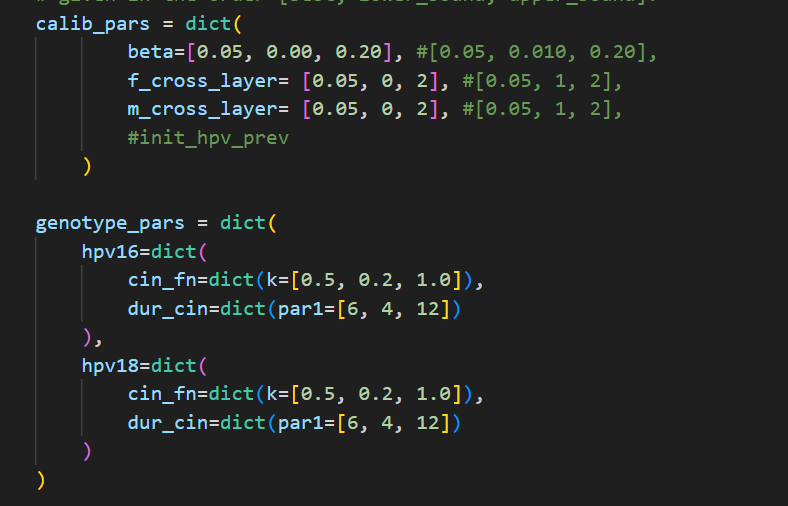
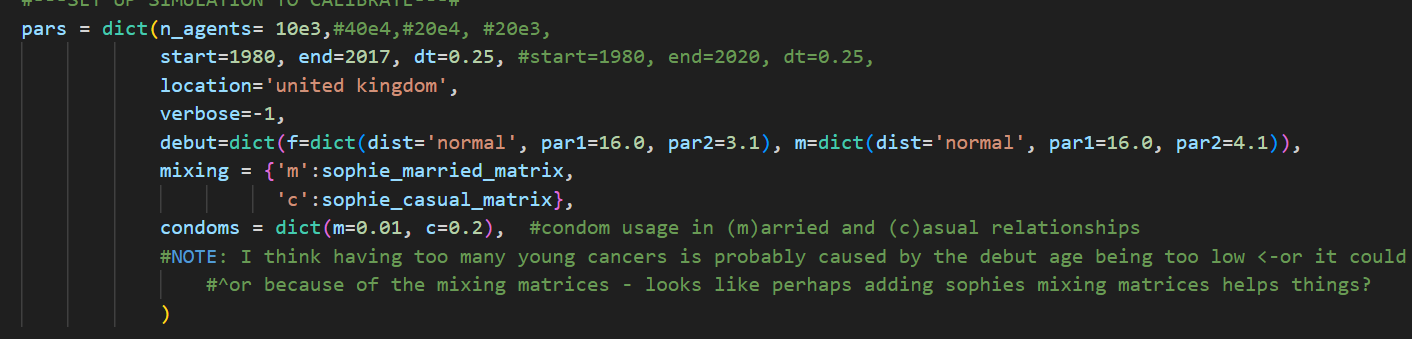
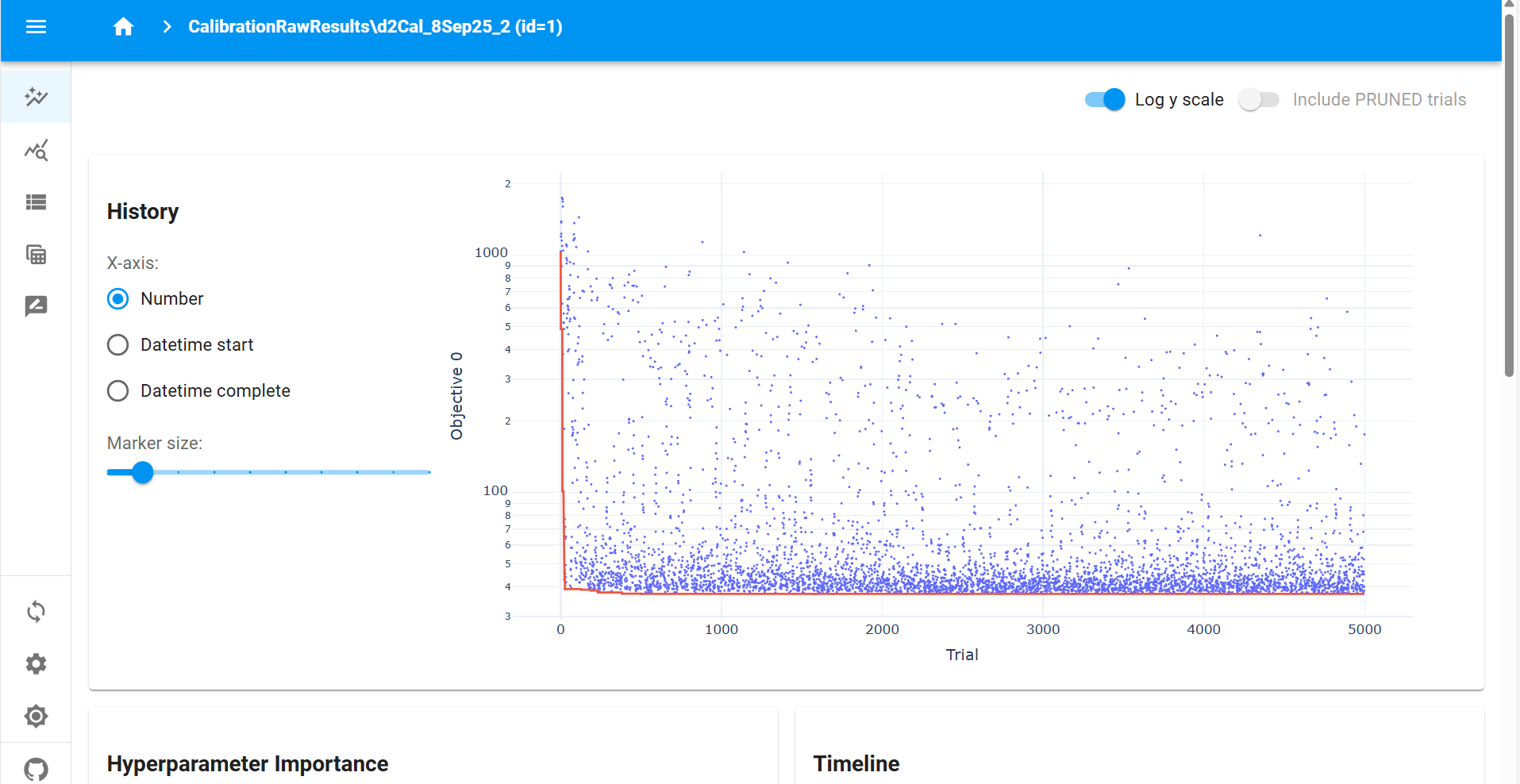
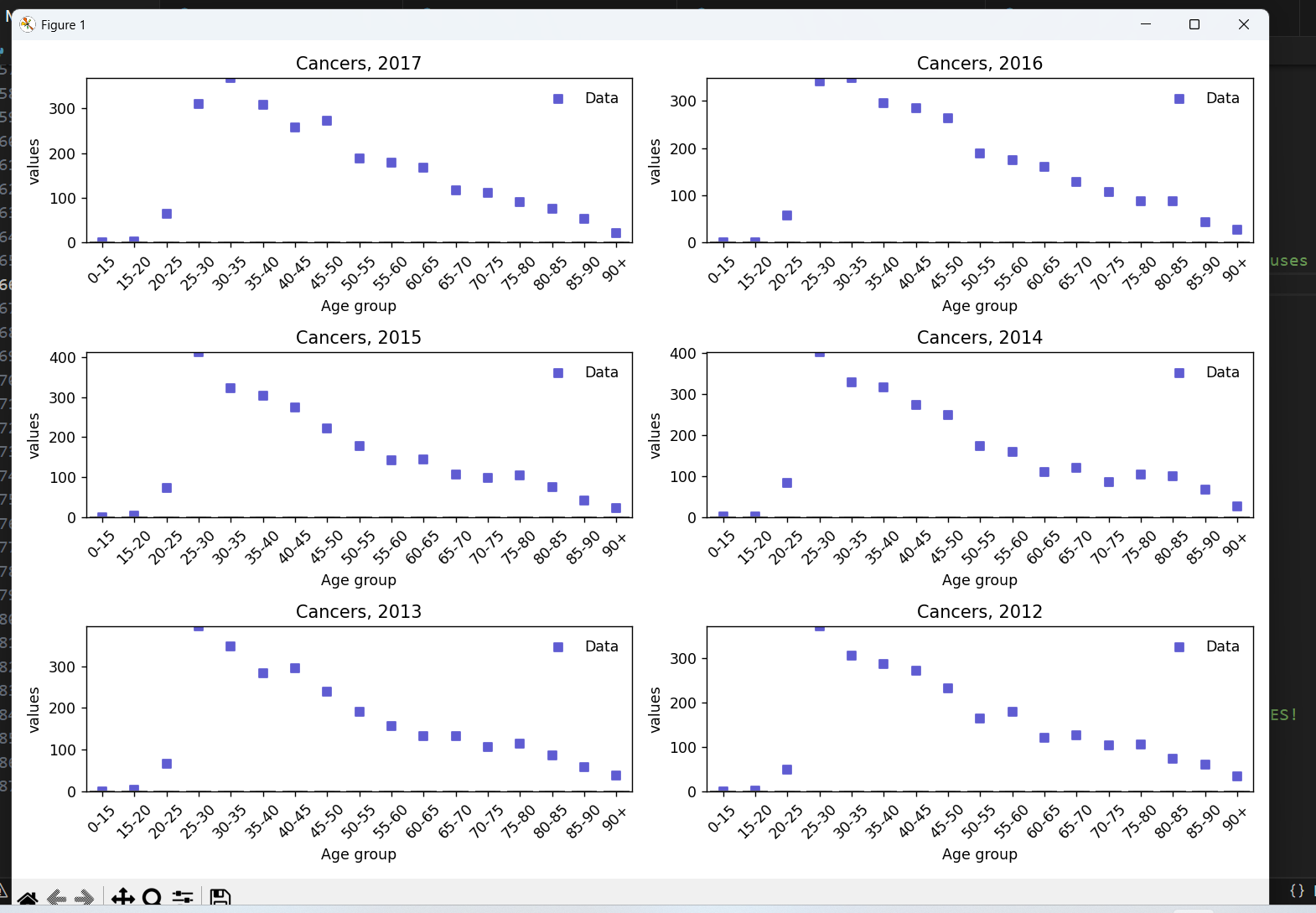
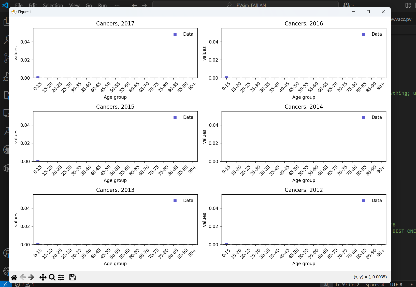
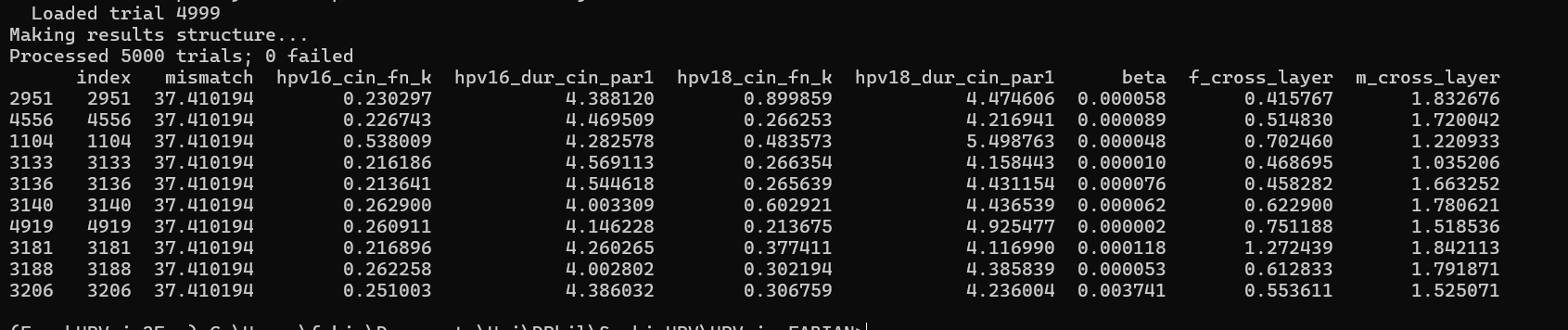
# Sophie Task 5 tracking –getting SOME good UK cal

**d2Cal\_8Sep25\_2**

* I want to try a calibration with the updated mixing matrices, which should allow for cancers to develop at a lower age. This should mean that I better reflect the trends in real-world data than before, where the model got a good fit for the later ages but massively underestimated how many cancers were for the lower ages
* I am not adding any of the NHS interventions right now, as it shouldn’t make a massive difference to us here (given data is 2012-2017) – of course, once I have gotten a decent calibration with no added interventions, I will model full NHS interventions to get the final cal I really want
* Only calibrating to dataset D2, so no genotype distribution: I do want to update init\_hpv\_dist to be a reasonable value (if I cant calibrate it, as calibrating it seems hard perhaps), and deffo calibrate to at least one year worth of genotype distributions
* 
* 
* It took c. 10mins to load the 5 000 trials up on Optuna Dashboard
  + 
* This calibration was utter crap. For the best 50 calibrations all the way down to the best single calibration, it is predicting absolutely 0 cancers for all age groups for all times, which of course doesn’t fit the data whatsoever
  + 
* It looks to me that the beta values for some reason are all absolutely tiny, and assuming that Optuna has done a thorough search of the parameter space (which, over 5000 trials it should have done, and it does look like it has done as we see decent variation in some other parameters – not the dur\_cin’s but i think they are all sticking rather close to 4 again to push the #cancers to 0, so that makes sense perhaps– so I think I can assume this), this means that the closest our model can get to our data is 0 cancers for each age bracket for each year
  + 
  + I think this is because there are **too few agents**. With a UK population of around 60 000 000, each of my 10 000 agents represents c.6 000 people (with cancerous agents representing 600). I suppose with the absolute values I am looking to fit to being around 10-350ish with a mean of around 150, it means the models with the best fit are just those which always predict 0 cancers.
    - If true, this further means that for dataset D2, a mismatch of 37.410194 means all model predictions are at 0 – so any lower mismatch should mean the model is doing something somewhat meaningful when fitting to the data!

**d2Cal\_8Sep25\_5**

* I am redoing the same as **d2Cal\_8Sep25\_2** with the only difference being I am using 100e3 agents rather than 10e3 agents (i.e. 100 000 rather than 10 000). If my reasoning for why **d2Cal\_8Sep25\_2** didn’twork well is correct, then this should mean I have enough granularity with cancerous agents representing around 60 people to at least get somewhat of a good fit to the data, very crudely.
  + If this is true, it is then time to either try a calibration with even more agents (maybe x4 so each cancerous agent represents around 15 people), or by grouping some of the age brackets for the data (perhaps into 10-year buckets), or both, to try and get a model which fits the data really nicely.