1. Evaluation

- focus on CRPS
- look at IQR or other ranges –> see whether 50 percent of the observations fall inside the IQR
- compare to 1 day-ahead-forecasts (i.e. multiples of 1 day ahead forecasts. Mail to Fabian Krüger to ask wether ratios or differences of crps are in any way meaningful)
- would be good to also have something more interpretable like MSE or absolute error for intuition.

2. Models and Estimation

- AutoAR
  - how many data points do we actually want to include? All past data points?
  - maybe have a mechanism that resets the number of past data points once predictions deteriorate? ⇒ probably should have a look how quickly the model is able to correct for changes in trends.
  - what about any kind of trend? AR assumes basically a constant mean.
  - should we maybe to classical time series analysis things like (p)AC plots, differencing, stationarity checks etc.?
  - do we have to pay attention to stationarity?
- external regressors
  - code certain shocks / interventions as 0/1?
  - use google searches?
    ⇒ general idea: can we make use of the delay between infection/symptom/report and use lagged predictors?
- think about multivariate time series? lags ⇒ probably makes less sense now with lock downs.
- Post-hoc correction of predictions
  - some models seem to be consistently underestimating R_t-values. Can we simply multiply these with a factor to make sure bias is equal to 0.5? One way would be to use model stacking, if we e.g. took the 90% quantile of the forecast in and treat it as a different model, that could draw all estimates upwards where needed.
  - Can we correct the predictive distributions so that they are well calibrated?
    ⇒ need to know about correlation between Bias and bias distribution over time.
- Model Stacking
  - use Regression where all parameters are constrained to sum up to one.
  - literature research: find other ways to do that.

3. Literature

3.1 A simple approach to measure transmissibility and forecast incidence

https://www.sciencedirect.com/science/article/pii/S1755436517300245

3.1.1 Aim

Forecasting of Ebola using R_t

3.1.2 Summary

The study does exactly the same thing we do, albeit probably less sophisticated. They take (simulated?) cases and a linelist, use the renewal equation to compute R_t-values using an averaging window of 2-4 weeks, then project these cases into the future using a branching process. I am not sure whether they just assumed the R_t to be constant in the future.

For evaluation, they calculated the number of true values outside the IQR (should be 50%) and used MSE.

3.1.3 Take away

"Our approach worked best when near-future patterns of incidence were well described by an exponential trend." Might be due to constant R_t?

I have the feeling the paper is interesting for us as a proof of concept, but ultimately pretty useless...

3.2 Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions

http://www.stat.cmu.edu/ ryantibs/papers/deltadens.pdf

3.2.1 Aim

Improve Flu forecasting by using delta-densities and ensemble models.

3.2.2 Summary

They use "delta densities", which assumes an autoregressive dependency structure, but uses a kernel density estimation approach to model these dependencies rather than the common choice of linear relationships plus Gaussian noise → similar to method of analogues (haven't researched it yet), also similar to Kernel Conditional Density Estimation forecasting. Also includes a lot of references / info on model stacking / averaging.

For evaluation, they use binned log scores (not sure why binned) and absolute errors.

3.2.3 Take Away

Very interesting, sophisticated and complicated. Might be useful to look the ensemble things more closely. Also the delta density seems very interesting. One problem with the paper for us is that for flu, past data from previous years is available. So not sure how good we can use their methods, as we don't have that.