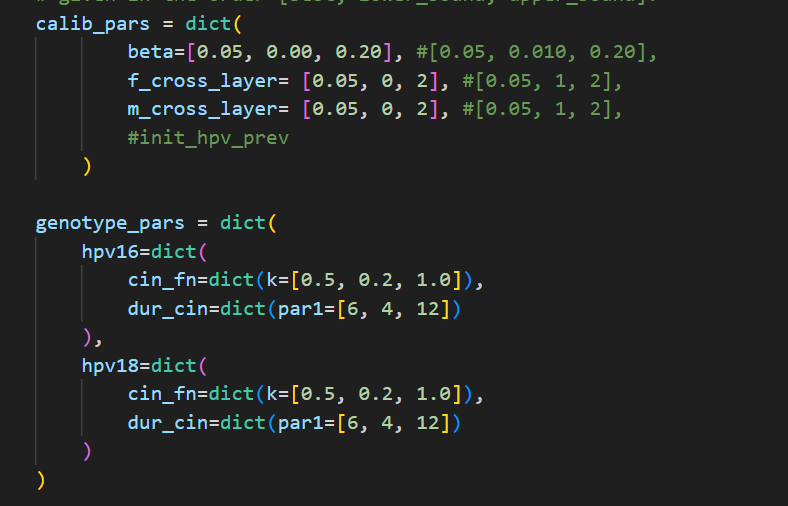
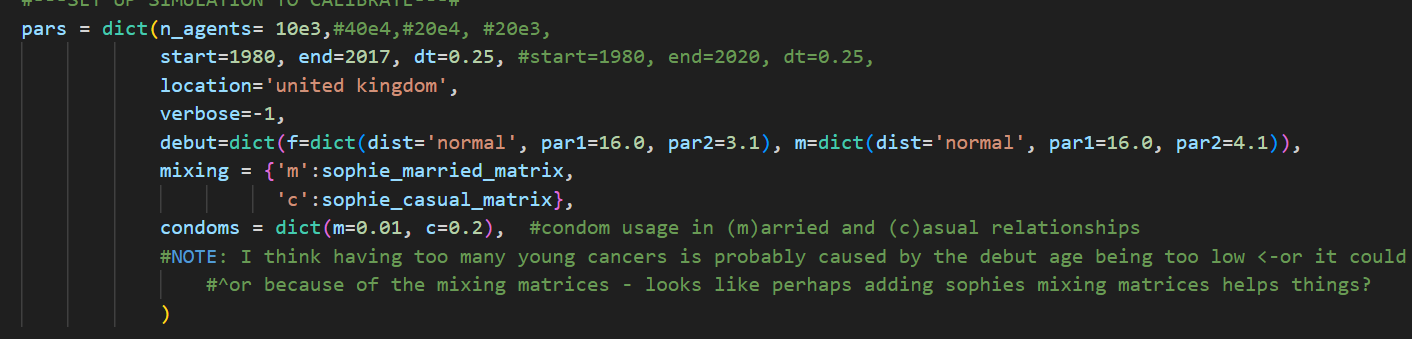
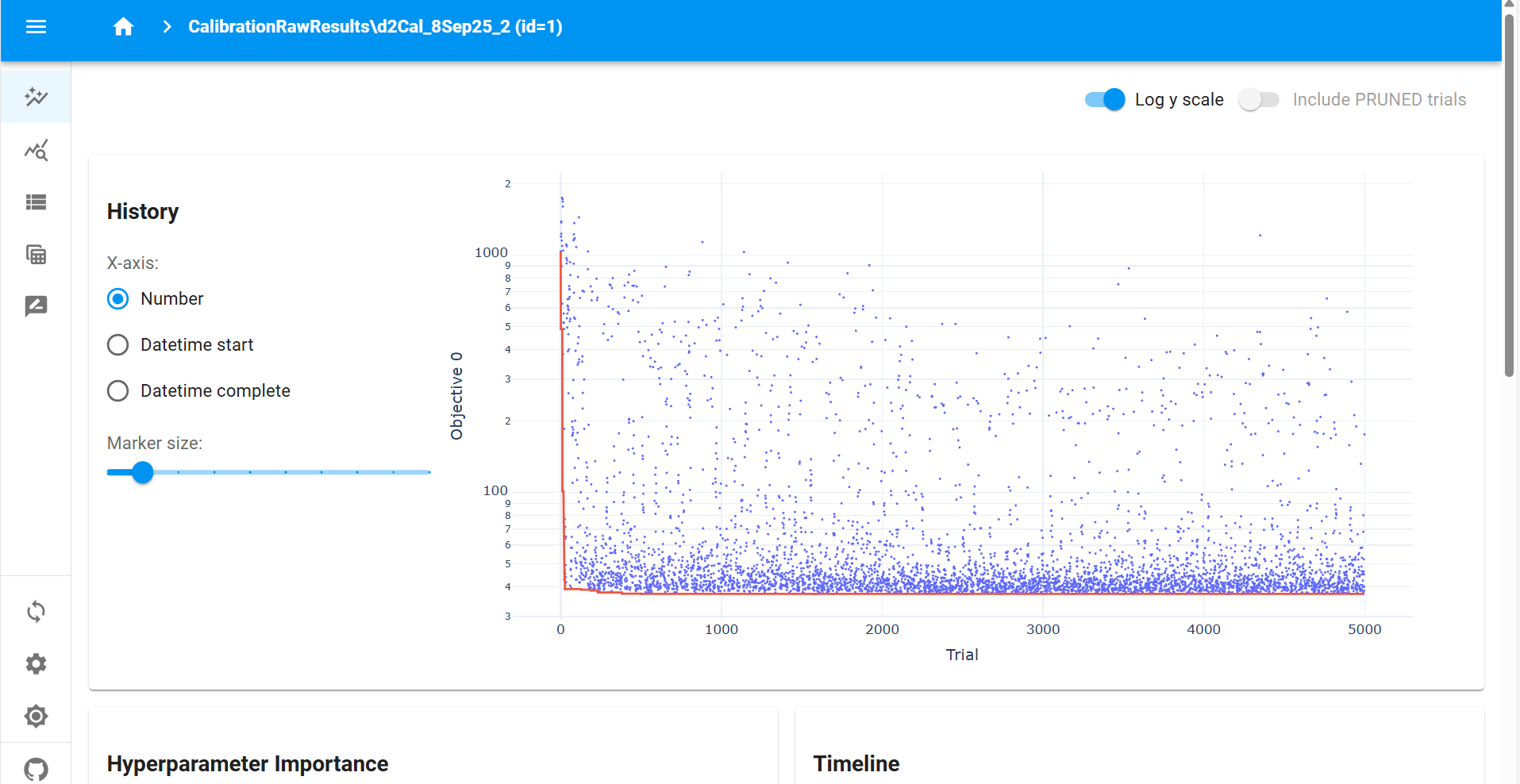
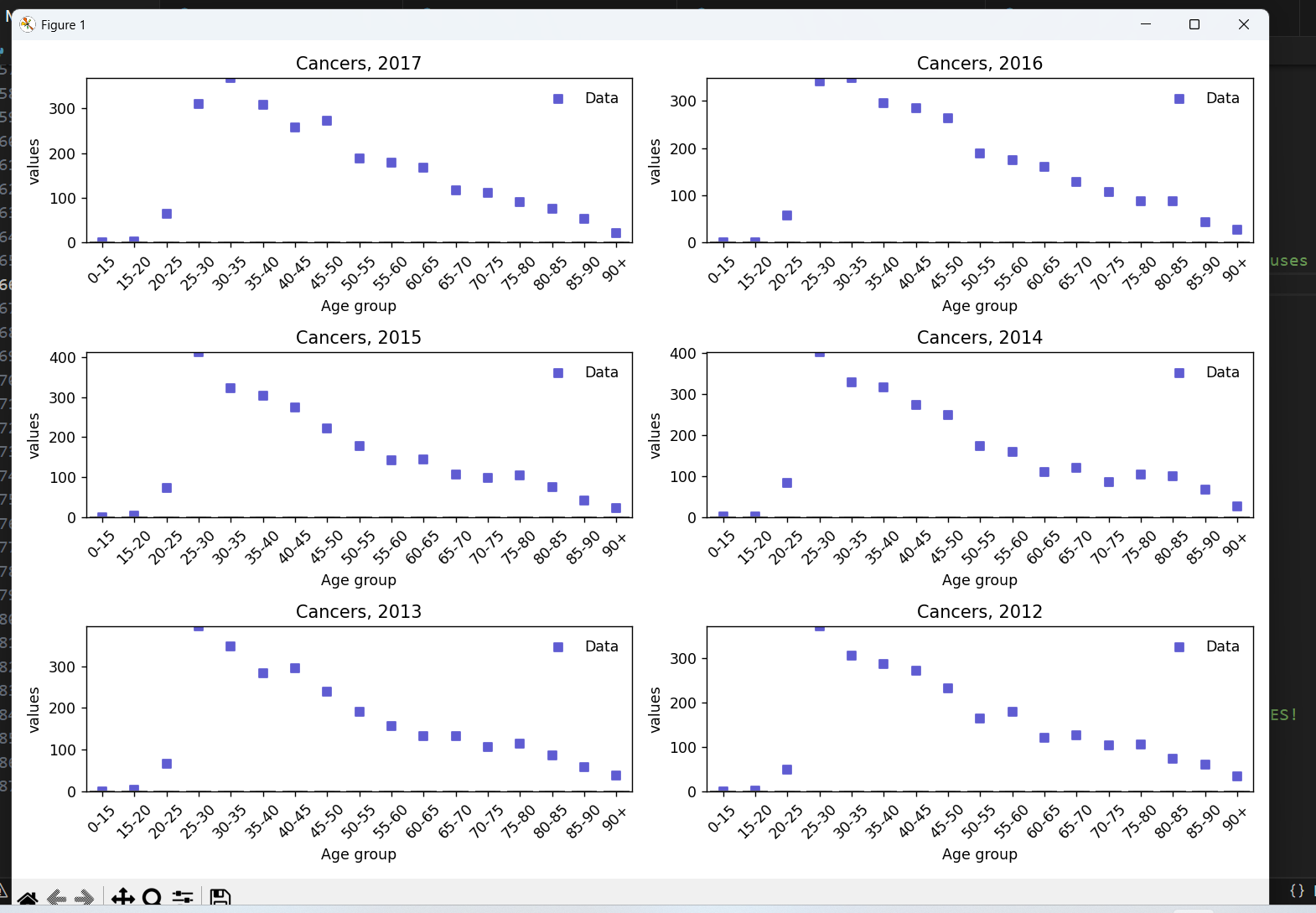
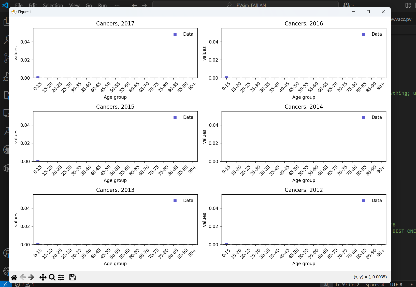
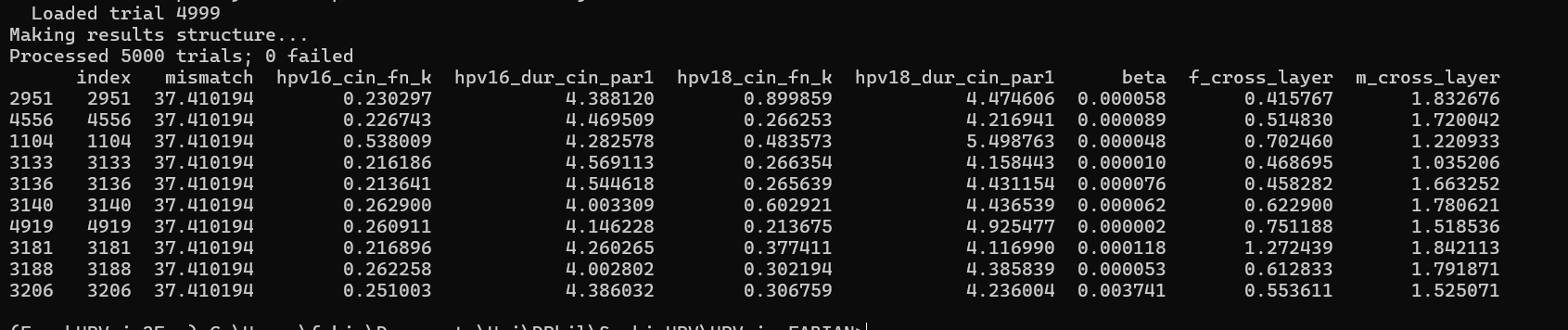
# Sophie Task 5 tracking –getting SOME good UK cal

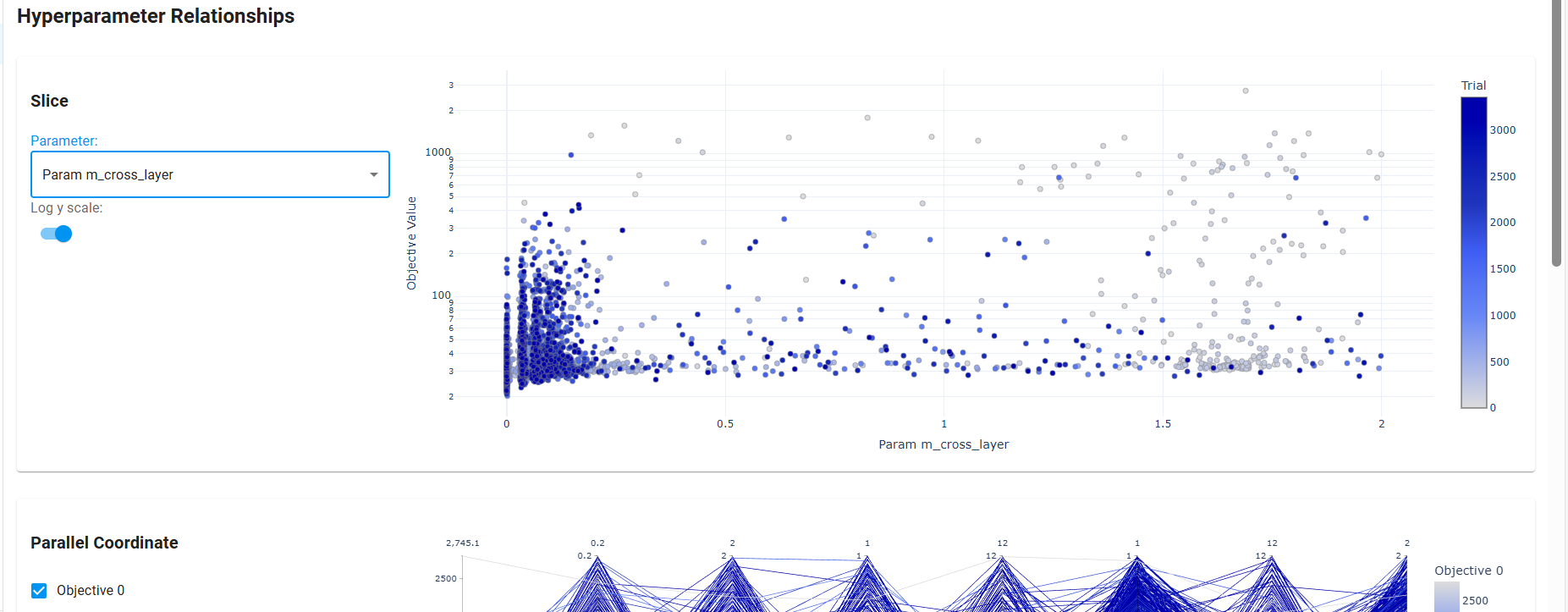
**d2Cal\_8Sep25\_2**

* 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!

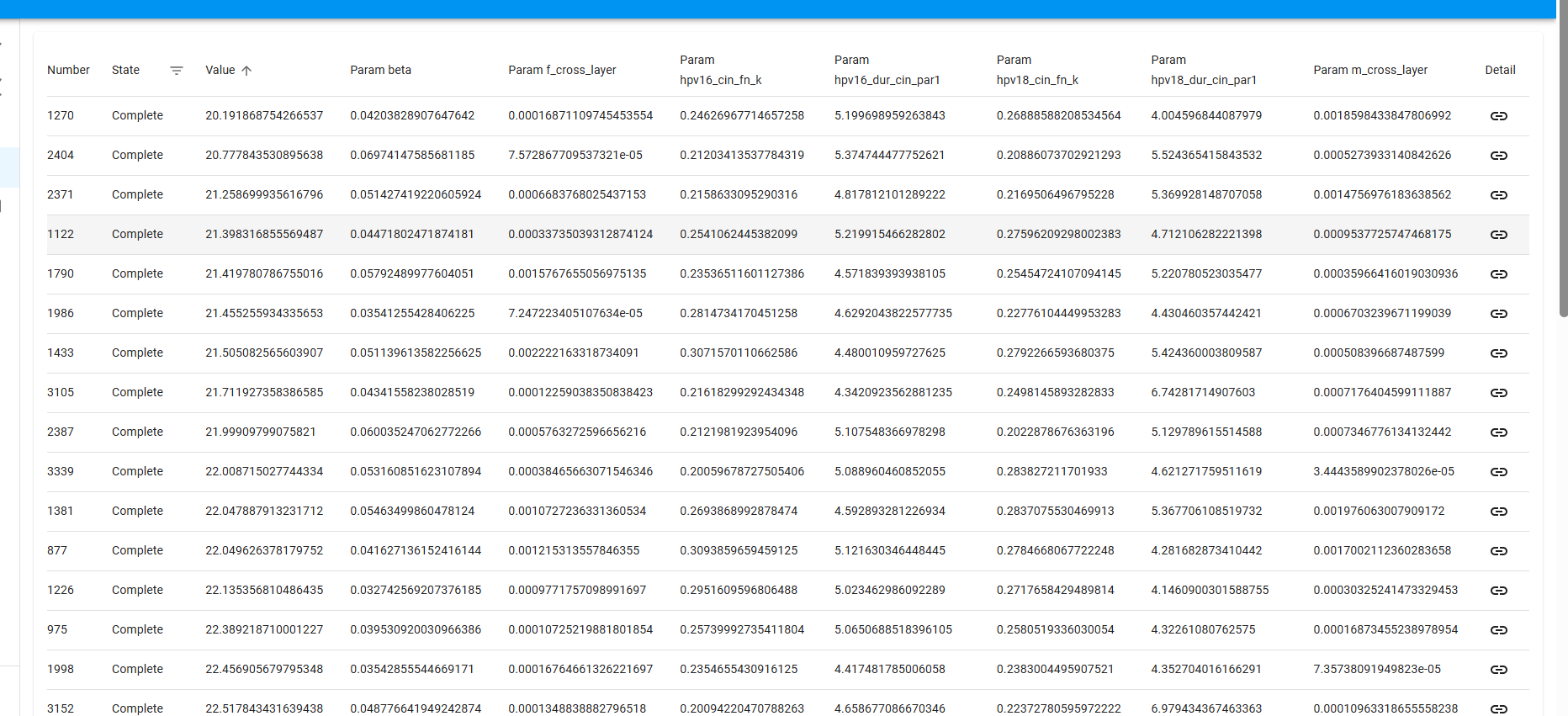
**d2Cal\_8Sep25\_5, d2Cal\_8Sep25\_6**

* 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.

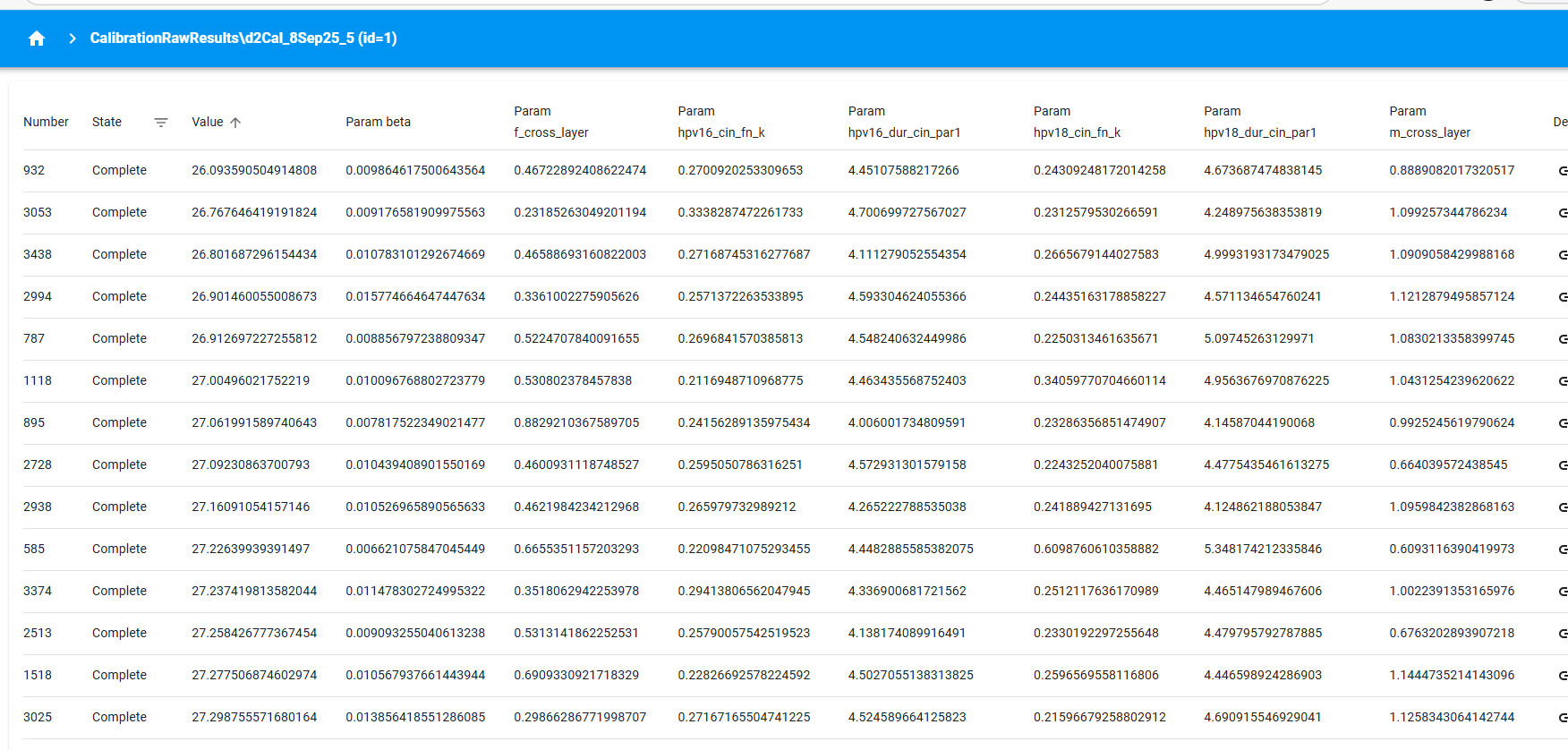
….. i think that maybe as trials continue, more data is stored up and maybe my computer ran out of memory and that is why it crashed at 4000 trials.. if so it is no matter as i can still load up the study in Optuna dashboard and get the best parameter values on there and rerun with these values and see the tightness of the fit. And either way, i should still use that calibration therefore! Even if the rerun somehow works because i have done die=False rather than die=True



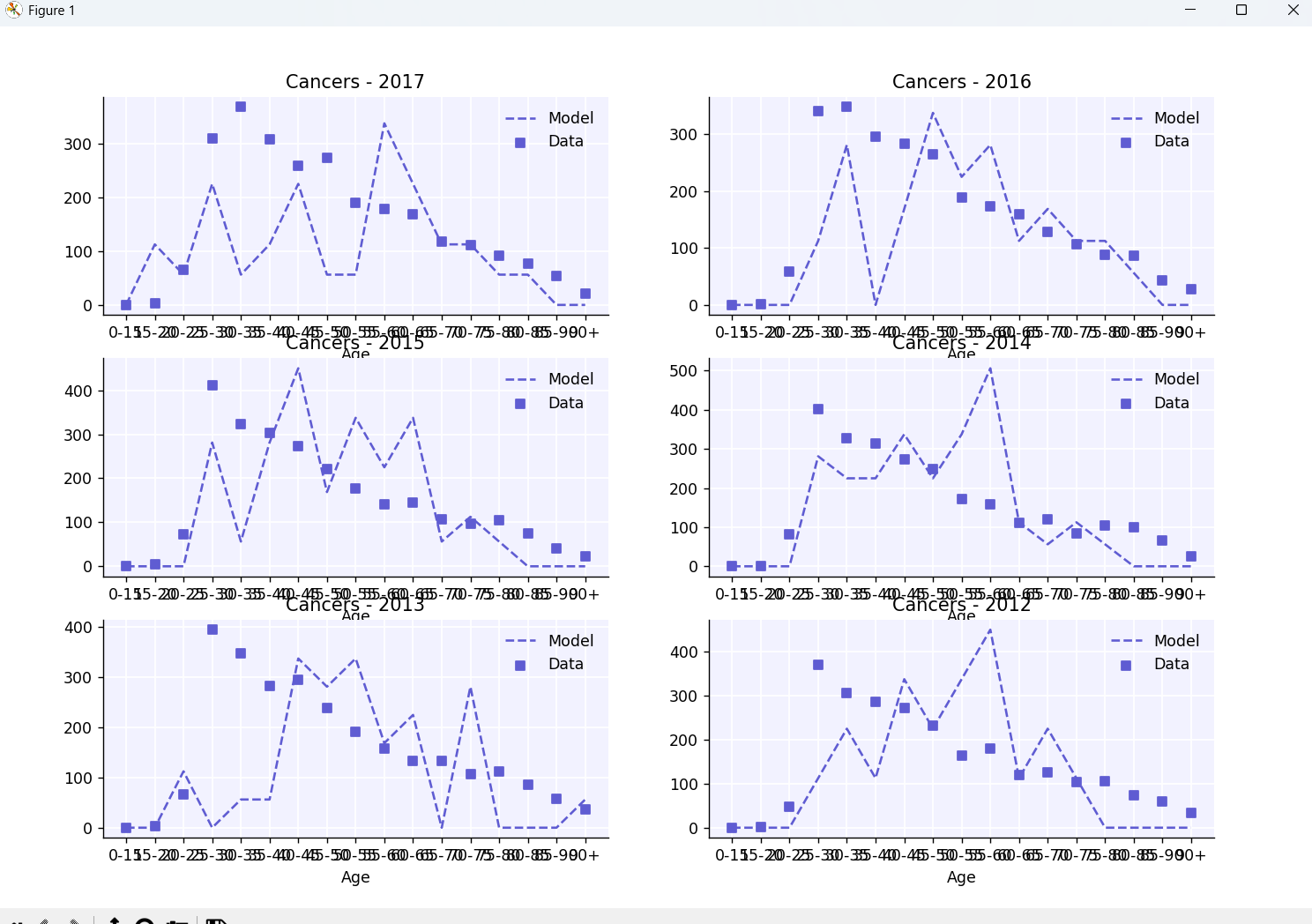
^its worth looking at this i think to see if i need to make ranges of stuff bigger , i think in cases like this where stuff is bunched to the side, probably yes!.



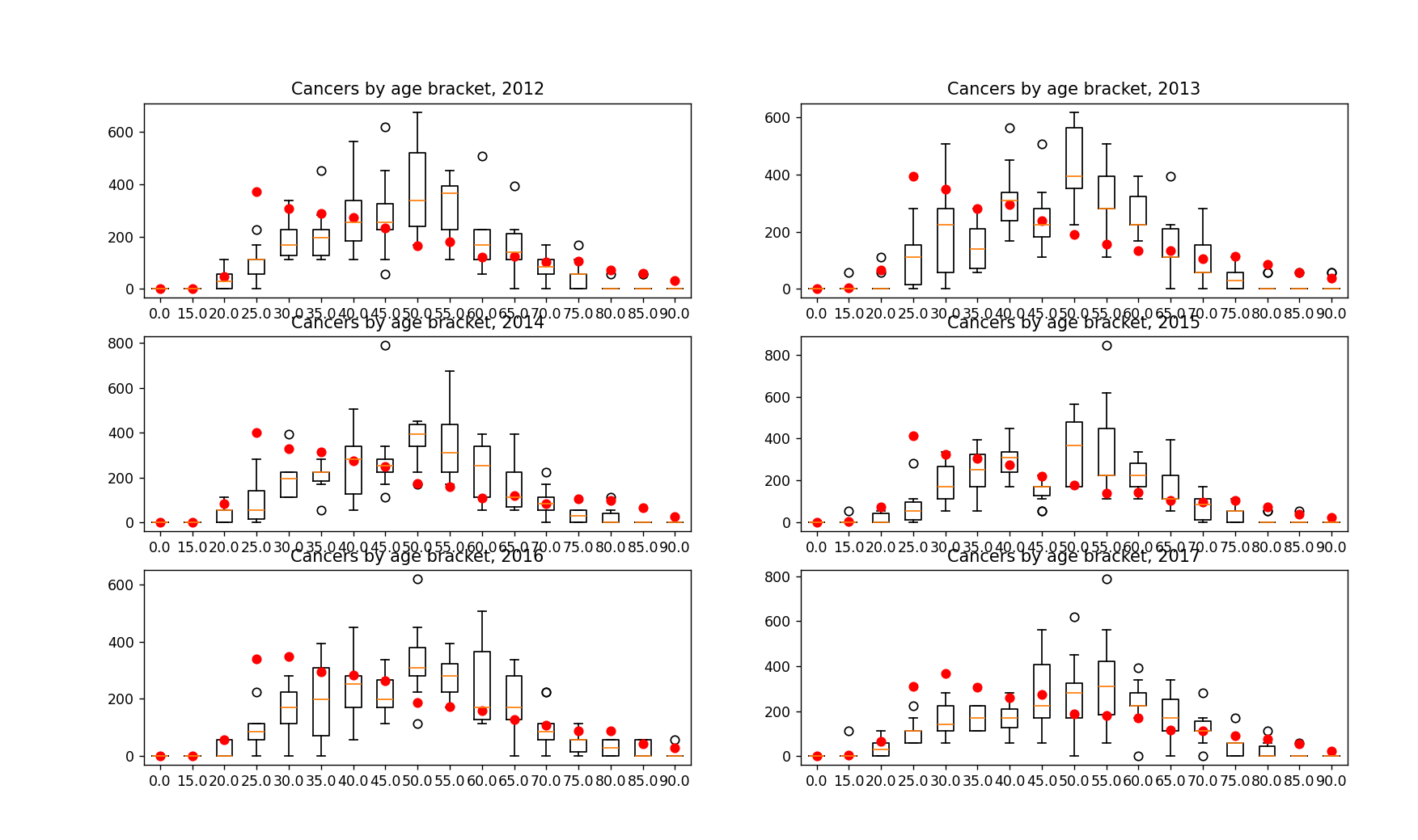
^ for number 6



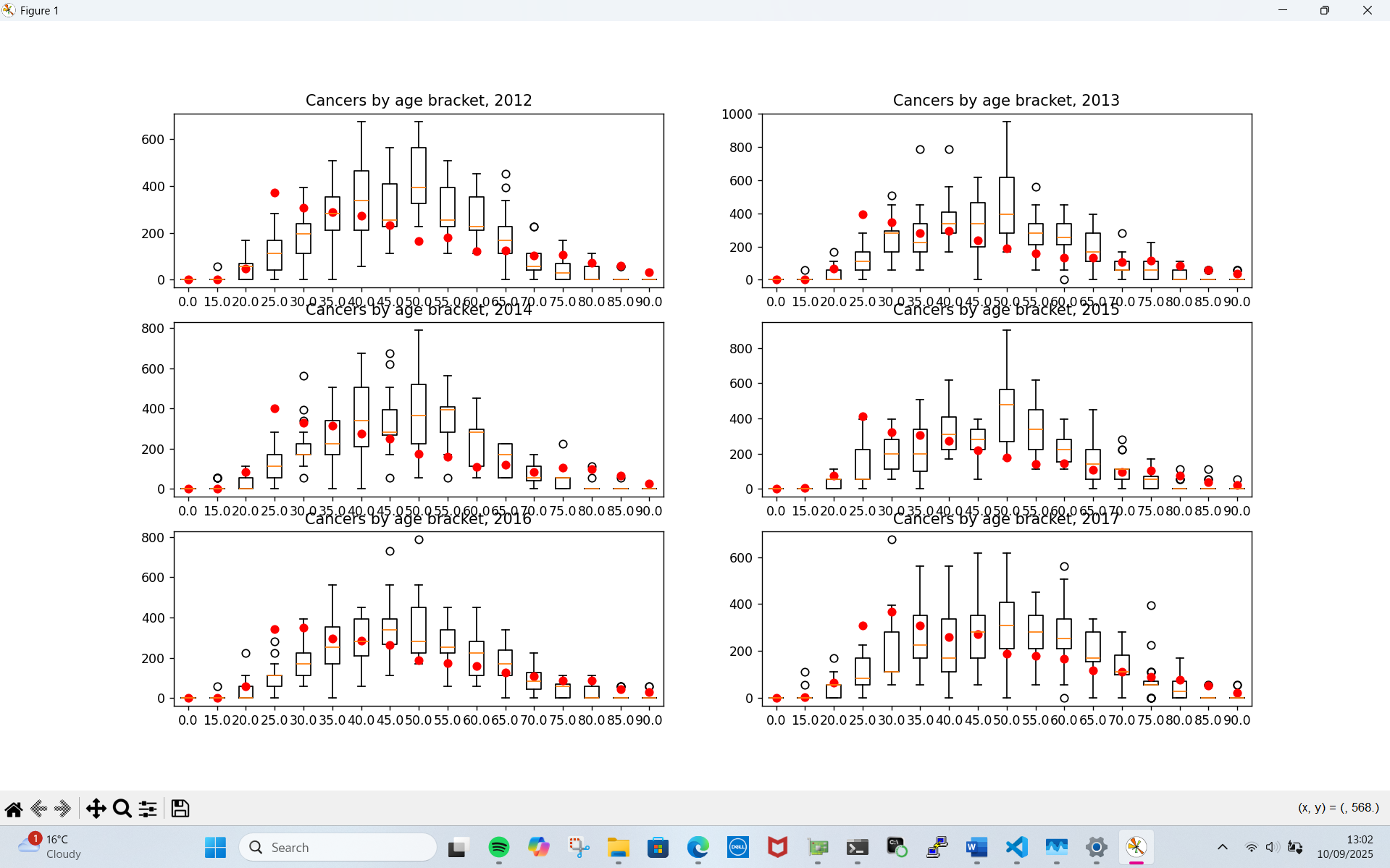
^for number 5



^ one run from best of number 6



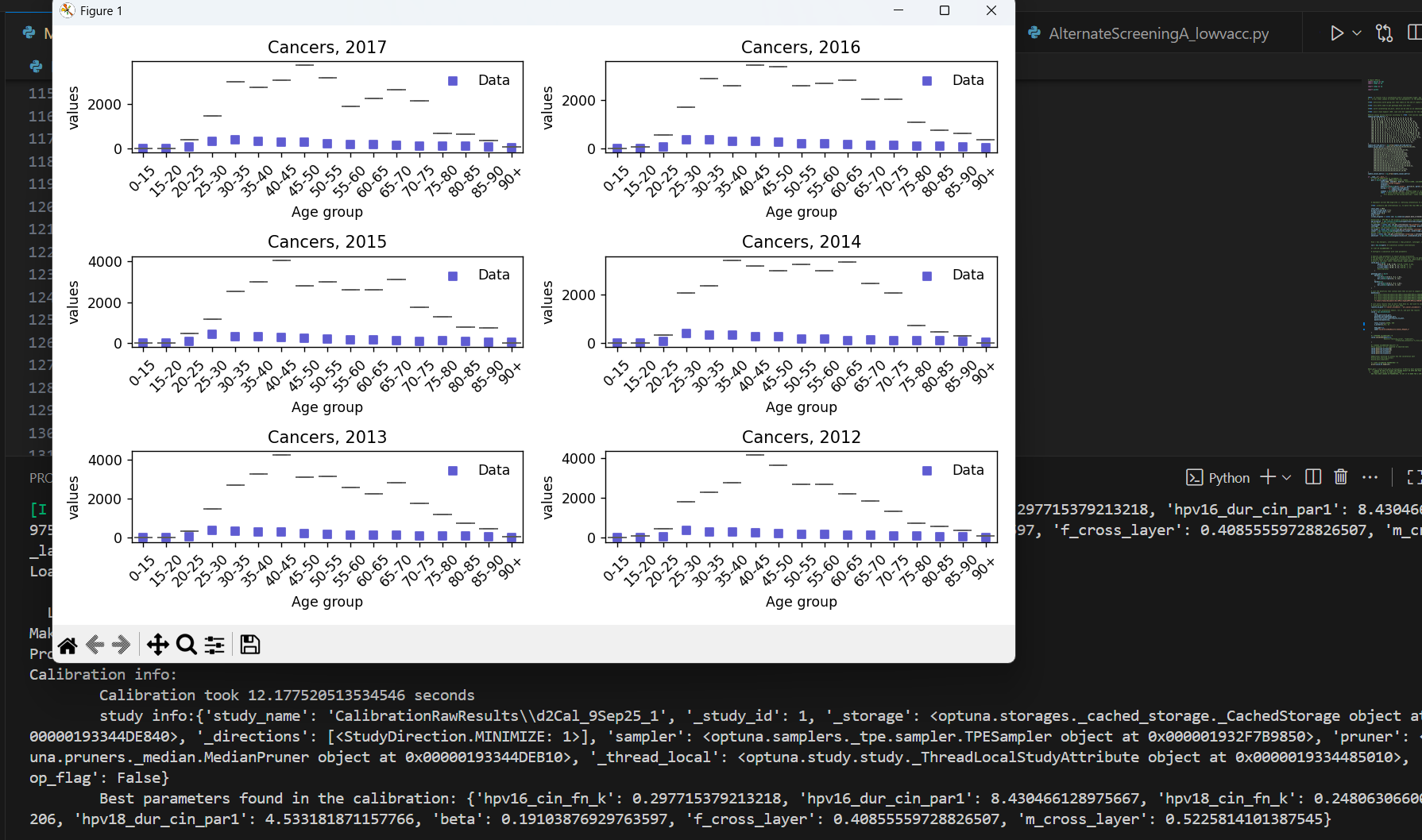
^ 10 runs from the single best of number 6



^ 5 runs for each of the 4 best fits of cal 6

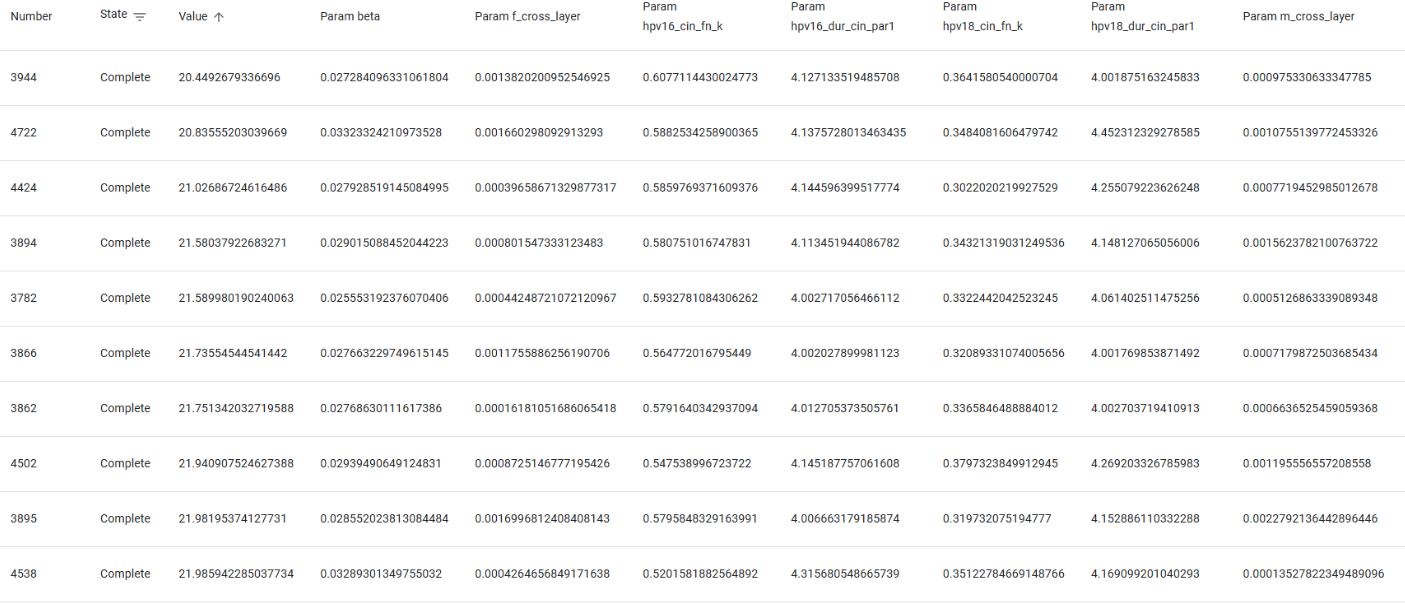
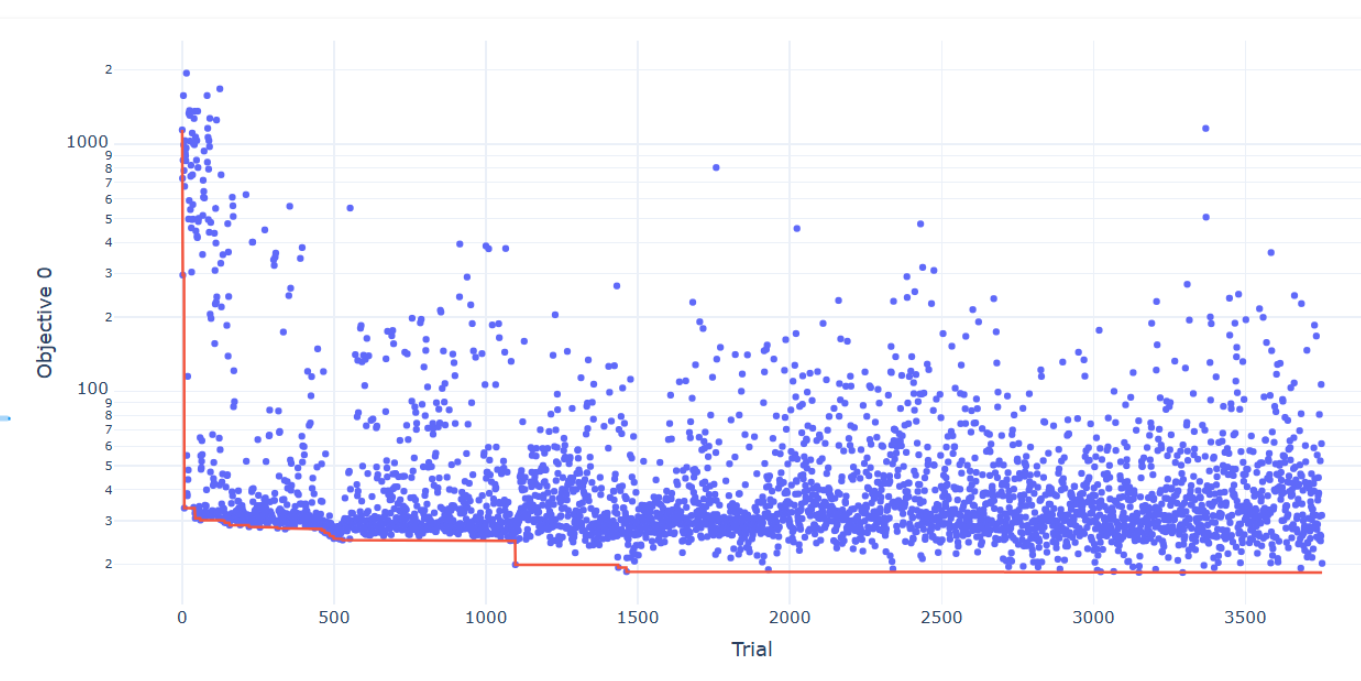
**d2Cal\_9Sep25\_1**

interestingly, when running a single trial cal i found:

* Optuna does not start off with our ‘best guesses’ for parameters, meaning I can’t use that to control the parameters used in a calibration, to retry them
* With the current config (as in **d2Cal\_8Sep25\_5, d2Cal\_8Sep25\_6** ) I am able to get cancers at a sufficiently low age, at least when beta is big enough (noted **d2Cal\_9Sep25\_2** has much smaller beta and no cancers below a certain age, while **d2Cal\_9Sep25\_3** hasa bigger beta of 0.01 at least and does have cancers below a certain age, alebit fewer. I think its fair to blame the differences on the beta)
* 

# C

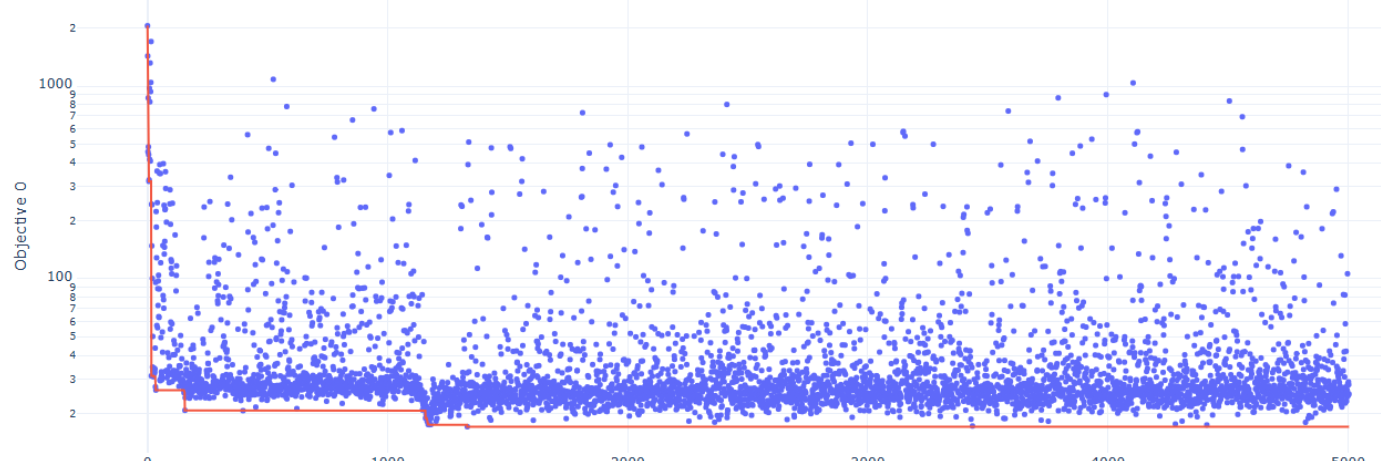
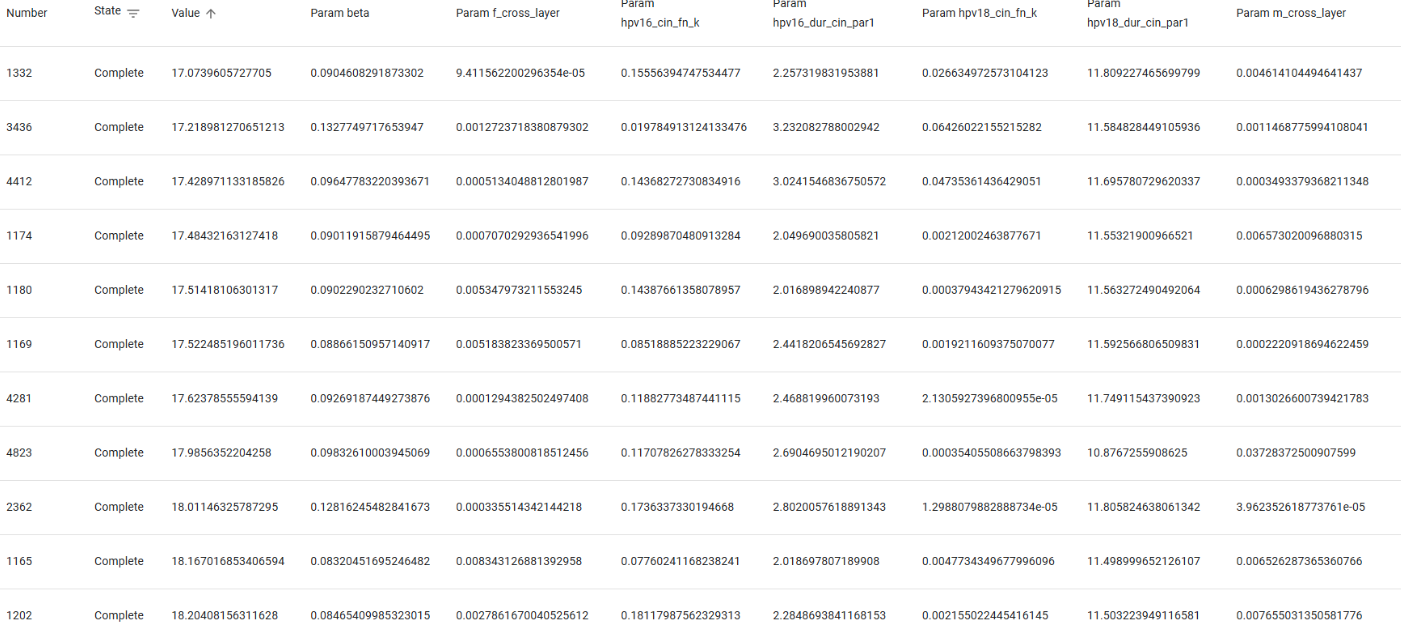
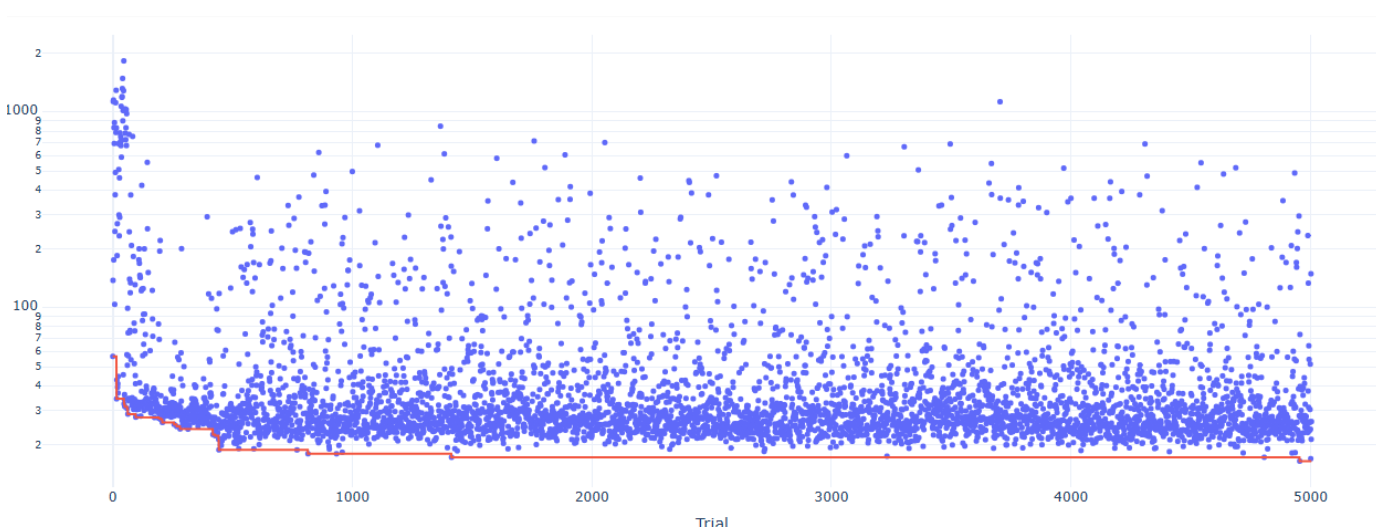
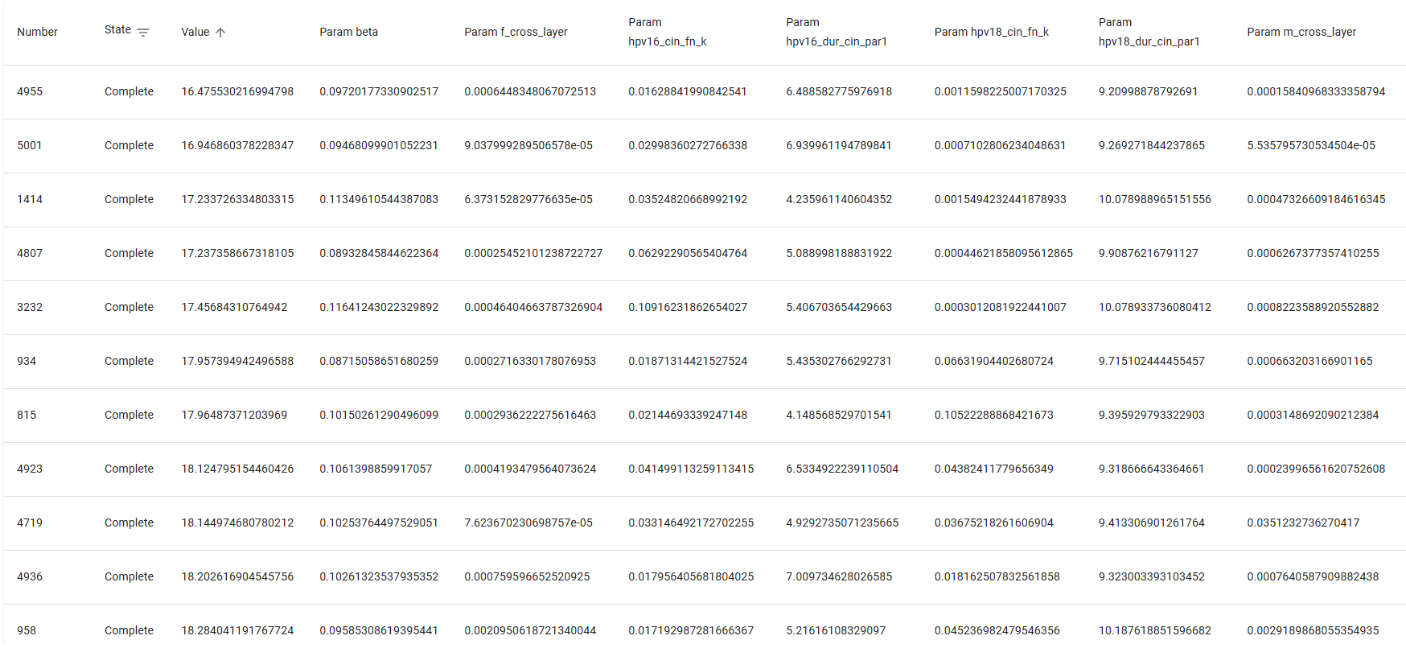
**d2Cal\_10Sep25\_1, d2Cal\_10Sep25\_2**

* It does look better with 100 000 agents, so now I am trying out 250 000 agents.
* As with 100 000 agents we seem to capturing the right level of granularity, i am not expecting a serious improvement if any
* I am running the cal twice to get some robustness against the randomness of the calibration process itself
* In both cals, we don’t see a noteworthy improvement upon our B calibrations
  + Each full trial of the C cals takes 2.5x as long as a trial of the B cals, and they require 2.5x as much memory at their max memory need, which means I can do many fewer in parelell, meaning we get a double whammy in slowdown.
  + … but if it is not signficiantly better than B, it is not worth the slowdown because I can do several B cals in the time of a C cal, and with the noise in calibration results, often trying many identical-set-up cals gets better results than one on-average-better cal that takes the same time
  + [also, given that extending the ranges as in D also gets very good results, we really cant justify increasing to 250 000 agents from 100 000]
* **d2Cal\_10Sep25\_1**
  + 
  + 
* **d2Cal\_10Sep25\_2**
  + 
  + 

# D

A tricky thing here is that different calibrations may end up exploiting different parts of the parameter space and settle on discovering different parts of the parameter space as the best regions, maybe getting as good GOFs as each other but with some parameters really differing. This makes a difference and means I shouldn’t be too hasty in reducing parameter ranges, but does mean I should think more about increasing them at least. Maybe reducing sometimes tho.

**d2Cal\_11Sep25\_ZB\_D2, d2Cal\_11Sep25\_ZB\_D3**

* Calibrating (ZB means on zenbook) identical to **d2Cal\_8Sep25\_5, d2Cal\_8Sep25\_6**, but with extended parameter ranges (CHANGES: 16\_cin\_fn [0.5,0,1], 16\_dur\_cin [6,2,12] and 18\_cin\_fn [0.5,0,1], 18\_dur\_cin [6,3,12])
* **d2Cal\_11Sep25\_ZB\_D2**
  + ****
  + ****
* **d2Cal\_11Sep25\_ZB\_D3**
  + ****
  + ****
* Great news – 100 000 agents with an extended range does seem to consistently get better results than even 250 000 with the old ranges. Nice!
* Looking at these two cals to see if I can extend the range even further, it looks like I could.

**d2Cal\_12Sep25\_ZB\_D4, d2Cal\_12Sep25\_XPS\_D5, d2Cal\_12Sep25\_XPS\_D6**

* CHANGES TO RANGES COMPARED TO A,B,C: 16\_cin\_fn [0.5,0,1], 16\_dur\_cin [6,1,12] and 18\_cin\_fn [0.5,0,1], 18\_dur\_cin [6,1,12]
* Hoping to see at least as good cals as before (that is, after 5000 trials under 20 consistently), if we can get even better, that would be great. (Then I am ready to improve the fixed parameters of the HPVsim model and carefully pick which parameters will get calibrated in the final cals, that is, doing E)