Priors for model parameters

Robert Link

3/27/2020

```
## Loading required package: foreach

## Loading required package: iterators

## Loading required package: parallel
```

Introduction

Our model has 16 parameters (counting the schedules each as a single parameter, and not counting the initial state parameters). The full list of parameters is given in Table 1. However, not all of these parameters are plausible MC parameters. Because T_0 is not directly observable, and because we don't have a great deal of data to use to infer it, we are considering that parameter to be deeply uncertain. We are arguably in the same situation for D_0. The schedules for changes in β and D are presumed to be affected by policy, and therefore should also be scenario-based. Finally, the various time delay parameters affect the timing of events, but their effect on the dynamics is small, so we can consider leaving them out for now (but they might be important when we start trying to apply observational constraints). That leaves us with just 6 parameters that we will want to include in MC simulations for the ODE model (the agent model undoubtedly has additional parameters of its own.)

Table 1: Table 1: List of model parameters

name	desc
T0	Initial infection doubling time
D0	Infection duration
beta_schedule	Schedule for changes in infection rate
gamma_schedule	Schedule for changes in infection duration
symptoFraction	Fraction of infected persons with symptoms
hospFraction	Fraction of symptomatic persons requiring hospitalization
ratioICUtoAcute	Fraction of hospital admissions that go to ICU; the rest are acute.
(Note that the name	of this parameter is a little misleading.)
fractionMV	Fraction of hospital patients requiring any ventilator
hospFractionNIPPV	Fraction of hospital patients requiring non-invasive PPV
hospFractionIMV	Fraction of hospital patients requiring invasive ventilator
timeToSympto	Time delay for symptoms
timeToAcuteHosp	Time delay for Acute admission
timeToICU	Time delay for ICU admission
timeToMV	Time delay for ventilator (all)
timeToNIPPV	Time delay for non-invasive ventilator
timeToIMV	Time delay for invasive ventilator

Determining priors

Most of the parameters we are looking at are fractions, suggesting a Beta distribution for our prior. If you estimate such a fraction by observing cases, and you see M positive outcomes and N negatives, then the distribution for the probability p of a positive outcome is $p \sim \text{Beta}(M+1, N+1)$. Several of the parameters are based on reports in the literature where the value was estimated in precisely this way, so this is our best option for a prior for these parameters. This, of course, only takes into account the statistical uncertainty in the parameters. Uncertainty arising from differences in the population or differences in the procedures in other medical systems are not captured.

Table 2: Table 2: Fractional parameter es

parameter	estimate	num.cases	source	notes
symptoFraction	0.43	23	http://dx.doi.org/10.15585/mmwr.mm6913e1	There are more studies a
hospFraction	0.03	NA	JV assessment	Need to discuss more wh
hospFraction (alternate)	0.30	2449	http://dx.doi.org/10.15585/mmwr.mm6912e2	It is unclear what the de
ratioICUtoAcute	0.24	508	http://dx.doi.org/10.15585/mmwr.mm6912e2	
fractionMV	0.32	201	doi:10.1001/jamainternmed.2020.0994	
hospFractionNIPPV	0.26	201	doi:10.1001/jamainternmed.2020.0994	
hospFractionIMV	0.07	201	doi:10.1001/jamainternmed.2020.0994	

Table 2 gives the fractional parameters, the number of cases they were based on, and their sources. The ventilator and ICU fractions seem to be well characterized in the studies we have used so far. For example, the `fractionMV` parameter would have a `beta(65,138)` distribution, which produces a 95% credible interval of $0.26 - 0.39$.

The remaining parameters are for the time being less well constrained. For the symptomatic fraction, I don't think the information we have is the last word on the subject; we should be able to find something more comprehensive. The hospitalization fraction seems to be the subject of some dispute, so we need to get that sorted out. The source of the `hospFrac` figure we are using was unclear from the comments in the original source code, so I have arbitrarily selected an N of 100 for purposes of estimating the uncertainty on that parameter. This gives us a 95% CI of $0.011 - 0.084$ for that parameter.

MC simulations

I went ahead with some simulations using the priors we have, for the 5-day and 7.4-day doubling times. Code and results are below.

```
## Produce baseline simulations
library(CovMitigation)
tmax <- 180
baseline5day <- run_scenario(tmax, list(T0=5), '5-day baseline')
baseline7p4day <- run_scenario(tmax, list(T0=7.4), '7.4-day baseline')
```

```
set.seed(867-5309)
## Run the Monte Carlo calculations
distparms <- list(
  symptoFraction = genbeta(11,14),
  hospFraction = genbeta(4,98),
  ratioICUtoAcute = genbeta(122,388),
  fractionMV = genbeta(65,138),
  hospFractionNIPPV = genbeta(53,150),
  hospFractionIMV = genbeta(15, 188)
)
```

```r
N <- 1024

mc7p4 <- run_mc(N, distparms, tmax, list(T0=7.4), '7.4-day MC')
mc5 <- run_mc(N, distparms, tmax, list(T0=5), '5-day MC')

library(ggplot2)
library(ggthemes)
makeplot <- function(baselines, mc, output, timevar, title)
{
  baselines[['time']] <- baselines[[timevar]]
  baselines[['value']] <- baselines[[output]]
  baselines <- select(baselines, time, value, scenario)

  ggplot(mapping=aes(x=time)) +
    geom_ribbon(data=mc, mapping=aes(ymin=q05, ymax=q95, fill=scenario), alpha=0.5) +
    geom_line(data=baselines, mapping=aes(y=value, color=scenario), size=1.25) +
    ylab('Daily new cases') + ggtitle(title) +
    theme_bw() +
    scale_color_solarized() + scale_fill_solarized()
}

baselines <- bind_rows(baseline5day, baseline7p4day)

plthosp <- makeplot(baselines, bind_rows(mc5$acuteHosp, mc7p4$acuteHosp), 'acuteHosp', 'daysToAcuteHosp
pltICU <- makeplot(baselines, bind_rows(mc5$icuHosp, mc7p4$icuHosp), 'icuHosp', 'daysToicuHosp', 'Daily
pltvent <- makeplot(baselines, bind_rows(mc5$imvHosp, mc7p4$imvHosp), 'imvHosp', 'daysToIMVHosp', 'Daily

print(plthosp)
```
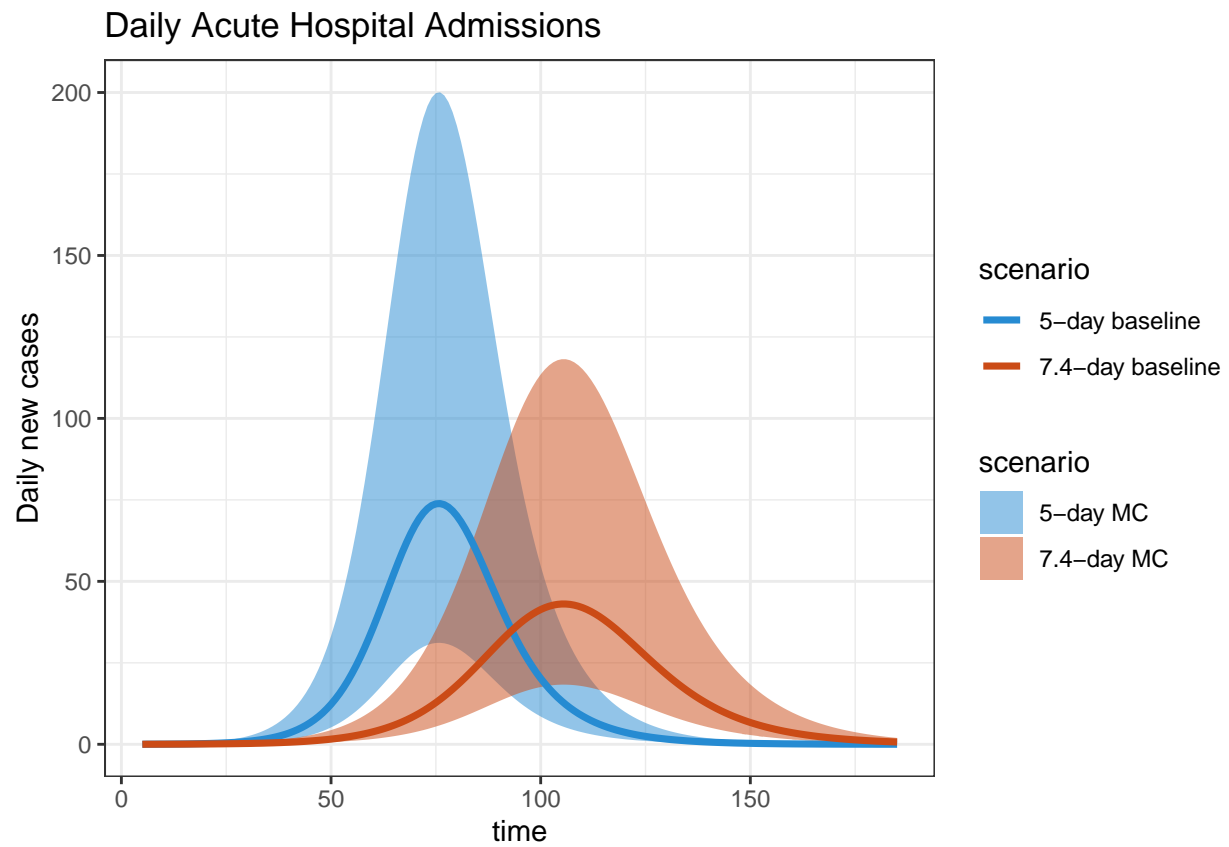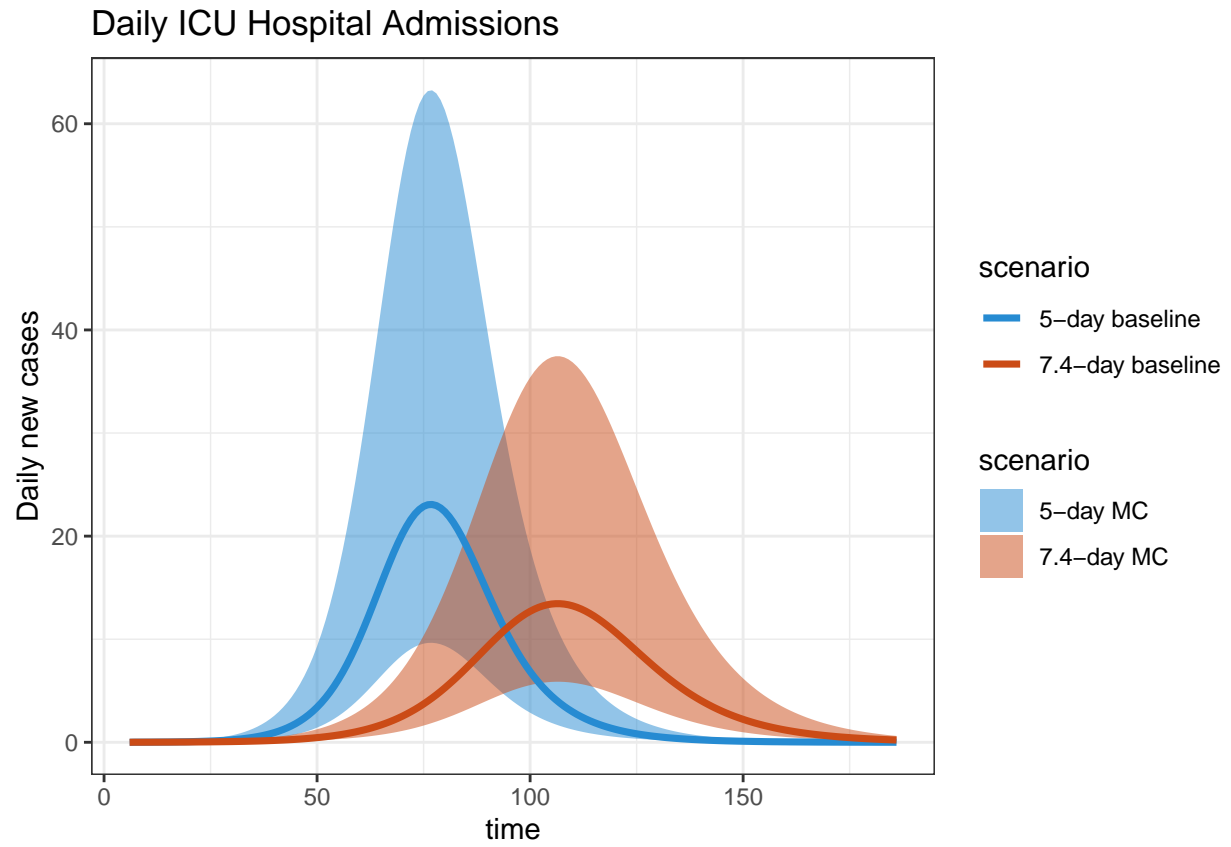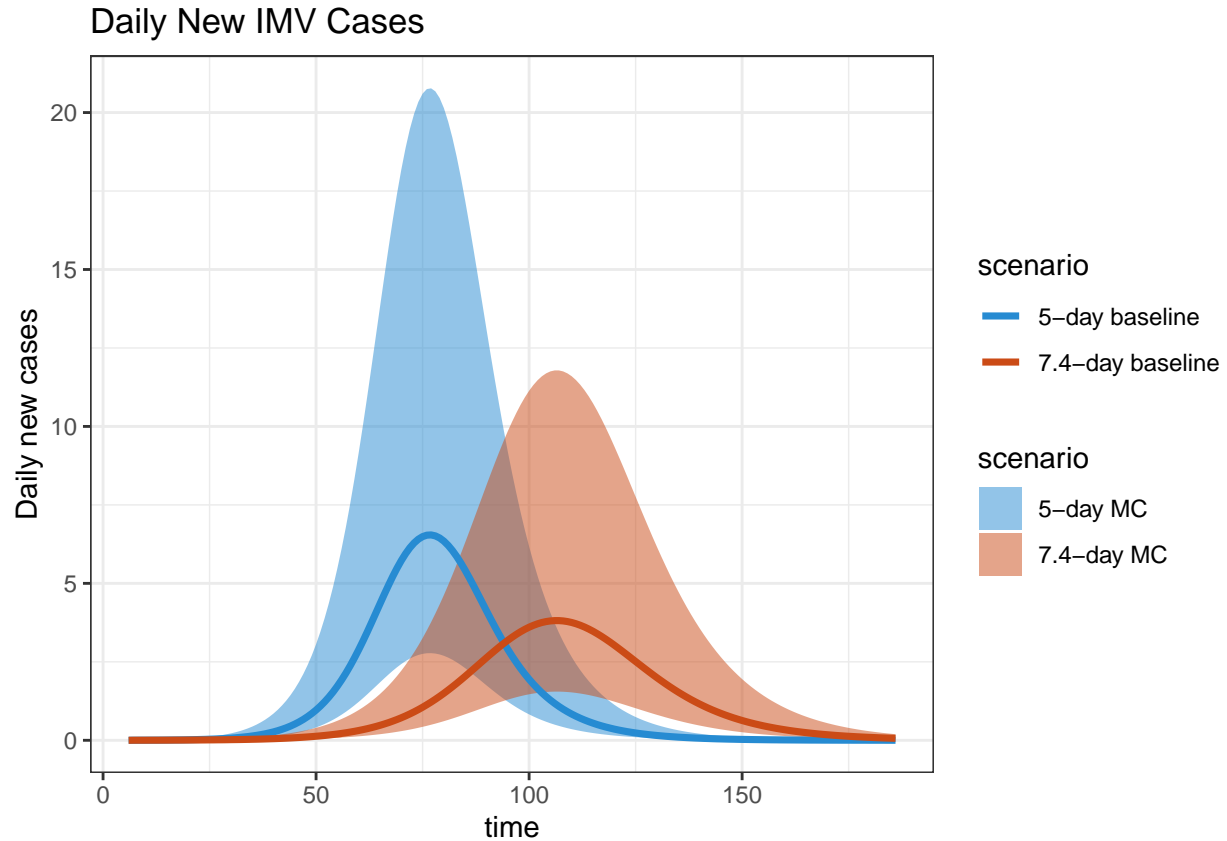
Daily Acute Hospital Admissions

```
print(pltICU)
```

Daily ICU Hospital Admissions

```
print(pltvent)
```

Daily New IMV Cases

What is the main driver of this uncertainty? Naively I would have thought it was `symptoFraction`, since it is currently based on the fewest observations, but we can apply each distribuion individually to see how much each one is contributing to the total. Each output is affected slightly differently by the parameters, so we will look at acute hospital admissions, ICU admissions, and IMV cases.

```
rununcert <- function(T0, scenName) {
  tmp <-
    lapply(names(distparms),
           function(parm) {
             dp <- distparms[parm]
             mc <-
               run_mc(N, dp, tmax, list(T0=T0), scenName)
             for (i in seq_along(mc)) {
               mc[[i]][['uncertParm']] <- parm
             }
             mc
           })
  ## We now have a nested list.  The outer list is different uncertainty parameters,
  ## and the inner list is different outputs.  We want a single list of merged results
  ## for each output.
  outvars <- names(tmp[[1]])
  reorg <-
    lapply(outvars,
           function(outvar) {
             bind_rows(lapply(tmp, function(l) {l[[outvar]]}))
           })
  names(reorg) <- outvars
```

```r
    reorg
}

uncert7p4a <- rununcert(7.4, '7.4-day MC')
uncert5a <- rununcert(5, '5-day MC')

makeuncertplot <- function(baselines, mc, output, timevar, title)
{
  baselines[['time']] <- baselines[[timevar]]
  baselines[['value']] <- baselines[[output]]

  baselines <- select(baselines, time, value, scenario)

  ggplot(mapping=aes(x=time)) +
    geom_ribbon(data=mc, mapping=aes(ymin=q05, ymax=q95, fill=scenario), alpha=0.5) +
    geom_line(data=baselines, mapping=aes(y=value, color=scenario), size=1.25) +
    facet_wrap(~uncertParm) +
    ylab('Daily new cases') + ggtitle(title) +
    theme_bw() +
    scale_color_solarized() + scale_fill_solarized()
}

pltuncert_acute <- makeuncertplot(baselines,
                                  bind_rows(uncert7p4a[['acuteHosp']], uncert5a[['acuteHosp']]),
                                  'acuteHosp', 'daysToAcuteHosp', 'Daily New Acute Admissions')

pltuncert_ICU <- makeuncertplot(baselines,
                                bind_rows(uncert7p4a[['icuHosp']], uncert5a[['icuHosp']]),
                                'icuHosp', 'daysToicuHosp', 'Daily New ICU Cases')

pltuncert_IMV <- makeuncertplot(baselines,
                                bind_rows(uncert7p4a[['imvHosp']], uncert5a[['imvHosp']]),
                                'imvHosp', 'daysToIMVHosp', 'Daily New IMV Cases')

print(pltuncert_acute)
```
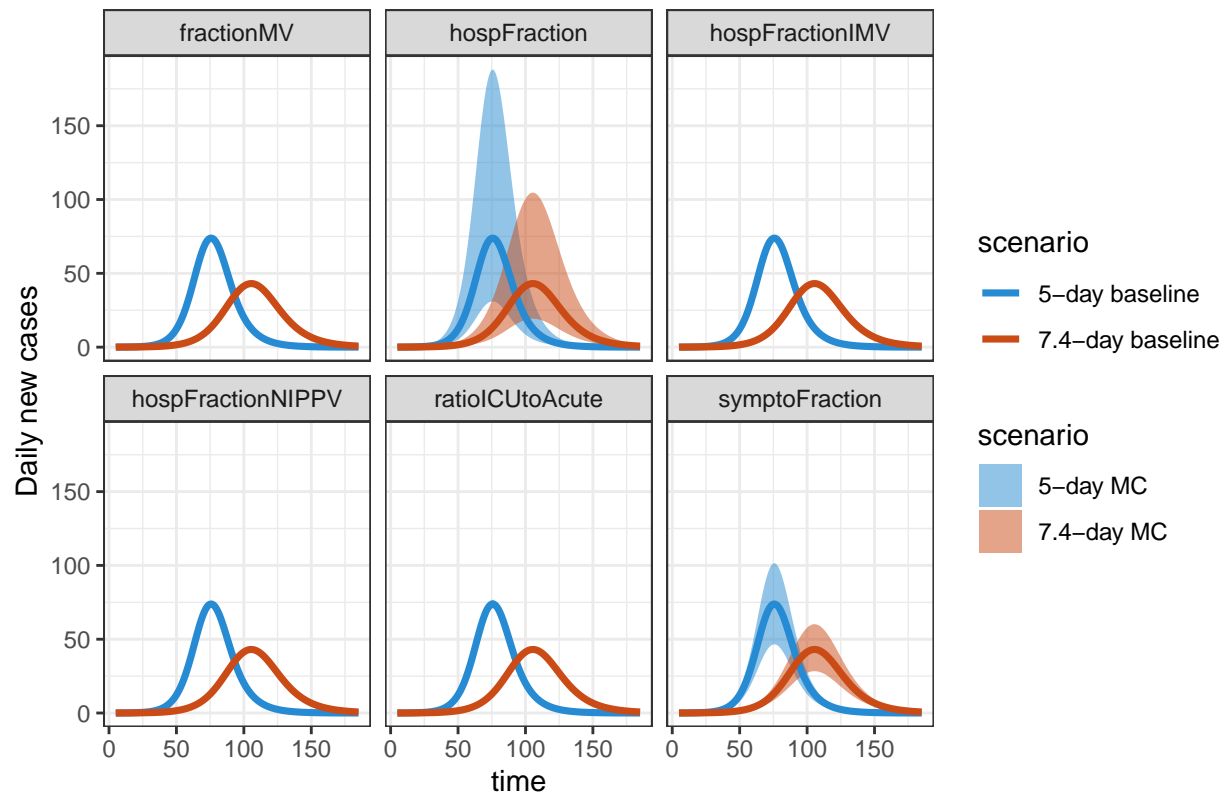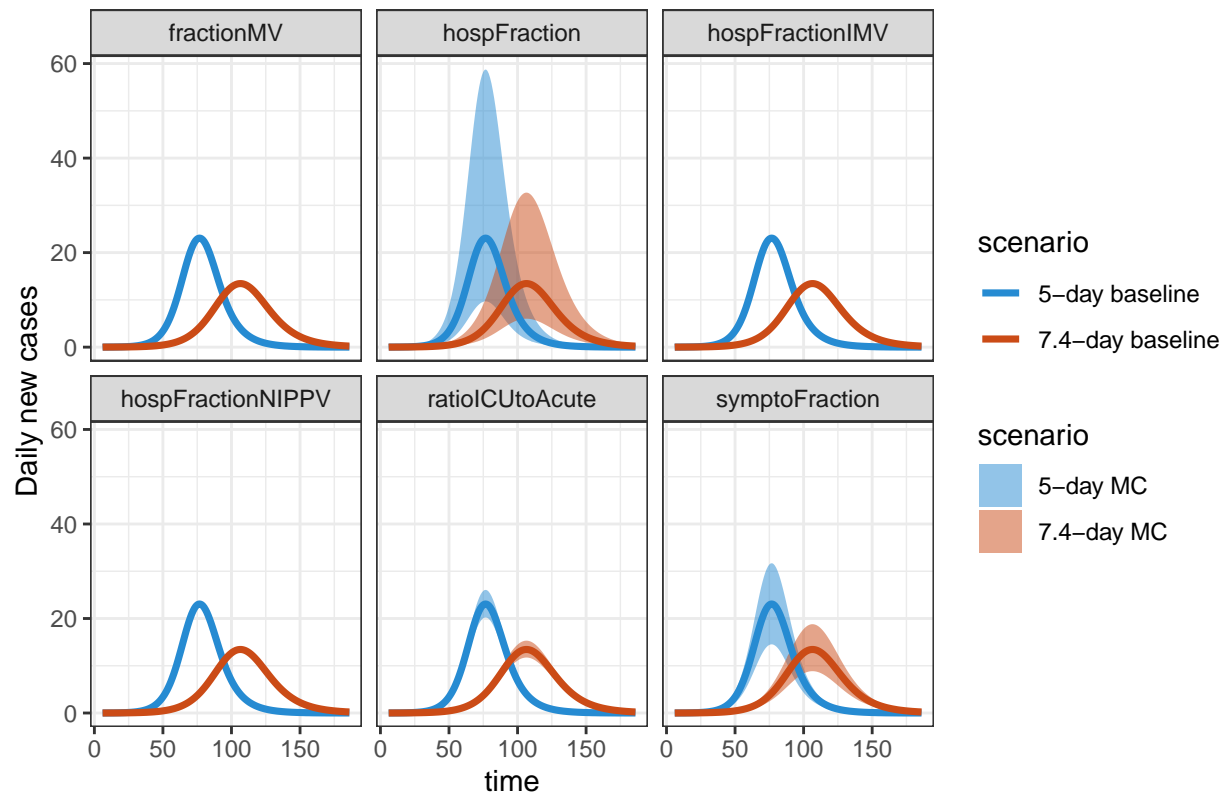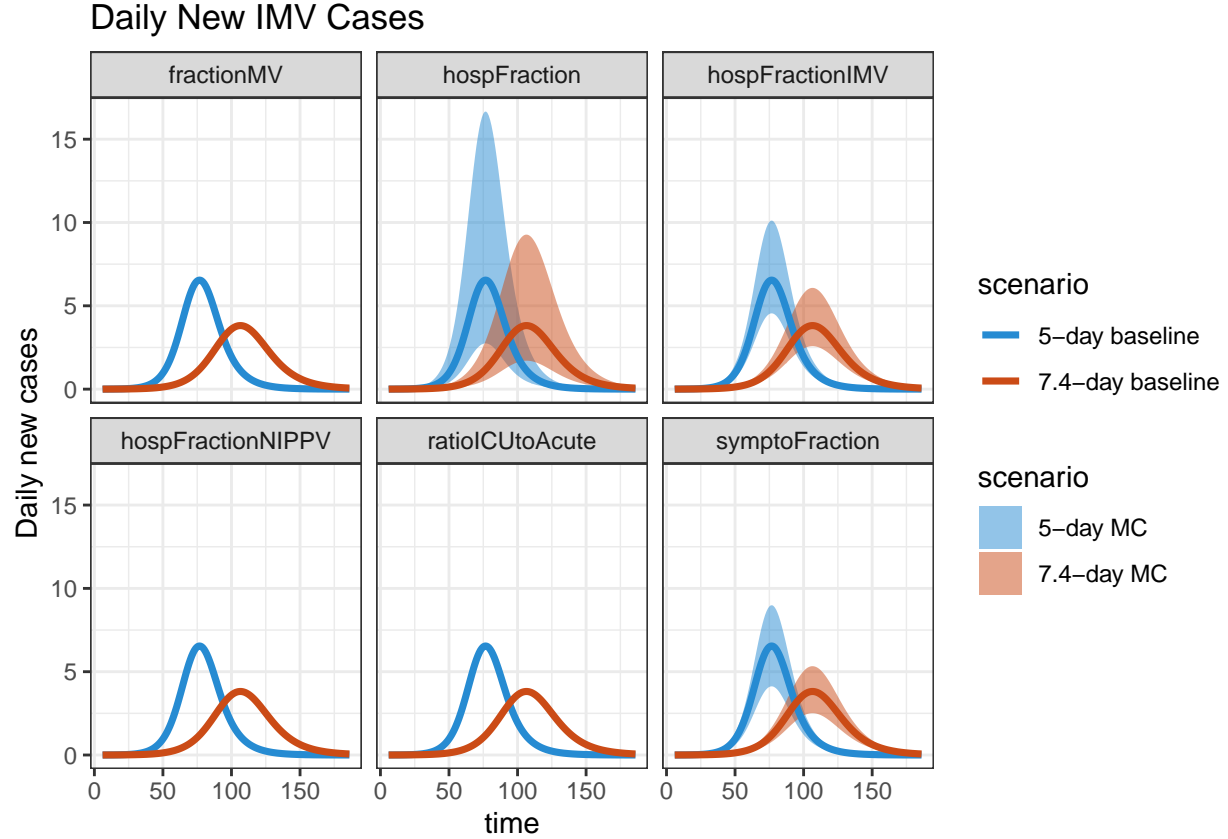
Daily New Acute Admissions

```
print(pltuncert_ICU)
```

Daily New ICU Cases


```
print(pltuncert_IMV)
```

Daily New IMV Cases

Surprisingly, `symptoFraction` isn't a big contributor to the uncertainty in any of the outputs. The main contributor in all cases is `hospFraction`. For both acute and ICU admissions, that is pretty much the end of the story. The fraction of patients that go to the ICU appears to be sufficiently determined that it's not a big contributor to the overall uncertainty. For IMV cases, `hospFractionIMV` is also an important contributor, though nowhere near as much so as the overall `hospFraction`

Conclusion

These results are for the ODE version of the model, but it seems likely that the relative importance of the uncertainty sources will carry through into the agent-based model. Right now, the most important source of uncertainty in our modeling appears to be the `hospFraction` parameter. Perhaps surprisingly, even a crude estimate of the fraction of symptomatic cases seems to be sufficient for our modeling purposes.

In terms of resolving uncertainty, the most important thing we can do is to nail down the estimate of the fraction of symptomatic cases that result in hospitalization. This parameter is upstream from everything we calculate, and it is currently very uncertain, making it our most critical information need right now.