**HBV Case study**

* Overall prevalence rate is: 0.4% (Source: Epidemiology of Hepatitis B in the United States)
* Age distribution by case is given by (Year 2017, United States): <https://gis.cdc.gov/grasp/nchhstpatlas/tables.html>

|  |  |  |  |
| --- | --- | --- | --- |
| Age | Cases | Population | Calibrated risk |
| 0-14 | 0 | 60,628,386 | 0.00% |
| 15-24 | 99 | 42,934,115 | 0.09% |
| 25-34 | 572 | 44,955,410 | 0.48% |
| 35-44 | 1,138 | 40,578,221 | 1.07% |
| 45-54 | 896 | 42,088,638 | 0.81% |
| 55-64 | 457 | 41,696,099 | 0.42% |
| 65+ | 242 | 50,505,383 | 0.18% |

* Calibrate the CDC data by multiplying with a constant for it to match the overall prevalence rate of the US, i.e., 0.4%. Assumption: Underreporting is uniform across age groups.
* The calibrated risk is shown in the table and is plotted below (highly heterogeneous).

* The paper by “Comparative analysis of triplex nucleic acid test assays in United States blood donors” gives data on fully infected pools for HBV (shown below):

|  |  |  |
| --- | --- | --- |
| n | k | Se\_tilde |
| 1 | 1 | 0.94 |
| 6 | 6 | 0.90 |
| 8 | 8 | 0.83 |
| 16 | 16 | 0.75 |

* Assuming a specificity of 0.95 (a source for Sp needs to be found) using the linear regression method.
* (This examples is for alpha=0.031117374, gamma=1.327264196) For the calibrated alpha and gamma parameters we can have a plot (x-axis=pmin and y-axis=pmax) that displays that region the will satisfy the condition. The plot is shown below:

Notice that all relevant points are above the y=x line (as pmax>=pmin). The green region is when it is satisfied while the red is when it is not satisfied. The figure reveals that the condition is satisfied for the vast majority of cases .The only cases is does not satisfies is when we have a huge disparity between pmin and pmax, which is unlikely to occur in most realistic settings. In our case study, we have that p\_min=0% and p\_max=1.07%, which is part of the green region

* A new calibration based on linear regression (see screenshot on linear regression). Doing see gives: alpha=0.053460023 and gamma=1.2467. The plot for the linear regression is (almost a perfect fit):