the authors findings. They show the ‘high responders’ (group 3 in 2003, and groups 3 and 4 in 2005-2006) the solid curves, the ‘susceptible’ group 1 in both periods (the dotted curves), and the ‘low responders’ group 2 in both periods (the dashed lines) that fell in between the 2 other groups. The graph clearly shows that the high responders show mostly children 2-5 years and older than 20 (whose immunity is likely due to having actually had the measles). In 2006 the height is around 1.5 years of age likely due to the focus of the immunization campaign to vaccinate at younger ages. Both periods show high responders in the younger age groups which is consistent with targeting of the younger age groups for immunizations. What the authors found surprising was the lower prevalence of high repsonders in the 10-20 year age group (10-20 in 2003 and 15-25 in 2005-2006). The authors suggest that this age group is a good target for vaccinations/2nd dose programs.

I am certainly not an expert in Bayesian modelling and I know very little about serological data, so I don’t feel comfortable saying things are erooneous or have issues. I will highlight some of my observations based on the things that we have learned this semester. First I’ll point out that I did not see any mention of model diagnostics (other than model comparison) such as convergence testing, checks for normality within the different components, or checks for outliers. The authors do point out that the inferences could be better if they had additional information, in particular if they knew whether or not each subject had had measles before or had the initial vaccination and 2nd dose vaccination. If this information were available, better inferences could be made (in particular for the low responders). While the authors make a point that the 2005-2006 data coming from just Florence shouldn’t matter due to many things, this is a possible issue that could have skewed the data.