COVID-19 modelling report from McMaster: **Ontario** analysis and forecasting

Ben Bolker, Jonathan Dushoff, Michael Li and David Earn
earn@math.mcmaster.ca

April 28, 2020 @ 20:48

Contents

[**DE:** *`Makefile` doesn't know to check for new data in Mike's repo.* `touch ontario_clean.R` *before compiling this document.*]

1 Summary of Ontario analysis

- Date range for calibration: 2020-01-30 – 2020-04-27
- Social distancing measures have reduced the effective reproduction number ($\mathcal{R}_t$) substantially, but do not appear to have reached the point of suppression (we still estimate $\mathcal{R}_t > 1$ with 95% confidence). [**BB:** *The fits are **still** quite unstable, even with only two break ponts ...*] [**DE:** *Would it be reasonable to add: "It is important to note that the fits are unstable, in the sense that relatively small changes in unknown parameter values (e.g., mean time from infection to symptom onset) can yield large changes in peak time and peak height."*]
- The peak of the epidemic could still be months away, after a slow and shallow rise.

- Because $\mathcal{R}_t$ is estimated to be close to 1, any modelling improvements are likely to lead to forecasting an earlier epidemic peak. If $\mathcal{R}_t$ is actually < 1 then we should see decline very soon; if $\mathcal{R}_t$ is larger than we currently estimate then the peak can be expected to be higher as well as sooner.

- Any relaxation of social distancing will likely speed the approach to the peak and increase its height.

2 Questions arising from modelling

- Can we figure out how to "plug the leaks"?
  - Where and why are people being infected? Can we quantify transmission (contributions to $\mathcal{R}_0$) by age? location? population density? ...
  - Are infected people not self-isolating?
  - Are travellers not self-quarantining?
  - Is PPE inadequate?
  - Are non-public data sufficiently detailed that we can potentially answer some of these questions? [**DE:** *Kevin Brown at PHO is looking at this; looks like Toronto is a problem; they believe (based on bluedot data) that people coming from the US are not self-quarantining.*]

- ICU counts continue to diverge fairly sharply from the model predictions, which is a phenomenon that we do not understand. This is not well explained by saturation of capacity – the current counts are still well below capacity. (The puzzling patterns are in relation to hospitalizations and ICU, not deaths.)
  - *Note:* The model assumes that a constant fraction of infections are severe, a constant fraction of severe infections are hospitalized, and a constant fraction of hospitalized cases are transferred to the ICU. Consequently, the model cannot predict a flattening of the ICU curve before everything else flattens.
  - Is it possible that the epidemic is actually burning out in LTCFs? If so, this could potentially explain the drop we're seeing. What fraction of cases are HCWs vs LTCF residents? How much under-reporting is there in LTCFs? If LTCF epidemic is peaking earlier, is this because infection rate was extraordinarily high in LTCFs, or that lockdown has a particularly strong effect on LTCFs?

3 Limitations and concerns

- Analysis presented here is based on the **public data** that is compiled and cleaned daily by Michael Li; see `https://wzmli.github.io/COVID19-Canada/`.

- The quality of the data is unclear. [**DE:** *Ask Kevin which columns can be trusted.*]

- Testing lags for HCWs are likely much shorter than for the general public.
- A large proportion of those who have died were residents of long term care facilities (LTCFs, which are not modelled explicitly).
- At present, we do not attempt to model any heterogeneities of transmission associated with age.
- Preliminary analysis of Ontario Laboratory Information System (OLIS) data from PHO reveals that many individuals have been tested more than once. The public testing data do not account for this.
- The mean lag between incidence (infection) and reporting a case is currently assumed to be 11 days. [**DE:** *DC (following advice from Kevin Brown) uses a 9 day lag between symptom onset and report. The lag for HCWs is "much shorter".*]

4 Analysis framework

- Compartmental mechanistic model: susceptible-exposed-infectious-removed (SEIR), with additional compartments for individuals in hospitals in acute care (H) or intensive care (ICU). Infections can be asymptomatic, mildly symptomatic or severely symptomatic. All symptomatic individuals are presumed to have had a period of pre-symptomatic infectiousness.
- Model calibration is performed using maximum likelihood estimation (MLE) by matching deterministic trajectories to reported cases, hospitalizations and deaths, assuming negative binomial observation error.
- Model forecasts take account of the inherent randomness of the processes of transmission, recovery, *etc.* ("process error").

5 Code improvements since last report

- Forecasts include demographic stochasticity ("process error"). [**DE:** *This is now possible, but currently not shown.*]

6 Current results

6.1 Estimated parameters

Estimated parameters are listed in Tables 2 and 4. [**DE:** *This needs to be organized better. $\bar{G}$ is not estimated or derived here, but is important to state.*] [**BB:** *see stuff in ontario calibration report, which includes what is calibrated how*]

Estimated parameters are listed in Tables 2 and 3. Table 3 shows estimated relative changes in $\mathcal{R}_t$ on the dates on which social distancing interventions were implemented in Ontario:

Table 1: Estimated parameter changes during each phase of social distancing.

	Start Date	r_0	$\mathcal{R}_0$	$\bar{G}$	T_2
1	2020-01-30	0.16	2.28	6	4.41
2	2020-03-23	0.15	2.20	6	4.65
3	2020-03-28	0.04	1.23	6	17.18

Table 2: Estimated Parameters. Times are given in days.

	symbol	value	meaning
1	$E(0)$	3.71	number of initially exposed individuals
2	β_0	0.679	transmission rate
3	μ	0.981	proportion of symptomatic infections that are mild
4	ϕ_1	0.413	proportion of severe infections that do NOT require ICU
5	$n.H$	1.02	
6	$n.report$	0.313	
7	$n.death$	0.849	

- 17 March 2020. Declaration of Emergency (schools ordered to stay closed after March Break)
- 23 March 2020. At-Risk Workplaces
- 23 March 2020. All Non-Essential Workplaces
- 28 March 2020. Gatherings of More Than Five People with Strict Exceptions

We assumed $\mathcal{R}_t$ changed abruptly on a subset of these dates and found the best fit changes. [**DE:** *Need to make clear how many and which break dates we ended up including and why we ignored the other(s). It is not clear that the state of emergency had a big impact. Kids were on March Break and were not distancing. Genuine closures didn't happen until 23 March.*]

Table 4 lists parameters derived from the fitted parameters. [**DE:** *Actually, how are these numbers calculated? r_0 and $\mathcal{R}_0$ don't agree with any phase of the outbreak. $\bar{G}$ is not estimated or derived here, but is important to state.*]

Finally, Table 1 shows the growth rates and reproduction numbers during phases of the outbreak during which different social distancing measures were in place. [**DE:** *suppressed warnings (caused by meaningless doubling time when growth rate is negative) in construction of Table 1*]

6.2 Forecasts

Figure 2 shows our current projections. [**DE:** *based on ignoring the first $\mathcal{R}_t$ change time. So why are there three relative $\mathcal{R}_t$ estimates (Table 3) and four absolute $\mathcal{R}_t$ estimates (Table 4)?*]

Table 3: Relative changes in effective reproduction number $\mathcal{R}_t$.

	Intervention Date	Relative change in $\mathcal{R}_t$
1	03/23/2020	0.96
2	03/28/2020	0.54

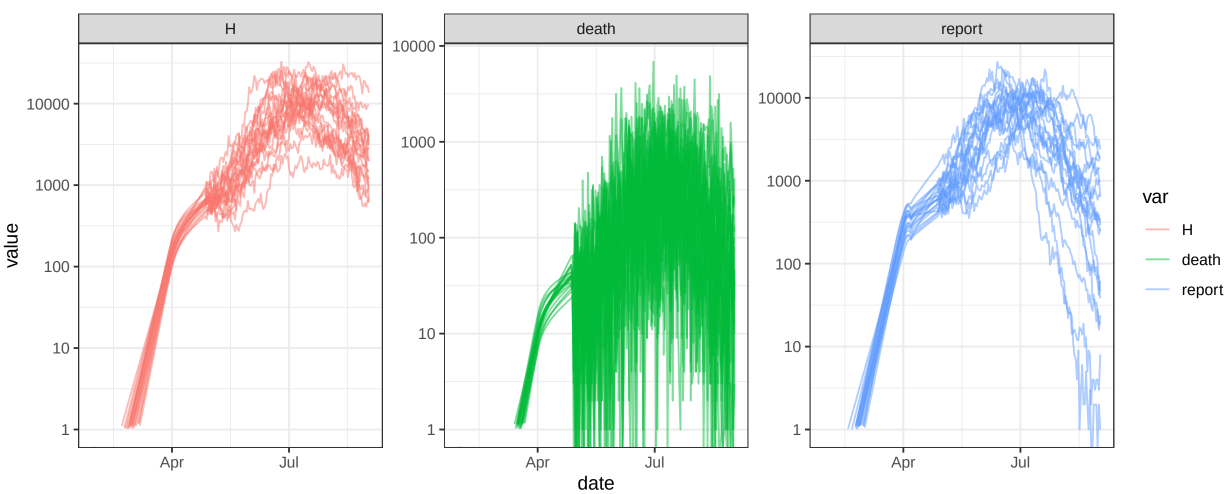
Table 4: Derived estimates. Times are given in days.

	symbol	value	meaning
1	r_0	0.158	initial epidemic growth rate
2	$\mathcal{R}_0$	2.29	basic reproduction number
3	$\bar{G}$	6.01	mean generation interval
4	T_2	4.4	doubling time

The vertical lines show the times of implementation of different social distancing measures (*cf.* Tables 3 and 1). The horizontal dashed lines represent COVID ICU capacity. There are no confidence bands on these plots and they would be very large. [**DE:** *Last week I said: We are working on that, and should have confidence bands on forecasts next week.*]

Using hospitalization-only fit (2 breaks, no priors) right now.

```
#> Warning:  Transformation introduced infinite values in continuous y-axis
```



CIs on peak/peak timing?

	2.5%	97.5%
H	2020-06-18	2020-08-18
death	2020-06-10	2020-08-18
incidence	2020-06-03	2020-07-21
report	2020-06-10	2020-07-27

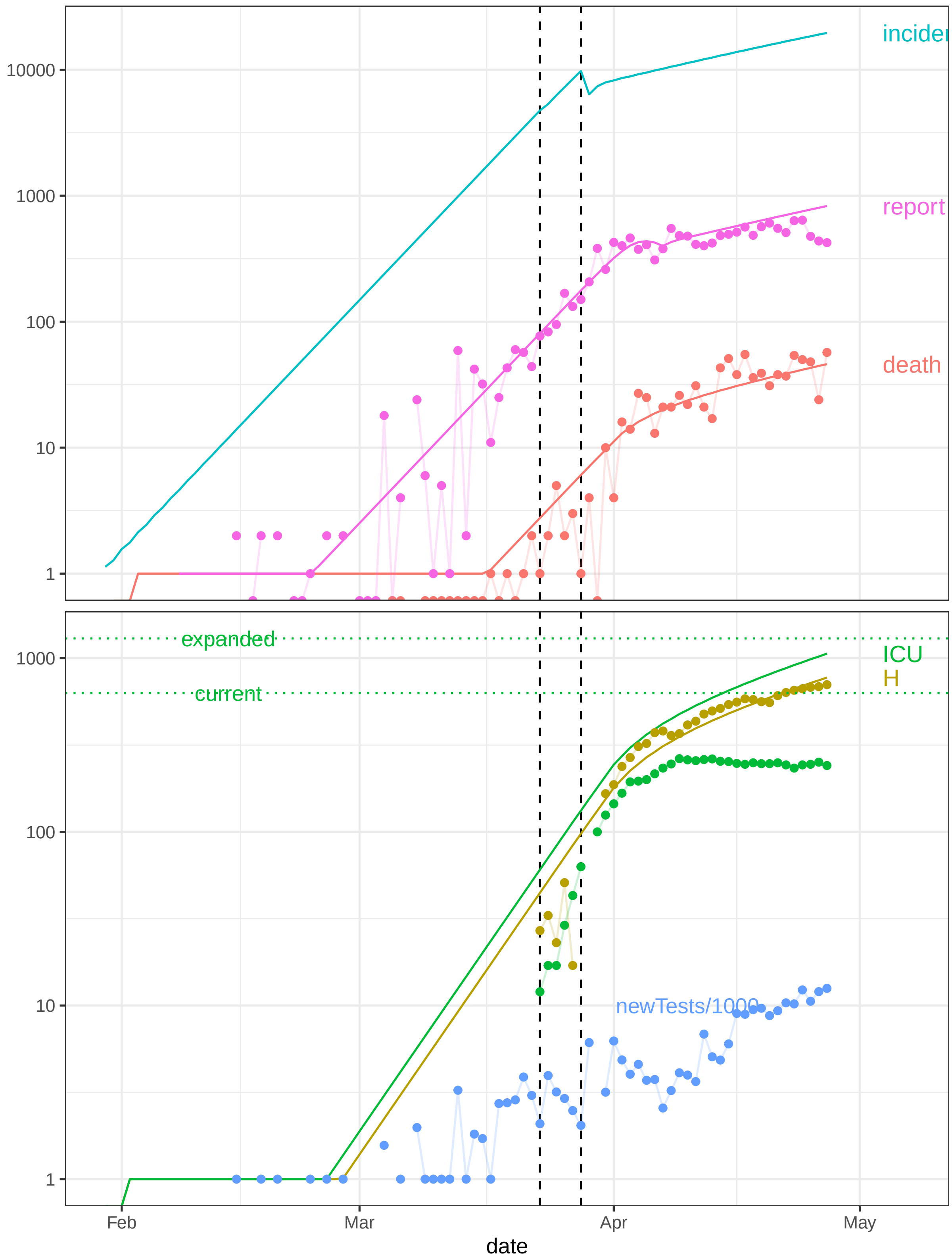


Figure 1: COVID-19 calibration. Dots show reported data. Curves show the fitted model trajectory. *Top panel:* Incidence, reported cases, deaths. *Bottom panel:* Hospitalizations in acute care (H) and intensive care (ICU).

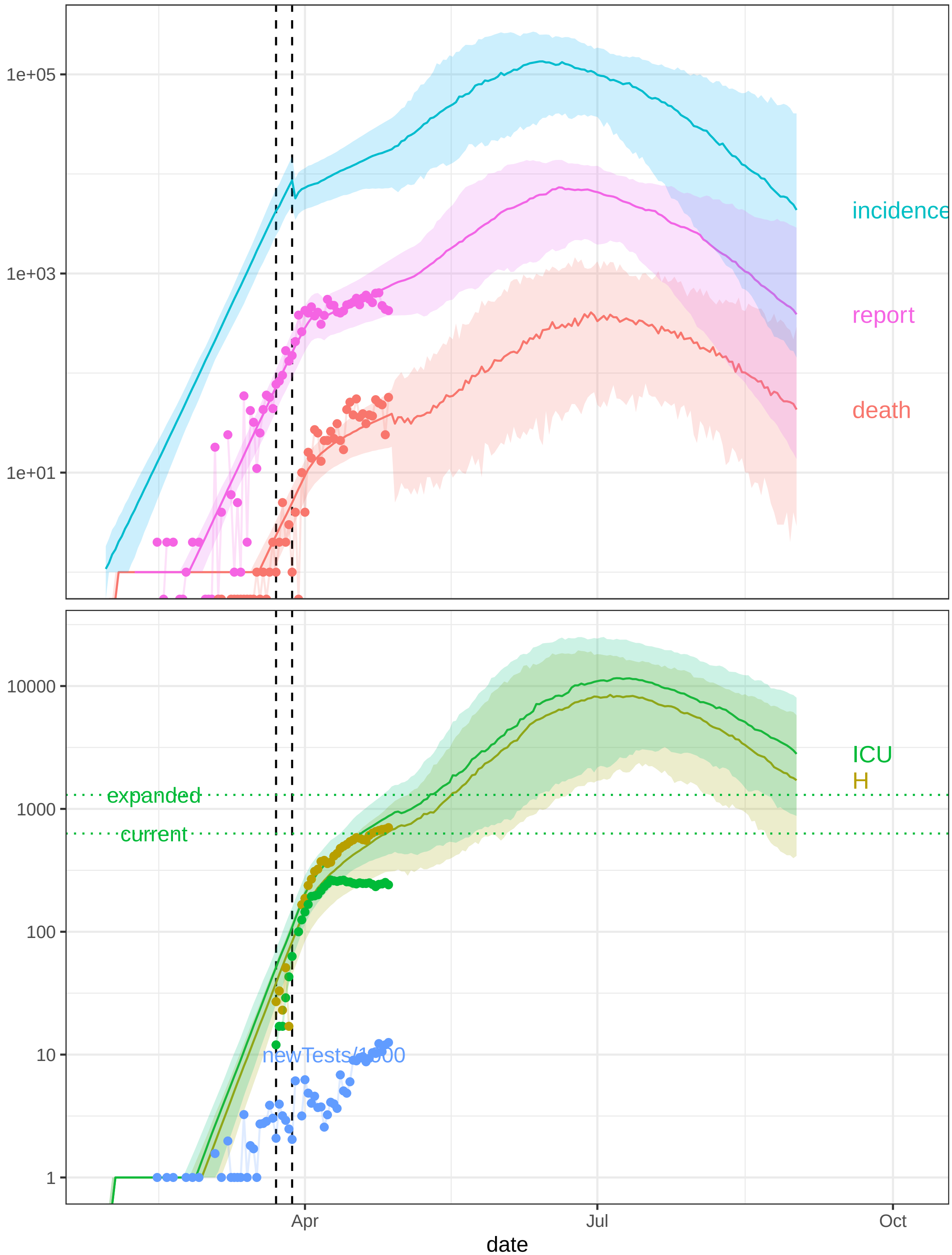


Figure 2: COVID-19 forecast. See caption to Figure 1.

	2.5%	97.5%
H	4576	32601
death	671	5921
incidence	126596	549296
report	5842	24822

caveats

- not doing importance weighting yet
- param CIs are probably too small …
- amount of noise is not calibrated