First part: Internship 2024

Paul PETIT

Based on the studies included in the Alene et al. (2021) and McAloon et al. (2020) meta-analyses, we will perform a linear regression on the estimated mean serial intervals and the estimated mean incubation period as a function of time (midpoint of the study period) during the first wave of the COVID-19 pandemic. To study this in more detail, we might, for example, repeat this analysis for M-Pox or look for other articles relevant to COVID-19 that could increase the power of the statistical analyses...

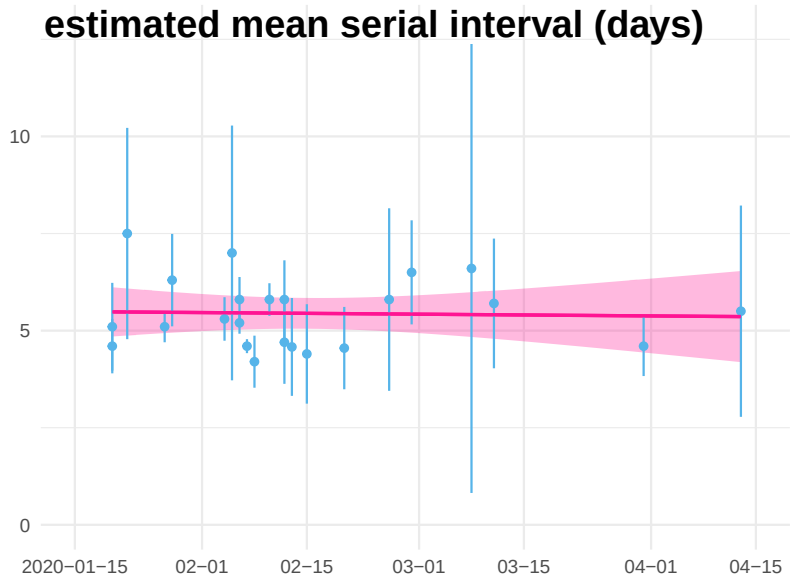
Serial interval linear regression

For serial intervals, only the Alene et al. (2021) meta-analysis was used.

```
OK: residuals appear as normally distributed (p = 0.135).

OK: Error variance appears to be homoscedastic (p = 0.352).

Warning: Autocorrelated residuals detected (p = 0.027).


Call:
lm(formula = Mean ~ mid_date, data = covid_serial_interval)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2549 -0.8181 -0.1606  0.3610  2.0206

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.835726 167.524925   0.190    0.851
mid_date     -0.001442   0.009151  -0.158    0.876
```

```
Residual standard error: 0.9149 on 21 degrees of freedom
Multiple R-squared:  0.00118,   Adjusted R-squared:  -0.04638
F-statistic: 0.02482 on 1 and 21 DF,  p-value: 0.8763
```

The linear regression assumptions are met, with the exception of the non-correlation of residuals. The p-value is between 2% and 3%, and since no clear pattern is evident in the data, it is acceptable to interpret the test results. Therefore, we can conclude that there is no significant time effect on the estimated mean serial interval in the studies analyzed by Alene et al. (2021) in their meta-analysis.



Incubation period linear regression

For incubation periods, both Alene et al. (2021) and McAloon et al. (2020) meta-analyses were used.

```
OK: residuals appear as normally distributed (p = 0.106).
```

```
OK: Error variance appears to be homoscedastic (p = 0.879).
```

```
OK: Residuals appear to be independent and not autocorrelated (p = 0.818).
```

```
Call:
lm(formula = Mean ~ mid_date, data = covid_incub_period)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6702 -0.9976 -0.1957  0.5794  2.3563

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -442.18898  227.97777  -1.940   0.0728 .
mid_date       0.02453    0.01246   1.968   0.0692 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.359 on 14 degrees of freedom
  (1 observation effacée parce que manquante)
Multiple R-squared:  0.2167,    Adjusted R-squared:  0.1608
F-statistic: 3.873 on 1 and 14 DF,  p-value: 0.06919
```

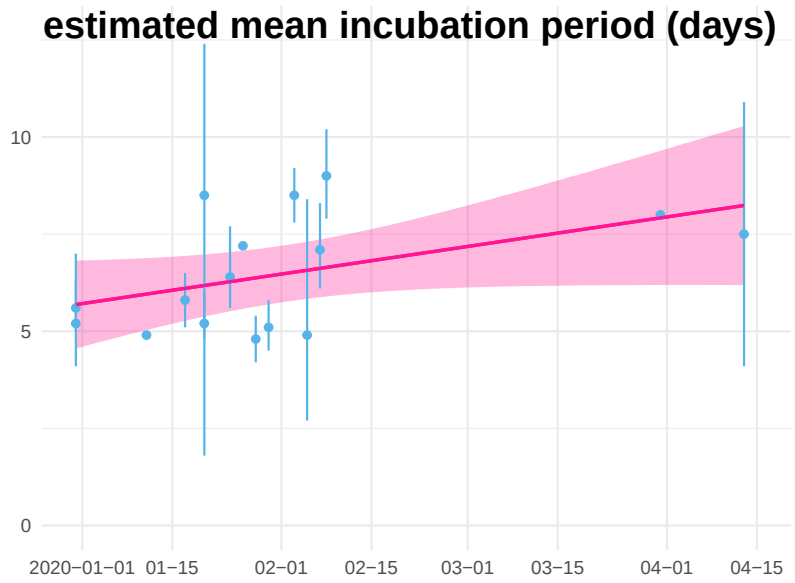
All the linear regression hypotheses are met, so we can interpret the tests results. While no significant association was found between the two variables ($p - value = 0.069$) if we keep a 5% threshold, the positive slope observed aligns with theoretical expectations (Britton & Scalia Tomba, 2019). This suggests a possible trend, though further data is needed to confirm the presence or absence of an effect. Also, the p-value is close to 5% and could be linked to a lack of statistical power. However, it is important to notice the low value or R-squared (0.22) which indicate that our linear regression isn't explaining so much our data.

Conclusion

At the onset of the COVID-19 pandemic, the situation was particularly challenging as we faced a new pathogen with unknown characteristics and that urgent daily decisions had to be made, often with limited data to guide them.

To better prepare for future pandemics, it is essential to understand uncertainties, address biases, and consider factors that might skew estimates of key control parameters, such as the instantaneous reproduction number. This understanding is critical for accurately assessing the impact of public health measures and interventions.

In this context, the theoretical insights provided by Britton & Scalia Tomba (2019) are valuable, but they must be tested in real-world scenarios to verify their influence. While we do observe, as they predicted, an increase in XXX in the estimated mean incubation



period—potentially explained by the factors they highlight (to be specified) — the anticipated relationship with the estimated mean serial interval does not hold. It is likely that adjusting the serial interval estimate accordingly would not enhance the accuracy of $\mathcal{R}_0$ or $\mathcal{R}_t$ estimations.

These findings, of course, need to be validated across different pathogens, such as M-Pox, Ebola, or influenza, where epidemiological data are available. Furthermore, a more comprehensive literature review could identify studies that Alene et al. (2021) and McAloon et al. (2020) may have excluded for reasons not relevant to our analysis. Finally, the same type of study could be extended to other epidemiological parameters to have a clearer view of factors to be taken into account during an emergency situation as the early stage of an outbreak situation.