Federico Lois

Twitter: @federicolois

Techies in Virusland

What is a performance guy doing in Virusland?



FOR SOME REASON, I FEEL A POWERFUL COMPULSION TO OWN ANY DEVICE WHOSE NAME ENDS IN "-OMETER".

If the only thing you know how to do is measure...

Guess what!!!

Source: https://xkcd.com/2060/

Epidemiology

is the study and analysis of the distribution, patterns and determinants of health and disease conditions in defined populations....

You are right!!! MEASUREMENTS, MEASUREMENTS, MEASUREMENTS

What are we going to talk about today?

- Epidemic behavior and modeling of infectious diseases.
 - A primer on SARS-Cov-2 epidemic behavior.
 - Some outcomes of our work with Levan Djaparitze.
- Why forecasting is CRAZY hard
 - Why we fail consistently to forecast.
 - What can we do about it.

Part 1: The disclaimer

HARDCORE

Not Applicable Sensitive Topic Forecasting is [H]ARD!!!

DOTNEXT CONTENT RATING

Part 2: The fundamentals

A primer in evaluation of evidence...



Not all evidence is created equal.

A primer in evaluation of evidence...



Let's talk about models Shall we?

A model is a group of equations, where you can choose what it is data and what is unknown.

- Levan Djaparitze

Models, models, models...

- Models are abstractions of processes.
- How do we know ...
 - ... we are modeling the right thing?
 - ... we aren't overfitting?
 - ... our theories are correct?
 - ... our calibration is correct?

Are we overfitting?



- Classic example.
- The problem with complex systems?
 - Overfitting is subtle.
 - Too many free parameters.
 - Solutions are usually non-linear
 - More on this later.
 - You will not know it until it is too late.
 - Induction problem.

Are we modeling the right thing?



- Huge problem in modeling.
- The problem with complex systems?
 - Is the simplification able to describe the system?
 - Can we measure the difference between predicted and measured values?
 - Are measurements:
 - trustworthy, unbiased, representative,
 - free of systematic observational error?

Is our calibration correct?



- Huge problem in epidemiology.
- When we measure, what are we measuring, REALLY!!! For ex,
 - R* (Reproduction number).
 - IFR or CFR
 - Cases or deaths
- This alone can break a model without us even knowing it.



All models are wrong, some are useful. The important part is that we forget that the rest are just plain garbage.

Traducir Tweet

4:06 p. m. · 5 abr. 2021 · Twitter Web App

+++

Modeling SARS-Cov-2

- Lots of questions, not many clear-cut answers:
 - ... are asymptomatic contagious?
 - ... what is the mode of transmission?
 - ... what is the spread factor (R knot)?
 - ... what is the lethality by age?
 - ... are there going to be second waves*
 - ... is seroprevalence a good measurement?
 - ... is the evidence any good?
 - ... does an optimal response exist? *

* We asked ourselves these questions in July and responded it by the 13th of October.

When modeling complex systems induction just fails...



What most scientist do not realize is: "Science is about proving yourself wrong, not right."

+ + +

Traducir Tweet

2:11 p. m. · 5 abr. 2021 · Twitter Web App

Modus Tollens to the rescue

Your most important tool for modeling processes.

 $\frac{P \rightarrow Q, \neg Q}{\therefore \neg P}$ From Latin: "mode that by denying denies"

Your most important tool for modeling processes.

P implies Q



Your most important tool for modeling processes.

$P \rightarrow Q, \neg Q$ P implies Qbut it is the case of not Q

 $\therefore \neg P$

Your most important tool for modeling processes.

 $P \rightarrow Q, \neg Q$

P implies Q but it is the case of not Q then we can conclude not P

What does this mean?

YOU ALWAYS HAVE



THE HIGH GROUND

You need just a single counter-example

How that works?

Let's say I have a theory

Claim: Obesity is a major factor in deaths of SARS-Cov-2

Evidence

Czech Republic: Obesity rate of 26% with 2603 deaths per million Switzerland: Obesity rate of 19.5% with 1201 deaths per million Argentina: Obesity rate of 31% with 1269 deaths per million Norway: Obesity rate of 23.1% with 126 deaths per million Egypt: Obesity rate of 32% with 120 deaths per million

Let's say I have a theory

Claim: Obesity is a major factor in deaths of SARS-Cov-2

Evidence

Egypt: Obesity rate of 32% with 120 deaths per million

 $P \rightarrow Q, \neg Q$ $\therefore \neg P$

What can we say about my pet theory?

We can conclude without risking to be wrong It's not correct

BUT!!! There is always a BUT

As always, the devil is in the details!!!

My pet theory may be:

- ... just plain wrong!!! *Usually, the most likely case.*
- ... unable to explain variance (a confounder)
- ... right, but only at a second or third order contribution.
- ... incomplete (**P** is missing clauses)
- ... underspecified (**Q** is ambiguous)
- ... (P, Q) are subjected to systematic observational error

Systematic Observational Error

Say you have a surveillance system:

- ... you do "randomized population sampling"
 - You test 2000 cases in the lab.
- ... you have 10+ years of data with positivity around 10%

Say some year you have:

- ... an abnormal spike in cases (say 10 times more)
 - Normally: 100.000 cases
 - This year: 1.000.000 cases (10x)
- ... you measure lab confirmed positivity
 - and it is compatible with your history.

If from 100,000 cases we send 2,000 to the lab, and 10% of the cases are influenza, then 10,000 cases are expected to be influenza, and 90,000 are expected to be untypified influenza-like diseases. Clearly no abnormality there, as positivity is compatible with history.

If from 1,000,000 cases we send 2000 to the lab, and 10% of the cases are influenza, then 100,000 cases are expected to be influenza, and 900,000 are expected to be untypified influenza-like diseases. Is this year compatible with history?

Can we conclude there is NO abnormality there?


Part 3: Modeling SARS-Cov-2

SARS-CoV-2 waves in Europe: A 2-stratum SEIRS model solution

By Levan Djaparidze and Federico Lois

Some important definitions (the watch-later glossary)

Definitions

SARS-Cov-2: RNA Virus from the Betacoronavirus genus.

COVID-19: The group of signs and symptoms which define the disease caused by SARS-Cov-2 **PCR+:** Individual with samples positive for RNA sequences of SARS-Cov-2 by Polymerase Chain Reaction

Ab+: Positive to antibodies specific to SARS-Cov-2 (IGg/IGm)

Humoral Immunity: Immunity conferred by antibody mediated response

Cellular Immunity: Immunity conferred by T-cells mediated response.

Seroprevalence: Prevalence of Ab+ at a population level

Infected: An individual carrying SARS-Cov-2 during the infectious period.

Case: A PCR+ individual detected or suspected by epidemiological criteria of carrying SARS-Cov-2 during the period it is thought to be infective.

Ro (Basic Reproduction Number): expected number of **new** infected directly generated by a single infected in the population

Definitions

Symptomatic: A PCR+ individual with inequivocal symptoms for COVID-19 **Presymptomatic:** A PCR+ individual which will be symptomatic in the future on a reasonable time period (5 to 10 days).

Mild-symptomatic: A PCR+ individual with non-inequivocal symptoms for COVID-19 **Asymptomatic:** A PCR+ individual without symptoms for COVID-19.

Suspected Death: An individual death which meets the definition of symptoms for COVID-19 **Confirmed Death:** A PCR+ individual death (no distinction) **Clinically Confirmed Death:** A symptomatic individual death

NPIs: Non-Pharmaceutical Interventions like social distancing, closed schools, masks, etc.

Definitions

Cohort: A population segment with a specific characteristic.

Compartmental models: Simplified mathematical models of epidemic behavior where the order of the letters shows the flow pattern between compartments.

Homogeneous models: Models in which individuals share the same parameters along the population in the cohort.

Heterogeneous models: Models in which individuals may have different parameters along the population in a cohort.

SIR Model: Susceptible-Infected-Recovered model [Kermack & McKendrick, 1927]. **SEIRS Model:** Susceptible-Exposed-Infected-Recovered-Susceptible model (SEIR with temporary immunity becoming susceptible again).

ABMs: Agent-based models where simulation can be anything, from an individual to an organization, or even a country where each cell is an individual agent interacting at different levels of details with other agents.

2-stratum SEIRS model

- 2 parallel feed-forward compartmental SEIRS models.
- Age stratified with 2 cohorts
 - Healthy under 60, Vulnerable (>60 or not healthy)
- Models explicitly NPIs through averaging (isolation).
 - For the math inclined, similar idea to mean-field theory.
- Both locations and viral parameters are fixed.
- Objectives:
 - Estimate total immunity on naïve populations.
 - Do they have enough susceptible to fuel another wave?
 - Disambiguate epidemic behavior
 - Is dynamic at Stockholm equal to Madrid's?



S + E + I + R = N, when N is constant for us

> $\beta = transmission rate$ $\sigma = latency$ $\gamma = recovery rate$



$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$





$$\frac{dE}{dt} = \frac{\beta IS}{N} - \sigma E$$





$$\frac{dI}{dt} = \sigma E - \gamma I$$





$$\frac{d\mathbf{R}}{dt} = \gamma I$$

2 stratum - SEIR model





A model is a group of equations, where you can choose what it is data and what is unknown.

- Levan Djaparitze

2 stratum – THE DATA

- Viral parameters are data (not unknowns).
 - Ro = 3.3, Do = 2 days, Eo= 5 days,
 - IFR_vul = 0.92%, IFR_non_vul = 0.0035%,
 - P_non_vul_una = 7%, etc.
- Location parameters are data (not unknowns).
 - Population, Population at risk, etc
- Reported daily deaths
- Seroprevalence ratio
 - This one is huge... more on this later

THE UNKNOWN: Total number of infected individuals from a fully naïve population

Can we forecast now?

The free parameters

- The model was designed to be fitted
 - Why fitting?
 - What data should we fit?
 - Avoid free parameters like the plague.
 - Aren't virus and location parameters, free parameters?
- 2-stratum free parameters
 - Average NPI level (isolation) for the vulnerable
 - Average NPI level (isolation) for the healthy

DEMO: Fitting Madrid



Healthy <60	Vulnerable	Fitted Deaths	
0.95	0.183	8391	
0.9	0.297 8391		
0.85	0.404 8391		
0.8	0.501	8391	
0.78	0.537	8391	
0.76	0.569	8391	
0.748	0.588	8391	
0.72	0.628	8391	
0.7	0.655	8391	
0.65	0.716	8391	
0.6	0.766	8391	
0.55	0.808	8391	
0.5	0.841	8391	
0.4	0.89	8391	
0.3	0.92	8391	
0.2	0.939	8391	
0.1	0.951	8391	

Let's look at what happen when we visually inspect them.



Healthy <60: 0.5 Vulnerable: 0.841 Healthy <60: 0.748 Vulnerable: 0.588 Healthy <60: 0.3 Vulnerable: 0.92

Are we there yet?



Healthy <60: 0.3 Vulnerable: 0.92



Healthy <60: 0.748 Vulnerable: 0.588

Final deaths: 13085



Healthy <60: 0.3 Vulnerable: 0.92

Final deaths: 16583



Healthy <60: 0.748 Vulnerable: 0.588



Healthy <60	Vulnerable	Vulnerable Final Deaths	
0.95	0.183	17440	
0.9	0.297	17260	
0.85	0.404	17065	
0.8	0.501	16843	
0.78	0.537	16746	
0.76	0.569	16647	
0.748	0.588	16583	
0.72	0.628	16430	
0.7	0.655	16310	
0.65	0.716	15971	
0.6	0.766	15569	
0.55	0.808	15086	
0.5	0.841	14520	
0.4	0.89	13085	
0.3	0.92	11304	
0.2	0.939	8987	
0.1	0.951	8400	

Fitting curve of acceptable solutions



Healthy <60	Vulnerable	Seroprevalence Ratio	Final Deaths
0.95	0.183	0.51	17440
0.9	0.297	0.58	17260
0.85	0.404	0.65	17065
0.8	0.501	0.72	16843
0.78	0.537	0.75	16746
0.76	0.569	0.77	16647
0.748	0.588	0.79	16583
0.72	0.628	0.83	16430
0.7	0.655	0.86	16310
0.65	0.716	0.93	15971
0.6	0.766	1	15569
0.55	0.808	1.07	15086
0.5	0.841	1.15	14520
0.4	0.89	1.29	13085
0.3	0.92	1.42	11304
0.2	0.939	1.53	8987
0.1	0.951	1.61	8400

If you don't fight back. It gets worse and worse.

Are we there yet?

Why overdetermination is important?



Stockholm



Madrid

Why overdetermination is important?



Stockholm

Seroprevalence Ratio: 1.7



Madrid Seroprevalence Ratio: 0.79

Why overdetermination is important?



Stockholm Seroprevalence Ratio: 1.7



Madrid Seroprevalence Ratio: 0.79

Are we there yet?

Part 4: Forecasting

SARS-CoV-2 waves in Europe: A 2-stratum SEIRS model solution

By Levan Djaparidze and Federico Lois

THE UNKNOWN: Total number of infected individuals from a fully naïve population

THE UNKNOWN: Immunity Estimation
Immunity Estimation

- Getting the right data is key.
 - Cases is a bad choice; date of reporting is just awful.
 - Confirmed Deaths by date of death is the most reliable.
 - High quality seroprevalence study:
 - Randomized sampling (representative),
 - Age stratified to recover the seroprevalence ratio
 - Find viral parameters that can explain all locations. (HARD)
 - Evaluating scientific literature is important.
 - Tollendo Tollens ALL THE WAY
 - For ex. Ro=3.3 can explain *all first waves.*

Immunity Estimation (Our results)

- Immunity estimation
 - Madrid: 41%
 - Catalonia: 23%
 - Paris*: 23%
 - London*: 33%
 - Brussels: 49%
 - Stockholm: 62% [1575

[Deaths per million] *if normal life*: [2461 / Million] [2766 / Million] [2297 / Million] [1930 / Million] [1814 / Million] [1575 / Million]

No location can return to normal life without having a second wave

Why use the seroprevalence ratio?



Source: https://twitter.com/dobssi

And Herd Immunity Threshold (HIT) is 10% higher now!!!!

Immunity Estimation (Our results)

Location	At July 4 th , 2020	Predicted	Actual
Madrid	1259	2461	2211
Catalonia	734	2766	1776
Brussels	854	1814	2056
Stockholm	968	1575	1715
London	760	1930	1725
Paris	620	2297	1731

What if the virus changes? The UK Variant (30% more transmissible) Ro=4.3

Immunity Estimation (UK Variant)

Location	At July 4 th , 2020	Predicted	Actual
Madrid	1259	2507 (+46)	2211
Catalonia	734	2896 (+130)	1776
Brussels	854	2042 (+228)	2056
Stockholm	968	2050 (+475)	1715
London	760	2046 (+116)	1725
Paris	620	2406 (+109)	1731

What if a virus doesn't kill anyone? The harmless virus IFR=0

[We know that is not true here, but play along]



Surprise!!!

- Madrid population
- No mitigation
- 404 deaths per million
- Seroprevalence Ratio: 0.88



- Madrid population
- Pandemic mitigation + Normal Life
- 368 deaths per million
- Seroprevalence Ratio: 0.79

More Surprises!!!

What is the probability of dying while positive from a virus that does not kill you?

P(dying | positive) P(positive)

Background mortality as a confounder

- There are 2 ways one can be positive at time of death.
 - A PCR+ diagnostic up to x days before death.
 - Viral shedding will cause PCR- to be delayed.
- We averaged both as window of positivity
 - We use 17 days; some countries use 28 days for the stats.

Sanity checks

If your model doesn't show deaths, you are doing it wrong.

15% of SARS-Cov-2 positive deaths are 'with' the virus

P(dying | other_cause) * P (dying | positive)

Östergötland, Sweden

(245 deaths)

- 60% contributing factor
- 25% dominant cause
- 15% unrelated ('with')

(148 deaths)(61 deaths)(36 deaths)

If natural mortality can impact the observations. Can it impact the immunity estimation?

How sensitive is the model to measurement error?

Sensitivity Estimation

	Variable values	Immunity value	Sensitivity
Variable name	Point (Min : Max)	Point (Min: Max)	(multiplier)
Ro	3.3 (2.5 : 4.0)	41% (41%:41%)	≈0
Do	2 (1.33:5)	41% (41% : 41%)	≈ 0
Eo	5 (4 : 6)	41% (41% : 41%)	≈ 0
RtoD	11 (9 : 13)	41% (41%:41%)	≈ 0
So	1.00 (0.9 : NA)	41% (51% : NA)	-2.44
IFR_vul	0.0092 (0.0077 : 0.0104)	41% (48% : 37%)	-0.85
IFR_non_vul	0.000035 (0.000029:0.000041)	41% (41% : 41%)	≈ 0
P_vul_u60	0.0342 (0.0274 : 0.0401)	41% (42% : 41%)	-0.06
P_non_vul_una	0.07 (0.05 : 0.09)	41% (41% : 41%)	≈0
T1	180 (150 : 210)	41% (41% : 41%)	≈0
PRLI	0.00 (NA: 0.50)	41% (41% : 41%)	≈0
Ma	7(1:14)	41% (41% : 41%)	≈ 0
A_060	0.21 (0.17:0.25)	41% (42% : 41%)	-0.06
A_u60	0.52 (0.42:0.62)	41% (40%:43%)	+0.21
TAK	0.35 (0.29:0.42)	41% (41% : 42%)	+0.07
APTP	17 (13 : 21)	41% (43% : 40%)	-0.17
ICU_pd	0.45	NA	NA
ICU_h_dur	10 days	NA	NA
vEff	0.77	NA	NA

Table A.1. Madrid immunity level estimation (Recovered/pop on July 2020)

sensitivity for virus parameters.

Fitted Immunity Estimation behaves differently than most expect.

Sensitivity Estimation

	Variable values	Immunity value	Sensitivity	
Variable name	Point (Min : Max)	Point (Min:Max)	(multiplier)	
Рор	6.662M (5.33M : 8M)	41% (51% : 34%)	-1.23	
P_060	0.233 (0.186:0.279)	41% (49% : 35%)	-1.00	
con_oD	(*0.8 : *1.2)	41% (33% : 49%)	+0.98	
sero_day	217 (-21 days : +21 days)	41% (41% : 41%)	≈ 0	
sero_u60_o60	0.79 (0.66 : 0.95)	41% (36% : 47%)	+0.75	
lso_1_real	99 (-5 days : +5 days)	41% (41% : 41%)	≈0	
lso_1_dur	102 (-21 days : +21 days)	41% (41% : 41%)	≈0	
PYDR_vul	0.031 (0.025 : 0.038)	41% (42%:40%)	-0.14	
PYDR_non_vul	0.0014 (0.001 : 0.0018)	41% (41%:41%)	≈0	

Table A.2. Madrid immunity level estimation (Recovered/pop) sensitivity for location parameters.

Fitted Immunity Estimation behaves differently than most expect.

Playing "What-if"



Madrid population Second wave with normal life Seroprevalence Ratio: 0.79 2487 / Million

Playing "What-if"



Madrid population Sweden like mitigation Seroprevalence Ratio: 1.55 1769 / Million

If we can estimate 'What-if' based on evidence. Can we find optimal strategies?

Death Minimizing

- Back in March we can find the optimal strategy.
- The objective is always to return to normal life.
- Sweden on the first wave was close, but not optimal.

Metropolitan area	Stockholm	Madrid	Catalonia	Brussels	Paris*	London*
Back in March 2020	1115	1269	1257	1011	1019	923
N-day death	(10 ICU)	(23 ICU)	(15 ICU)	(10 ICU)	(9 ICU)	(9 ICU)
minimizing + NL	97-day	102-d ay	100-day	101-d ay	113-d ay	104-d ay
	(0.941,	(0.941,	(0.941,	(0.941,	(0.941,	(0.941,
	+0.17)	+0.16)	+0.07)	+0.23)	+0.18)	+0.26)
Fitted + 90-day death	1579	2426	2325	1765	1911	1798
minimizing with 0.00	(2 ICU)	(14 ICU)	(25 ICU)	(8 ICU)	(19 ICU)	(16 ICU)
to healthy < 60 + NL	(0.25,	(0.37,	(0.71,	(0.35,	(0.78,	(0.60,
	0.00)	0.00)	0.00)	0.00)	0.00)	0.00)
Fitted + 90-day death	1454	1834	1630	1297	1384	1314
minimizing + NL	(2 ICU)	(15 ICU)	(34 ICU)	(9 ICU)	(28 ICU)	(17 ICU)
	(0.941,	(0.941,	(0.941,	(0.941,	(0.941,	(0.941,
	-1.98)	-0.84)	-0.43)	-0.96)	-0.35)	-0.50)

Table 8 – Final deaths per million (ICU/100K), isolation to vulnerable (with 0.94 maximum), and isolation to healthy <60 for various strategies (beginning on lockdown day) and followed by normal life.

Death Minimizing



Can we find optimal strategies EVEN with vaccines?

The Vaccine Gamble

- From game theory:
 - A gamble on the expectation of final deaths.
- It is solved though an oracle mechanism.
 - The Oracle 'knows' when a vaccine will be available.
 - The System calculates the expectation of final deaths.
- Assumptions:
 - The vaccine is ve+ 100% [No deaths after vaccine]
 - It can be inoculated to the whole population in 1 day.

There is NO better scenario for a vaccine!!!

After the first wave, what do we do?



120-day fitted + 180-day (death minimizing for no vaccine) (0.94, -0.24) + NL. Total final deaths 11034 (1656) (4 ICU/100K).

120-day fitted + 180-day (0.65,0.4) + Normal Life. Total final deaths 16573 (2487) (19 ICU/100K)

120-day (death minimizing for no vaccine) (0.94, 0.17) + Normal Life. Total final deaths 8420 (1264) (15 ICU/100K)

What if we are wrong? What if it is more lethal than we estimated?

A final thought...

Can an ebola patient avoid mandatory isolation?

The problem

- Not a new problem, Mary Mallon [1869-1938]
 - Chronic typhoid fever patient.
 - Required a ruling from the Supreme Court.
 - Not the only case.
- Solution requires to solve many extremes.
 - Bubble-Boy: Doesn't damage anyone, everyone damages him.
 - Chronic-Ebola: Damages everyone, nobody damages him.
 - Novel-Virus: Unknown damage function.



Our proposal: Isolation Exemption Insurance

One more thing...

If from 100,000 cases we send 2,000 to the lab, and 10% of the cases are influenza, then 10,000 cases are expected to be influenza, and 90,000 are expected to be untypified influenza-like diseases. Clearly no abnormality there, as positivity is compatible with history.

If from 1,000,000 cases we send 2000 to the lab, and 10% of the cases are influenza, then 100,000 cases are expected to be influenza, and 900,000 are expected to be untypified influenza-like diseases. Is this year compatible with history? What can you say about the distribution of viruses present on the part of the sample that is untypified?

Did they all grow at the same rate as influenza?

Do both years have the same hidden distribution?

Do all viruses find immunity weaknesses all at once?

Did you really think this was hypothetical? ③



NEVER ASSUME THE DISTRIBUTION OF THE VARIABLES YOU DON'T MEASURE

Want to know more?

Start here!!!

- SARS-CoV-2 waves in Europe : A 2-stratum SEIRS model solution: <u>https://www.medrxiv.org/content/10.1101/2020.10.09.20210146v3</u>
- SARS-CoV-2 waves in Europe simulator: <u>www.sars2seir.com/paper-12-2020/</u>
- Detection of Respiratory Viruses in Deceased Persons, Spain, 2017 <u>https://wwwnc.cdc.gov/eid/article/24/7/18-0162_article</u>
- Modeling strict age-targeted mitigation strategies for COVID-19 <u>https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236237</u>
- Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19 <u>https://pubmed.ncbi.nlm.nih.gov/32979941/</u>
- Virological assessment of hospitalized patients with COVID-2019 <u>https://www.nature.com/articles/s41586-020-2196-x</u>
- Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study <u>https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31483-5/fulltext</u>

Thank you for coming! I will be at the discussion zone to answer questions.