This file, the slides: https://www.math.leidenuniv.nl/~gill/Corona.pdf
Video of second part of talk (including discussion):
https://video.leidenuniv.nl/media/t/1_egjftqok

Corona statistics, Corona uncertainty
Fun and challenges with Corona statistics

Richard Gill, Thursday 7 May, 2020

Natuurwetenschappelijk Gezelschap Leiden

Contents

Corona statistics, Corona uncertainty

NRC, 21 March, 2020

A Fokke & Sukke cartoon from 10-30-2007 in the Dutch newspaper NRC Next. The text was kindly translated into English for us by the creators of the cartoon: Reid, Geleijnse and Van Tol. Lucia de Berk was still in prison at that time. The two ducks are defending a family guardian, accused of being responsible for the death of the girl Savanna, who died by suffocation. The accused woman was in fact acquitted (with another defence). What counselor Sukke is saying corresponds to what the law psychologist H. Elffers told the court: "Honoured court, this is no coincidence. The rest is up to you."

4

LIVESCIENCE

Live Science is supported by its audience. When you purchase through links on our site, we may earn a

Husband and wife poison themselves trying to self-medicate with chloroquine

By Kimberly Hickok - Reference Editor   March 24, 2020

An Arizona man is dead and his wife is hospitalized after both of them self-medicated with chloroquine.

f 🐦 🔴 📌 📮 ✉   💬 Comments (10)

An Arizona man is dead and his wife is hospitalized after both of them self-medicated with chloroquine phosphate, a chemical used to treat fish for parasites, in an effort to ward off the novel coronavirus.

The couple, both in their 60s, listened to President Donald Trump tout chloroquine, a decades-old antimalarial drug, as a very promising treatment for COVID-19 in a recent press conference. The woman, who asked not to be named, said she was familiar with the chemical because she used it to treat her koi fish.

"I saw it sitting on the back shelf and thought, 'Hey, isn't that the stuff they're talking about on TV?'" she told NBC News on Monday (March 23). "We were afraid of getting sick," she said.

So, the couple mixed a small amount of the parasite remover with a liquid before drinking the solution. Within 20 minutes, both of them became ill. The woman started vomiting and her husband experienced severe respiratory problems, NBC News reported. They called 911, and soon after arriving at the hospital the man died from cardiac arrest. The woman was initially in critical condition but is now stable and expected to fully recover.

This evening's menu

It's an ill wind …

2·The beginning: the SEIR model, R_0, the model of the RIVM-CIB

3·How many deaths will there be? Study of excess deaths?

4·Exponential, or a power-law?

5·A controversial treatment: Hydroxychloroquine & azithromycin

https://gabgoh.github.io/COVID/index.html

https://www.nemokennislink.nl/publicaties/relativerend-rekenen-aan-covid-19

https://doi.org/10.1016/j.ijantimicag.2020.105949

2. R_0 and all that, how does RIVM do it?

The SEIR model

Susceptible -> Exposed -> Infectious ->Removed

- "Susceptible": Not yet carrying the virus

- "Exposed": Infected, but not yet infectious ("incubation period")

- "Infectious": Can cause infections of susceptible persons

- "Removed": died, or recovered and became immune, or in permanent isolation; in any case, no longer infectious or susceptible

- We start with a few "exposed" individuals and many "susceptible". Per small time interval, each "exposed" has the same small chance to become "infectious", each "infectious" to become "removed", and each "infectious" has the same small chance to infect each "susceptible" person.

The SEIR model

Susceptible, Exposed, Infected, Removed

- This is a Markov process with state space the quadruples of positive integers

- One might start the process with all but one individual susceptible and one infected

- Wait and see... the epidemic might die out, might gradually grow till everyone is removed.

- The **expected values** at any time of number of individuals in any state follow a system of ordinary differential equations. The epidemic initially grows exponentially fast if R_0, the expected number of individuals infected by one exposed individual, is bigger than 1. We have a nice formula for R_0 in terms of the parameters of the model; and for the rate of exponential growth.

Epidemic Calculator

Day
218

☐ **Susceptible**
Population not immune to disease.
Σ 6,138,598 (87.69%)
Δ -76 / day

☑ **Exposed**
Population currently in incubation.
Σ 528 (0.01%)
Δ -26 / day

☑ **Infectious**
Number of infections *actively* circulating.
Σ 342 (0.00%)
Δ -17 / day

☐ **Removed**
Population no longer infectious due to isolation or immunity.
Σ 860,532 (12.29%)
Δ -196 / day

☐ **Recovered**
Full recoveries.
Σ 832,700 (11.90%)

☑ **Hospitalized**
Active hospitalizations.
Σ 9,007 (0.13%)
Δ -281 / day

☑ **Fatalities**
Deaths.
Σ 16,479 (0.24%)
Δ 25 / day

☑ linear scale

Intervention on day | to decrease transmission by
100
◁‖▷ (drag me) | 66.67%
⟵ $\mathcal{R}_0 = 2.20$ | $\mathcal{R}_t = 0.73$ ⟶

200,000

150,000

100,000

50,000

0

Day 0   20   40   60   80   100   120   140   160   180   200

First death

Peak: 74,738 hospitalizations

Transmission Dynamics

Population Inputs
Size of population.
7,000,000

Number of initial infections.
1

Basic Reproduction Number $\mathcal{R}_0$
Measure of contagiousness: the number of secondary infections each infected individual produces.
2.2

Transmission Times
Length of incubation period, T_{inc}.
5.20 days

Duration patient is infectious, T_{inf}.
2.9 Days

Clinical Dynamics

Mortality Statistics
Case fatality rate.
2.00 %

Time from end of incubation to death.
32 Days

Recovery Times
Length of hospital stay
28.6 Days

Recovery time for mild cases
11.1 Days

Care statistics
Hospitalization rate.
20.00 %

Time to hospitalization.
5 Days

11

The SEIR model

Susceptible, Exposed, Infected, Removed

This calculator implements a classical infectious disease model — **SEIR** (**S**usceptible → **E**xposed → **I**nfected → **R**emoved), an idealized model of spread still used in frontlines of research e.g. [Wu, et. al, Kucharski et. al]. The dynamics of this model are characterized by a set of four ordinary differential equations that correspond to the stages of the disease's progression:

$$\frac{dS}{dt} = -\frac{\mathcal{R}_t}{T_{\text{inf}}} \cdot IS, \qquad \frac{dE}{dt} = \frac{\mathcal{R}_t}{T_{\text{inf}}} \cdot IS - T_{\text{inc}}^{-1}E, \qquad \frac{dI}{dt} = T_{\text{inc}}^{-1}E - T_{\text{inf}}^{-1}I, \qquad \frac{dR}{dt} = T_{\text{inf}}^{-1}I$$

In addition to the transmission dynamics, this model allows the use of supplemental timing information to model the death rate and healthcare burden.

http://gabgoh.github.io/COVID/index.html

Model Details

The clinical dynamics in this model are an elaboration on SEIR that simulates the disease's progression at a higher resolution, subdividing I, R into *mild* (patients who recover without the need for hospitalization), *moderate* (patients who require hospitalization but survive) and *fatal* (patients who require hospitalization and do not survive). Each of these variables follows its own trajectory to the final outcome, and the sum of these compartments add up to the values predicted by SEIR. Please refer to the source code for details. Note that we assume, for simplicity, that all fatalities come from hospitals, and that all fatal cases are admitted to hospitals immediately after the infectious period.

The model of the RIVM-CIB

Refined (stratified) SEIR

- Split Netherlands into strata (**categories**) according to known demographics (age, region, sex, ...**??**).

- Solve system of very many ordinary differential equations

- Notice: assumption of homogeneity: an individual in a given **category** (when infectious) has a constant rate at which they infect an individual in another **category** (rates depend on "source" and "sink" category, but not on time)

- Many equations with many *parameters*

- The model is truly a *model*: a very highly simplified description of reality. It *has* shown its worth in past epidemics

- There are so many parameters that one can always guestimate values so that it fits to existing data of an ongoing epidemic. It will then make decent predictions with a horizon of a few days or two. But maybe you don't need any "model" to do that

- Problem: lack of data, inadequacy of model

- **Fortunately, small variations of parameters lead to very different predictions, hence one can show policy-makers that actually – we actually don't have much idea!**

R_0 is a "lie for children"

Reverse engineering

- One sees, in initial stages of the epidemic, an exponential growth of e.g. deaths of people **with** Corona [note: I don't write: deaths caused **by** Corona]

- That would mean a certain R_0, which depends also on some other parameters

- That's the number you tell the journalists and the parliament

- What you should mean by R_0 for more complex models is not so clear

- O Diekmann, J.A.P. Heesterbeek, J.A.J. Metz (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* **28** (4), 356–382.

Wikipedia: "In populations that are not homogeneous, the definition of R_0 is more subtle. The definition must account for the fact that a typical infected individual may not be an average individual." … "When calculated from mathematical models, particularly ordinary differential equations, what is often claimed to be R0 is, in fact, simply a threshold, not the average number of secondary infections. There are many methods used to derive such a threshold from a mathematical model, but few of them always give the true value of R0. This is particularly problematic if there are intermediate vectors between hosts, such as malaria."

3. Unruly data

Data problems
Cause of death on death-certificate is a "fiction"

- Sick people die *with* perhaps several contributing "conditions", of which a recent or ongoing Corona infection may be just one

- Hospitals report (quickly) deaths in ICUs of patients who (recently) tested positive for Corona as *Corona deaths*

- Nursing homes belatedly report deaths of persons who were already in a very bad condition anyway, without testing if those persons had Corona

- Tests have false positives, false negatives

- Tests change in time; reporting policies change in time

- Data repeatedly shows a huge dip in number of deaths over a weekend!

- In ICU you only die when doctors decide to switch off life support systems

Data problems, model problems

Events in the SEIR model are not observed, the model is a fiction, anyway

- Hospital workers are continuously exposed to Corona virus coming from many patients in their care; they are also overworked and overstressed

- They infect persons in their households and they infect persons in the transport systems which they use

- We do not know the time of becoming "exposed" and we do not know the time of becoming "infectious". The time of being "removed" is also unclear. There are very many "asymptomatic" cases and not much testing of people who are not sick, so you don't know how many "infectious" there are around

- Meanwhile, there are care homes and nursing homes ... family come visiting ...

The good news, perhaps

Admit the uncertainties

- Even if one uses a wrong model, if one is honest about the uncertainty in its parameters one will see that actually, one cannot predict very far into the future, anyway

The Observer
Coronavirus outbreak

Coronavirus statistics: what can we trust and what should we ignore?

The flurry of figures, graphs and projections surrounding the pandemic is confusing. Two experts guide us through the maze
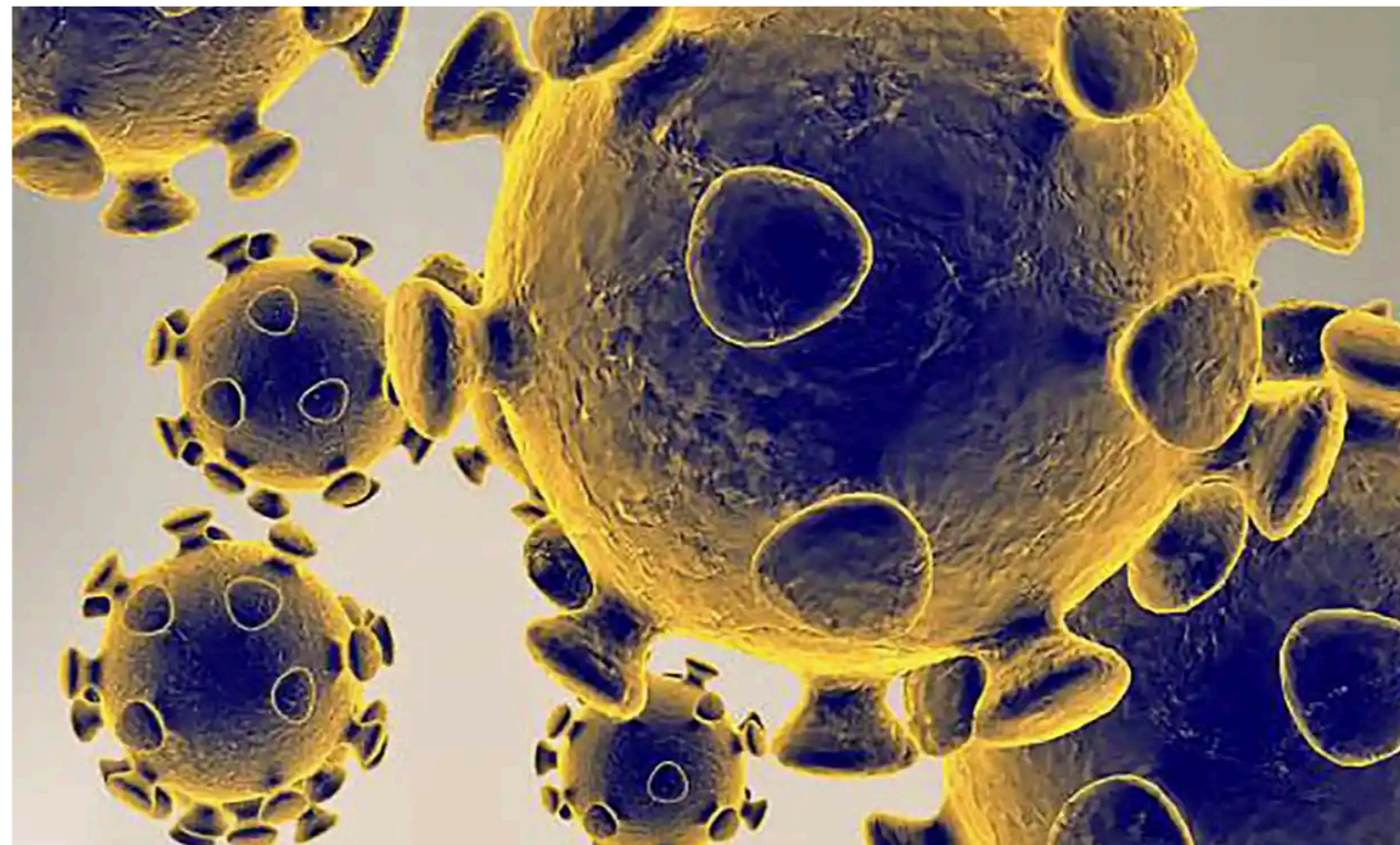
- **Coronavirus – latest updates**
- **See all our coronavirus coverage**

Sylvia Richardson *and* **David Spiegelhalter**

Sun 12 Apr 2020 08.16 BST

f 🐦 ✉

1,867

▲ The coronavirus, Covid-19. Photograph: US Food and Drug Administration/AFP via Getty Images

The past few weeks has seen an unstoppable epidemic … of statistics. The flood threatens to overwhelm us all, but what do all these numbers mean? Here are eight statistics you may see, with some warnings about how much we might trust them.

19

Sylvia Richardson *and* **David Spiegelhalter**

Sun 12 Apr 2020 08.16 BST

7 The lethal risks of being infected These vary dramatically with age and frailty, just as "normal" risks do. In fact the current estimates for the general public (rather than healthcare workers) seem remarkably similar to the risks we face anyway each year – but all packed into a few weeks.
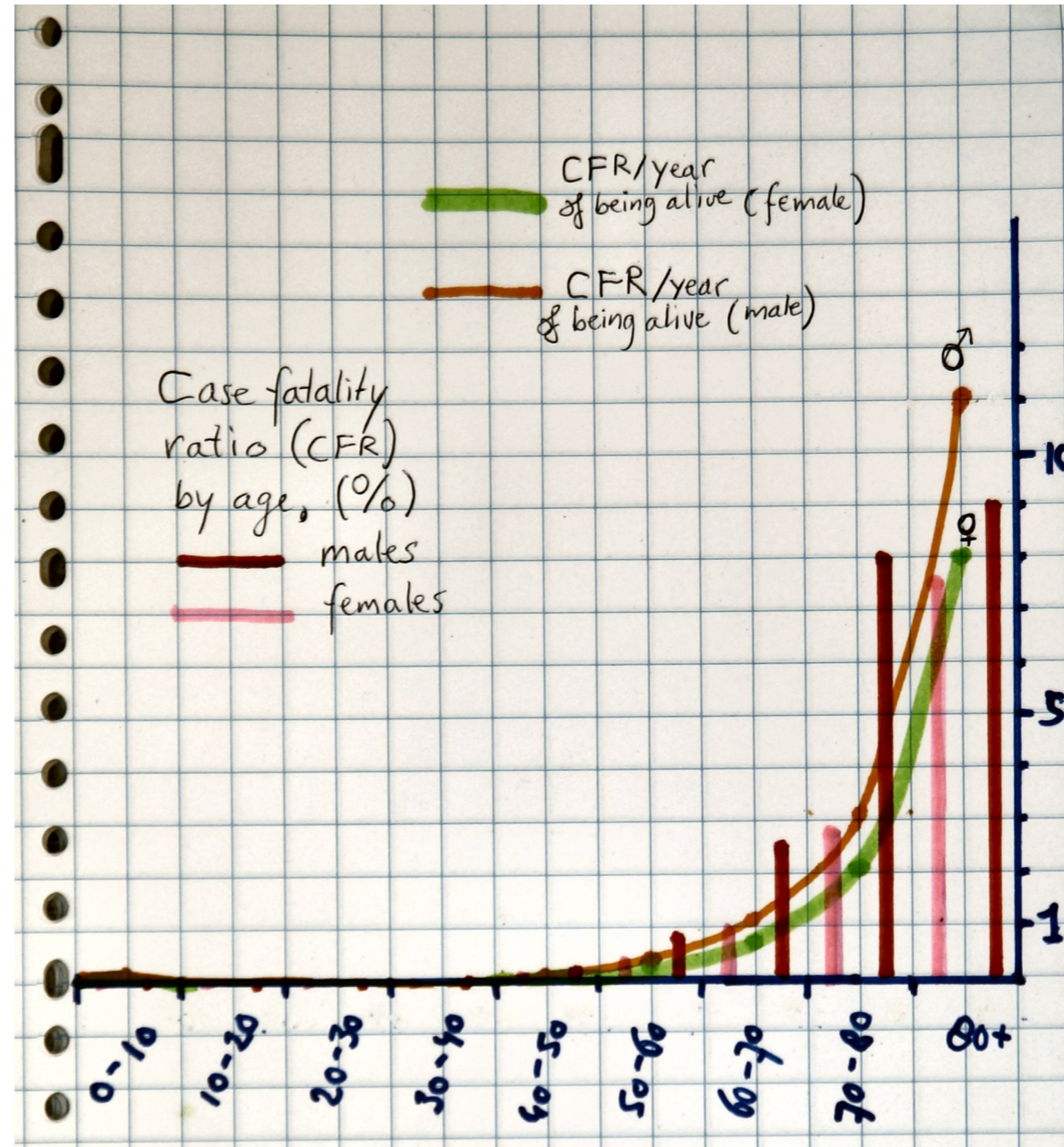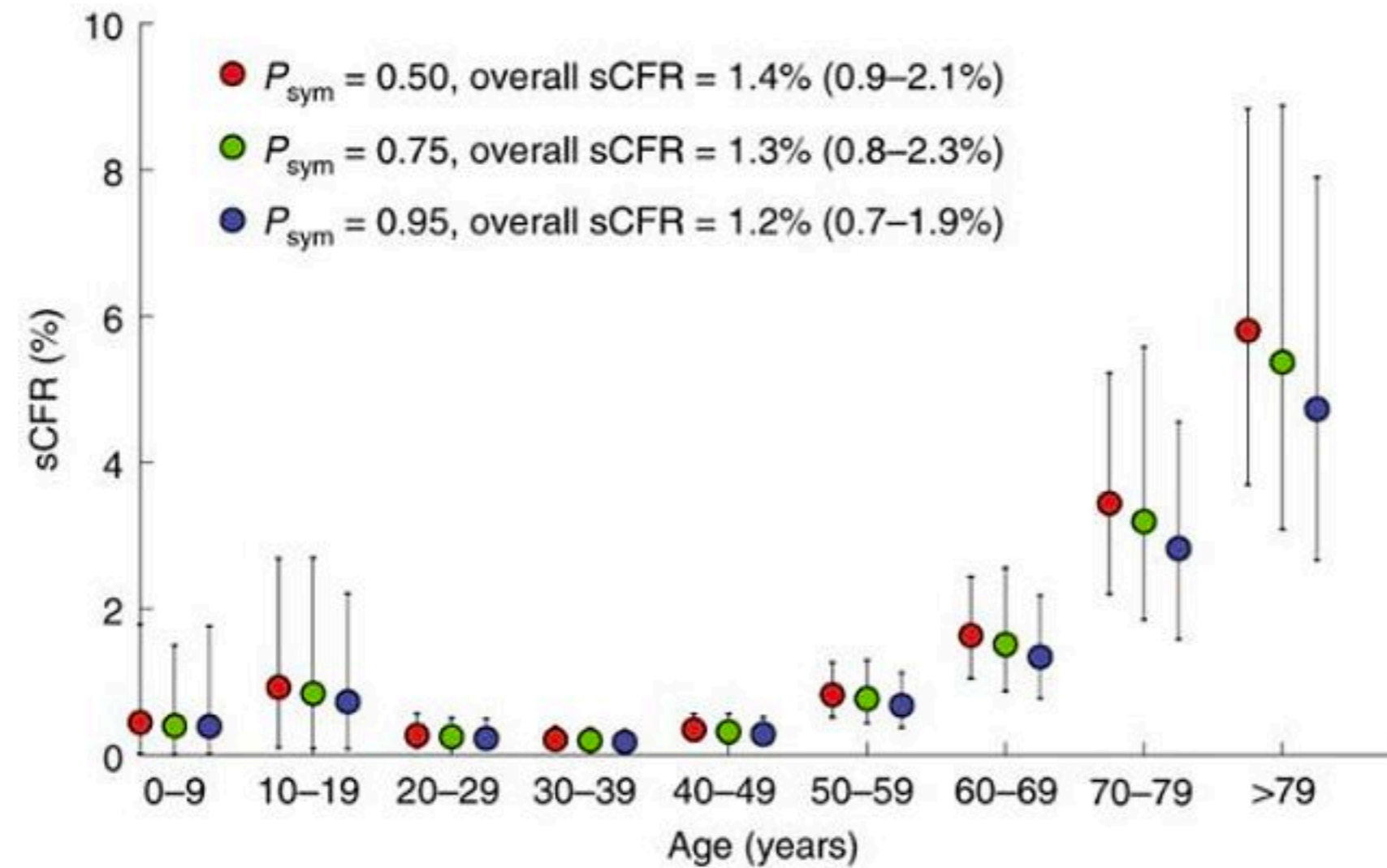

Illustration: Arnout Jaspers, a month earlier

Het aantal sterfgevallen in Wuhan als percentage van het aantal gediagnosticeerde ziektegevallen, per leeftijdsgroep. Mensen tot 40 jaar oud lopen vrijwel geen gevaar te overlijden aan Covid-19. De kleine bult in de grafiek bij mensen jonger dan twintig stelt weinig voor, omdat in die leeftijdscategorie bijna niemand ziek wordt. De drie kleuren staan voor verschillende aannames over het percentage mensen dat geïnfecteerd raakt en ook ziektesymptomen krijgt.

Wu e.a., Nature Medicine

https://www.nemokennislink.nl/publicaties/relativerend-rekenen-aan-covid-19

De relatieve kans om ziek te worden in Wuhan per leeftijdscategorie, waarbij de categorie 30-39 per definitie relatieve kans '1' heeft. Voor mensen jonger dan 20 is de kans vrijwel nul. De drie kleuren staan voor verschillende aannames over het percentage mensen dat geïnfecteerd raakt en ook ziektesymptomen krijgt.

Wu e.a., Nature Medicine
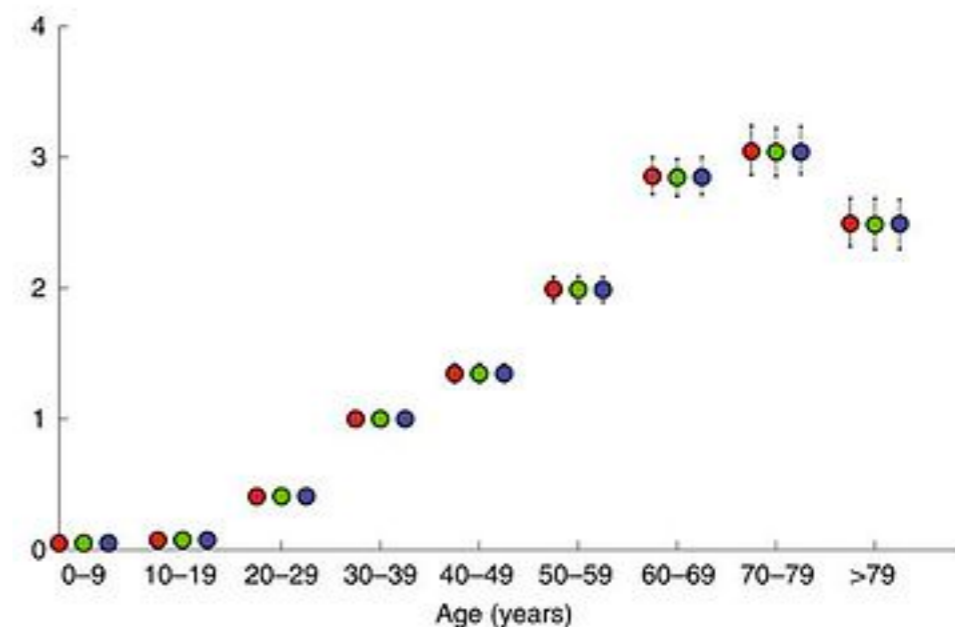
Annual risk of death in the **UK**

Annual risk of death (1 in X) by age and sex
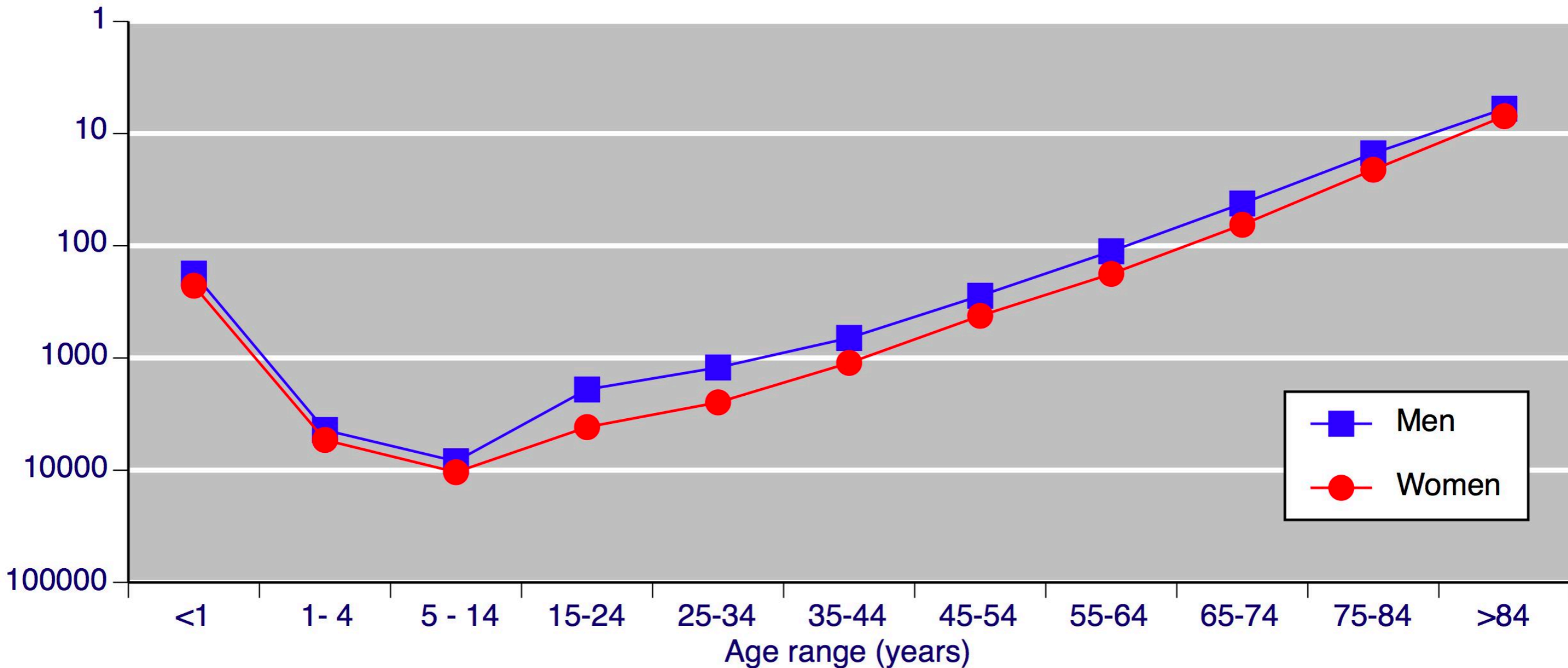
Mortality statistics 2005. National Statistics Series DH2 no.32 (www.statistics.gov.uk)

I conclude that "your" Corona risk this year is very similar,
at least for women, maybe worse for men, to "your" annual risk "without Corona"
Note: this is *my* guess of *your* risk when I only know your age and sex.
It isn't really *your* risk at all. Do you smoke? Live in a very polluted area? Have
some heart problems?

22

After the epidemic has hit

Statistics of what we observe

- Compare death rates by (e.g.) age and sex and month, with average death rates in last couple of years

- Conclusion: this year, everyone's chance of dying is indeed roughly doubled

- Equivalently – roughly, we all lose one year of life

- **Proviso**: Covid-19 doesn't mutate to something worse; and you can't get sick of it several times

4. A different approach

Is it really exponential?

Messages from statistical mechanics and from computer science

- Physicists studying interacting particle systems, and computer scientists studying (e.g.) spread of computer virus's in computer networks, know that *heterogeneity* can dramatically change the general picture

- Real human networks are almost fractal in nature, heterogenous in space and in "level"

- Cf. also "black swans" theory

- Consequences: long range dependence, heavy-tailed distributions

- The actual behaviour is qualitatively different from what the behaviour would have been with all parameters constant (e.g., put equal to their mean values)

- The good news: the epidemic spreads according to a power law, not exponentially

- We should expect to see straight lines in **double logarithmic** plots [both axes in log scale] not in **semi-logarithmic** plots [Y-axis in log scale]

- The bad news: we can expect to see randomly occurring new local outbreaks ("hotspots")

5. Chloroquine + antibiotic
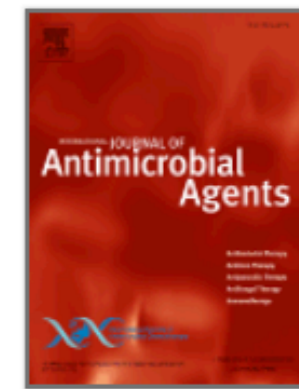
Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial

Philippe Gautret [a, b, $], Jean-Christophe Lagier [a, c, $], Philippe Parola [a, b], Van Thuan Hoang [a, b, d], Line Meddeb [a], Morgane Mailhe [a], Barbara Doudier [a], Johan Courjon [e, f, g], Valérie Giordanengo [h], Vera Esteves Vieira [a], Hervé Tissot Dupont [a, c], Stéphane Honoré [i, j], Philippe Colson [a, c], Eric Chabrière [a, c], Bernard La Scola [a, c], Jean-Marc Rolain [a, c], Philippe Brouqui [a, c], Didier Raoult [a, c] ⚲ ✉

⊞ **Show more**

Get rights and content

The Marseilles trial

Facts

- Didier Raoult (last author) is a famous & controversial figure with many enemies

- In Raoult's clinic, 26 patients were all treated with hydroxychloroquinine, a few also with azithromycin (on top of standard care procedures)

- **hydroxychloroquinine**: a very common, cheap, anti-malarial drug, few known side effects (on healthy people)

- **azithromycin**: a common (cheap, generic) antibiotic

- In several other clinics in the neighbourhood (Nice, Aix-en-Provence?) 16 patients were treated in a standard way

- This was the very beginning of the Corona epidemic in France, these were the very first patients who turned up at those hospitals suspected of having Corona

- In Wuhan, Chinese researchers already reported success with this treatment for patients in early stages of the illness

The Marseilles trial

Issues

- The Marseilles study reports success rates for two treatments. Individual patient data seems not to be available. There is anecdotal evidence that the two groups are comparable (e.g. mean age is about the same). The "Marseilles group" (treatment group, group B) does much, much better than the "Nice group" (control group, group A).

- Two patients are dropped from the Marseilles group "because they were transferred to ICU within 3 days so did not get 3 days of the treatment"!!! One died soon after!

- One patient was dropped from the Marseilles group because they just walked out of the hospital after 1 day!

- There is no statistician in the long list of authors. There is one epidemiologist: a very young Vietnamese researcher

- The treatment killed two patients in Brazil [admittedly, they accidentally gave 100x the recommended dose]

- The treatment made a lot of already seriously ill US veterans even more seriously ill, in a double blind randomised trial

- People are calling for criminal measures against Prof. Roualt and his team ...

The Marseilles trial

Recent developments

- With Leila Schneps (a pure mathematician from Paris) and with a Chinese student from Boston University, we wrote to Philippe Gautret and received fairly complete data of the 16 + 26 patients (i.e., including those dropped from the analysis in the published paper)

- I have performed an "intention to treat" analysis, using as target variable "Covid-19 infection has gone (according to daily PCR tests), patient still alive, not in ICU, on Day 6"

- I did both a frequentist and a Bayesian analyis. For the Bayesian analysis, I used a "slab and spike" prior.

- I did logistic regression using age and sex as covariates, intend to add also measures of "severity of existing conditions" and "severity of infection on admission"

The Marseilles trial

Recent developments

- The Marseilles treatment is deliberately engineered to prevent the rapid Covid-19 infestation in an infected person which often occurs *before* they are particularly sick (or even before they know they are sick at all)

- We know that this period can be fairly long and that such persons can be highly infectious before feeling particularly ill, and sometimes never actually get ill

- The infection is typically "beaten" by the patient's own immune response which, for some patients, goes on to start destroying the patient's own organs! (Many patients in ICU are actually already free of Corona virus).

The Marseilles trial

Histogram of AGE
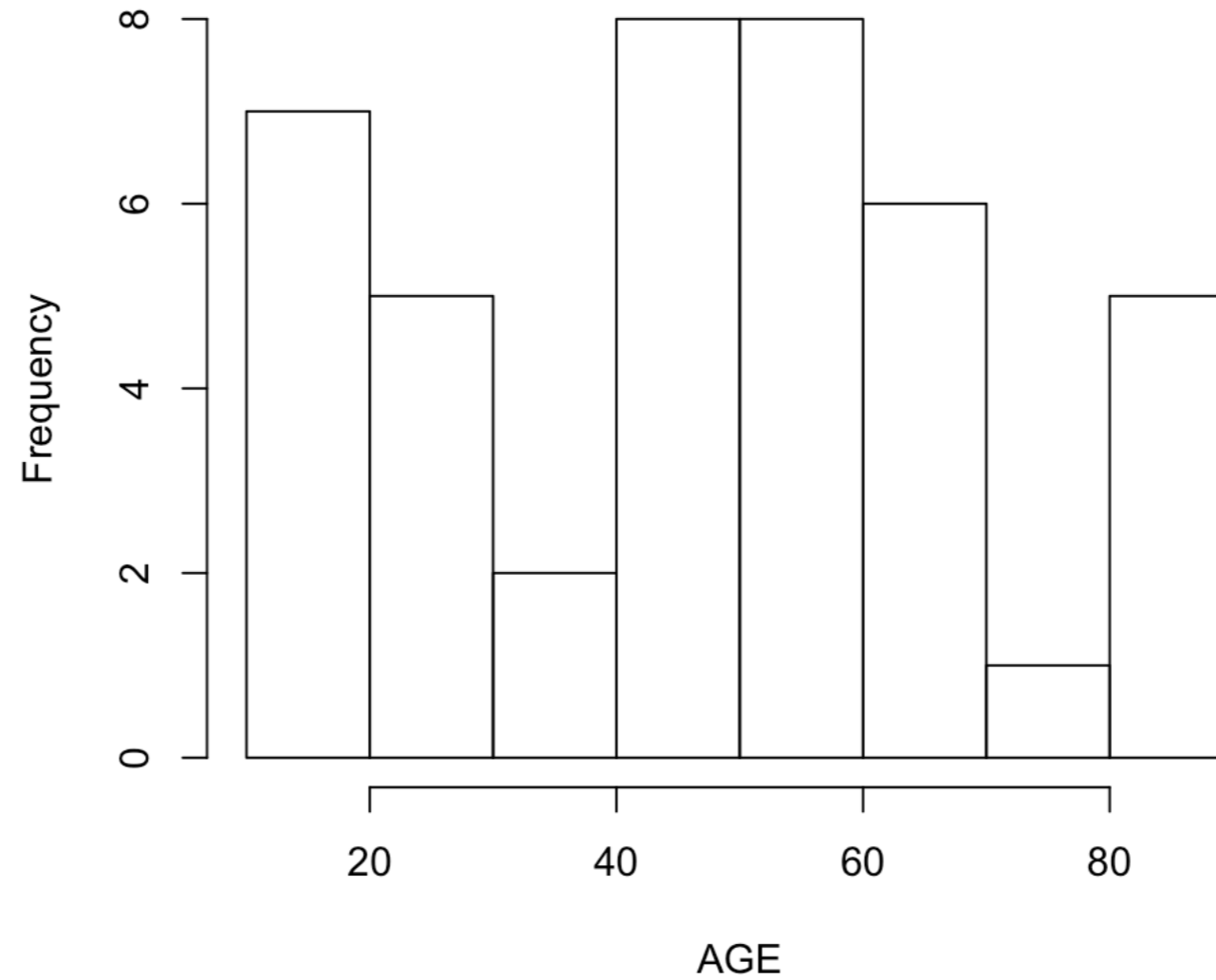
```
attach(data)

FINAL <- PCR_J6
(1:42)[is.na(FINAL)]
```

```
## [1] 40 42
```

```
FINAL[42] <- 1
FINAL[40] <- 0
sum(FINAL == 0)
```

```
## [1] 18
```

```
sum(FINAL == 1)
```

```
## [1] 24
```

```
summary(glm(FINAL ~ PLQ + AGE + SEXE, family = binomial))
```

Standard (frequentist) logistic regression

```
##
## Call:
## glm(formula = FINAL ~ PLQ + AGE + SEXE, family = binomial)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -1.7484   -0.8997    0.3589    0.8098    1.8038
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.76465    0.99688   0.767   0.44306
## PLQ         -2.84540    0.98960  -2.875   0.00404 **
## AGE          0.03064    0.02015   1.521   0.12824
## SEXE        -0.33862    0.75776  -0.447   0.65497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 57.364  on 41  degrees of freedom
## Residual deviance: 44.524  on 38  degrees of freedom
## AIC: 52.524
##
## Number of Fisher Scoring iterations: 4
```

The Marseilles trial

```
summary(samp)
```

```
##
## Iterations = 11001:31000
## Thinning interval = 1
## Number of chains = 3
## Sample size per chain = 20000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##             Mean       SD   Naive SE Time-series SE
## beta[1]   0.74459 0.97379 3.975e-03      0.0161000      Intercept
## beta[2]  -2.90447 0.96533 3.941e-03      0.0130108      Treatment
## beta[3]   0.03232 0.01985 8.104e-05      0.0003565      Age
## beta[4]  -0.33841 0.74543 3.043e-03      0.0075452      Sex
##
## 2. Quantiles for each variable:
##
##                2.5%      25%      50%      75%     97.5%
## beta[1]   -1.074023  0.08410  0.71068  1.36940   2.75742
## beta[2]   -4.945215 -3.52125 -2.85295 -2.23630  -1.15479
## beta[3]   -0.005019  0.01878  0.03185  0.04537   0.07267
## beta[4]   -1.801855 -0.83909 -0.33942  0.16513   1.11858
```

Bayesian logistic regression, MCMC & BUGS, RJAGS

Standard weakly informative prior - each beta is N(0, 10)

Conclusions

More work to do

- In my opinion, Covid-19 is a "warning shot"

- I hope we will grasp the opportunity to right some wrongs in society (in our societies, world-wide)

- At least, next time, our epidemic models and our statistical models will be better. Lots of new communication lines have been opened up. Exciting new research is going on now, both in the modelling of epidemics, in the statistical analysis of relevant data, in quantifying uncertainty and evaluation of policy choices, in the communication of science to the public