Uncertainty quantification for the modeling of the Covid-19 outbreak

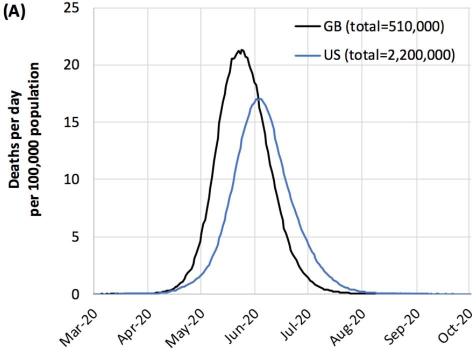
Josselin Garnier (Ecole polytechnique)

July 3, 2020

Overview of the current situation

• Models aim at predicting the evolution of the Covid-19 outbreak. Example: Neil Ferguson, Imperial College COVID-19 Response Team.

Unmitigated epidemic scenario March 12, 2020



• Typical models: Compartmental (SEIR-type) models with geographical and age stratification. A lot of free parameters.

• These predictions are very uncertain ! How to quantify and reduce the uncertainties ?

Classical SEIR model

- Four compartments:
  - $S(t)$ = proportion of susceptible (healthy) individuals at time t ($S(0) = 1$),
  - $E(t)$ = proportion of exposed individuals at time t,
  - $I(t)$ = proportion of infected individuals t,
  - $R(t)$ = proportion of recovered (and immune) or dead individuals at time t.

$$
\begin{aligned}
S'(t) &= -R_0 S(t) I(t) / T_{\rm i}, \\
E'(t) &= R_0 S(t) I(t) / T_{\rm i} - E(t) / T_{\rm e}, \\
I'(t) &= E(t) / T_{\rm e} - I(t) / T_{\rm i}, \\
R'(t) &= I(t) / T_{\rm i}.
\end{aligned}
$$

- Parameters:
  - Mean incubation time: $T_{\rm e}$
  - Mean infectious time: $T_{\rm i}$
  - Basic reproduction number (average number of individuals infected by one person): R_0

Classical SEIR model

- SEIR models can be stratified in age and region.
  - [1] Imperial College COVID-19 Response Team, Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries, March 30, 2020 (*Ferguson*).
  - [2] L. Di Domenico et al., Expected impact of lockdown in Ile-de-France and possible exit strategies, April 12, 2020 (*Colizza*).
  - [3] H. Salje et al., Estimating the burden of SARS-CoV-2 in France, Science, May 13, 2020 (*Cauchemez*).
- These models are based on strong hypotheses:
  - (exponential) distribution of the incubation time (with mean $T_{\rm e}$) and infectious time (with mean $T_{\rm i}$),
  - homogeneity of the population (within strata).
- Possible origins of the heterogeneity:
  - biological reasons: some individuals are more vulnerable or resistant than other ones.
  - social reasons (contact network): some individuals have many more contacts than other ones.

A heterogeneous SEIR model

Eight compartments:

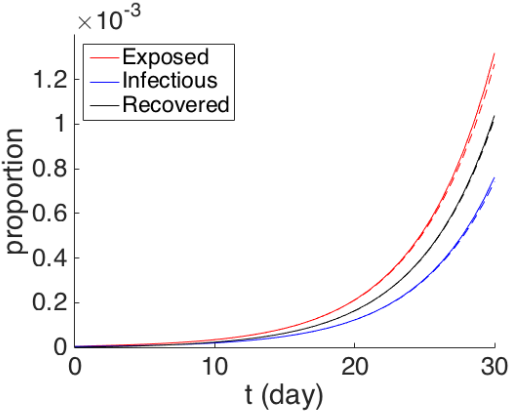
- $S_j(t)$ = proportion of susceptible (healthy) individuals, of type $j = 1$ "vulnerable" or $j = 2$ "resistant",
- $E_j(t)$ = proportion of exposed individuals, of type $j = 1, 2$,
- $I_j(t)$ = proportion of infected individuals, of type $j = 1, 2$,
- $R_j(t)$ = proportion of recovered or dead individuals, of type $j = 1, 2$.

$$
\begin{aligned}
S_j'(t) &= -R_0\, r_j\, S_j(t)(I_1(t) + I_2(t))/T_{\rm i}, \qquad j = 1, 2, \\
E_j'(t) &= R_0\, r_j\, S_j(t)(I_1(t) + I_2(t))/T_{\rm i} - E_j(t)/T_{\rm e}, \qquad j = 1, 2, \\
I_j'(t) &= E_j(t)/T_{\rm e} - I_j(t)/T_{\rm i}, \qquad j = 1, 2, \\
R_j'(t) &= I_j(t)/T_{\rm i} \qquad j = 1, 2,
\end{aligned}
$$

with $S_1(0) = f$, $S_2(0) = 1 - f$, $r_1 f + r_2 (1 - f) = 1$.

Lack of robustness of SEIR models

- Homogeneous SEIR model with $T_{\rm i} = T_{\rm e} = 4$ days and $R_0 = 3$.
- Heterogeneous SEIR model with $T_{\rm i} = T_{\rm e} = 4$ days, $R_0 = 3$, $f = 10\%$ of "vulnerable" individuals, 1-$f = 90\%$ of "resistant" individuals, $r_1/r_2 = 10$.



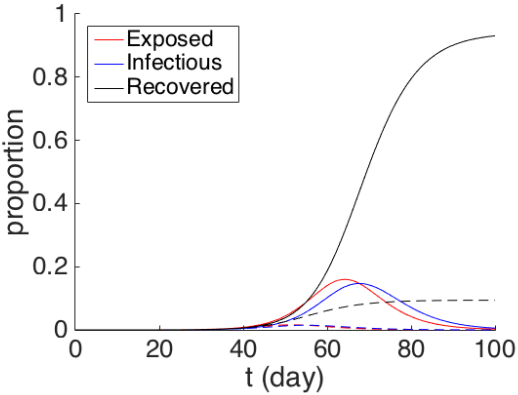
Solid lines: homogeneous SEIR
Dashed lines: heterogeneous SEIR

The two models can fit the same data.

$\hookrightarrow$ same R_0 in both models.

Lack of robustness of SEIR models

- Homogeneous SEIR model with $T_{\rm i} = T_{\rm e} = 4$ days and $R_0 = 3$.
- Heterogeneous SEIR model with $T_{\rm i} = T_{\rm e} = 4$ days, $R_0 = 3$, $f = 10\%$ of "vulnerable" individuals, 1-$f = 90\%$ of "resistant" individuals, $r_1/r_2 = 10$.

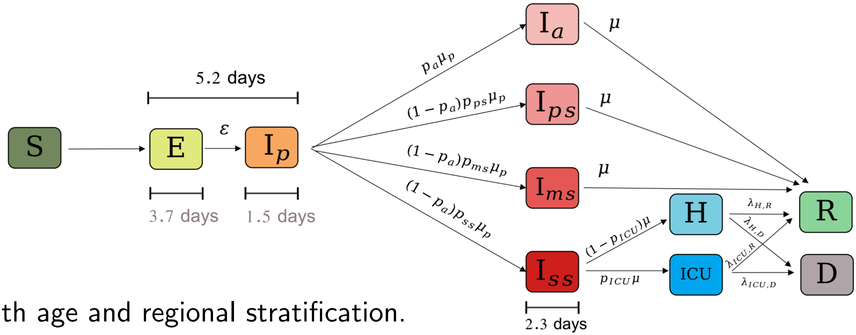


Solid lines: homogeneous SEIR
Dashed lines: heterogeneous SEIR

The heterogeneous model predicts that herd immunity is achieved with $\sim 10\%$ *seropositive.*

→ The two models have the same likelihood (w.r.t. data) but they give different predictions.

Uncertainty quantification strategy



with age and regional stratification.

1. Calibrate the parameters of the model by least-square (or maximum likelihood) fit.
2. Predict the evolution of the outbreak with the estimated parameters.
3. Determine by sensitivity analysis the "important" parameters.
4. Determine by Bayesian analysis the a posteriori distribution of the "important" parameters.
5. Propagate the uncertainty of the "important" parameters into the predictions.

Conclusions

- A lot of uncertainties:
  - Questions on the homogeneity of the population.
  - Need for uncertainty quantification.
- With the available data:
  - The epidemiological models are very good to fit the data.
  - The epidemiological models are very bad to make predictions.

$\hookrightarrow$ need for more data, different from the data collected today, in particular, *surveys from representative, random samples of the entire population*.

[1] S. Stringhini et al., Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study, The Lancet, June 11, 2020.