Making Useful Use of Particle Physics Experience in Epidemiology

Philip Bechtle

Jonas Dehning, Sebastian Mohr, Sebastian Contreras, Philipp Dönges, Emil Iftekhar, Oliver Schulz, Viola Priesemann

Physikalisches
Institut

13.04.23

PUNCH4NFDI
and what does that have to do with Epidemiology?

Examples for Applying Particle Physics Methods

How I finally started to be interested in soccer

PUNCH4NFDI and Epidemiology

UNIVERSITÄT BONN

PUNCH4NFDI

The National Research Data Infrastructure (in original German: Nationale Forschungs-Daten Infrastruktur NFDI) has the objective to systematically index, edit, interconnect and make available the valuable stock of data from science and research.

PUNCH4NFDI also offers services for the efficient scientific exploitation of research data. In doing so, PUNCH4NFDI collaborates closely with its European and international partners at the forefront of research in scientific data management.

https://punch4nfdi.de

UNIVERSITÄT BONN

Most Particle Physicists think in the context of *Events*

▶ In the area of epidemiology, that would mean that **every** single detected case was a separate **event** and we would know

- ▶ Gender, age, location, occupation, recent travels, movement data, contact history, . . .
- ▶ living conditions, eventual family, . . .
- ▶ Medical history, vaccination history, infection history . . .
- ▶ Testing history, sequencing, viral load tracking, exact history of COVID19 symptoms, medical treatment history, time until recovery or worse, . . .

- ▶ All observables were readily calibrated for known inefficiencies and delays in a central processing and then analysed
- ▶ A lot of detailed control/testbeam experiments/measurements would be performed, all auxiliary/environmental data would be tracked
- ▶ All data would be treated uniformly and shared worldwide, *but (if unlucky) only within 2 competing collaborations*

▶ For obvious reasons of privacy and data protection, this is neither possible nor attractive

UNIVERSITÄT BONN

Let's Measure Currents instead of Particles

▶ What situation do we encounter with epidemiological data?
  ▶ Date, gender, age, location is resolved in pre-defined bins and data is reported in an *aggregated* and *anonymized* way
  ▶ It is often unclear what the reference date means – reported, symptoms onset, sample taken...
  ▶ Different reporting delay structures for different data or different countries
  ▶ No official recording of the detailed implementation of political measures
  ▶ Movement, contact, cohort study, medical/hospital data... is reported in separate independent studies
  ▶ The binning or the time resolution is not guaranteed to be coordinated between any of the above
  ▶ The most sensitive auxiliary data (contact networks from cell phone GPS data) is not publicly available
  ▶ No continuously analyzed cohorts in all countries/regions, analysis of incredibly useful auxiliary data like waste water not common, not a lot of sequencing, unclear whether enough testing capacity, definitely not good enough tracing, lack of consistent communication, scientific statements are commonly interpreted as negotiation standpoints...

UNIVERSITÄT BONN

Therefore, let's try to be helpful in a dire situation

▶ We identified three areas where PUNCH4NFDI could be helpful and in March 2021 contacted the NFDI4Health consortium and Viola Priesemann

▶ Naive ideas wher ewe could be helpful:
  ▶ Data Management and Distributed Data Analysis
  ▶ Statistical Tools for Inferrence
  ▶ Machine Learning Tools for Anomaly Detection

▶ Let's see how it turned out

UNIVERSITÄT BONN

Examples for Applying Particle Physics Methods

In Collaboration with NET CHECK

- ▶ Together with Sten Rüdiger and Steven Schulz from NET-CHECK
- ▶ Contact networks influence how pandemics spread
- ▶ **Effective contact network** is different for different transmission paths
- ▶ Conventional wisdom: Need to know the transmission path to know the effective contact network
- ▶ Try to turn this around: If we knew the physical contact network, observing the pandemic spread would allow to extract the dominant transmission path (works if enough auxiliary parameters are known)
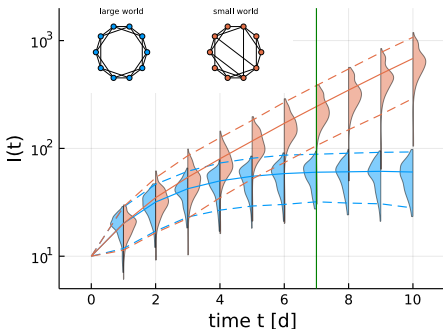
Example: Calibrating in the Most Sensitive Way

▶ From Oliver Schulz and Arne Gottwald
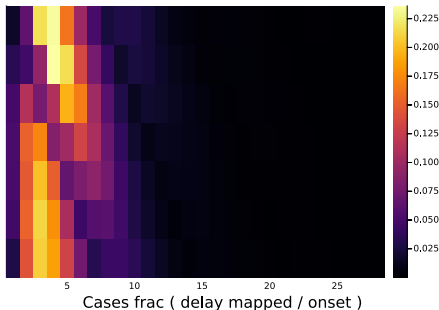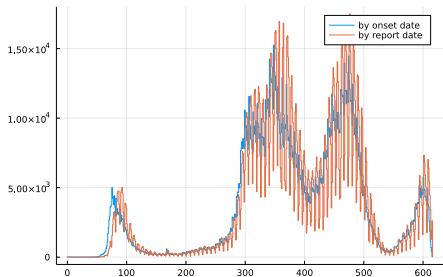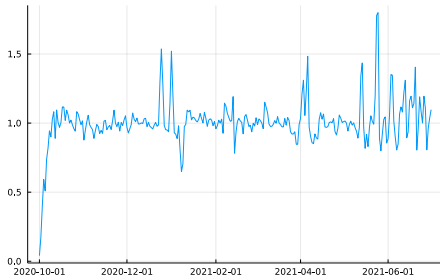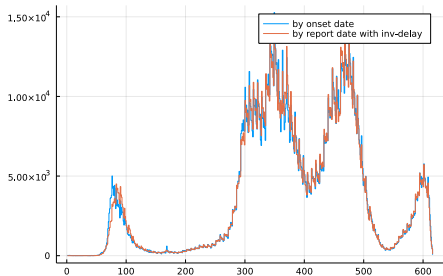
Bayesian Inference Using Hamiltonian Monte Carlo

▶ Hamiltonian Monte Carlo, developed in the context of QCD on the lattice, is widely employed to sample the posterior space of Bayesian Compartmental Models – here using `BAT.jl`
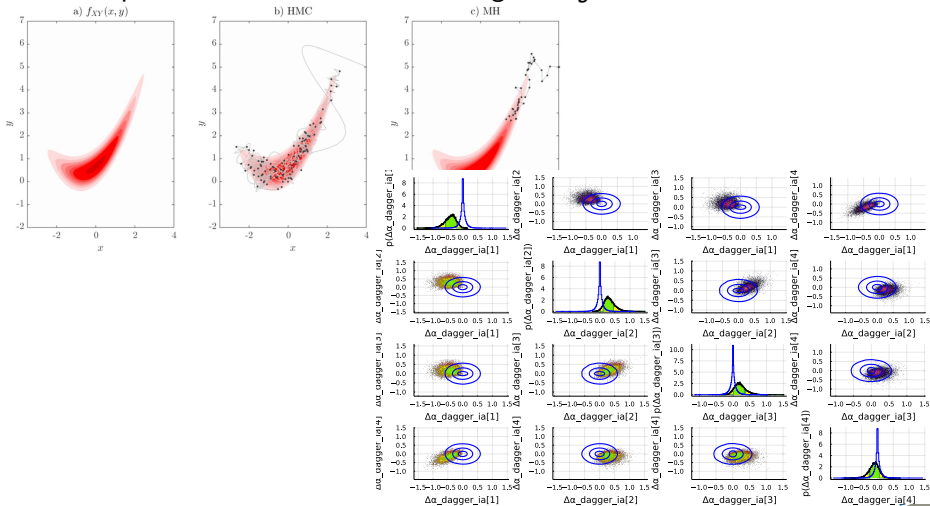
UNIVERSITÄT BONN

BAT.jl

`docs stable` `docs dev` `license MIT` `CI passing` `codecov 53%` `DOI 10.5281/zenodo.2587213`

Welcome to BAT, a Bayesian analysis toolkit in Julia.

BAT.jl offers a variety of posterior sampling, mode estimation and integration algorithms, supplemented by plotting recipes and I/O functionality.

BAT.jl originated as a rewrite/redesign of BAT, the Bayesian Analysis Toolkit in C++. BAT.jl now offer a different set of functionality and a wider variety of algorithms than its C++ predecessor.

```
https://github.com/bat/BAT.jl
```

UNIVERSITÄT BONN

How I finally started to be interested in soccer

UNIVERSITÄT BONN

Was the EURO 2020 (in 2021) safe?

Impact of the Euro 2020 championship on the spread of COVID-19

Jonas Dehning, Sebastian B. Mohr, Sebastian Contreras, Philipp Dönges, Emil N. Iftekhar, Oliver Schulz,

Philip Bechtle ✉ & Viola Priesemann ✉

https://www.nature.com/articles/s41467-022-35512-x.

UNIVERSITÄT BONN

Just to introduce the methods:
A friendly SIR Pandemic Compartmental Model

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\lambda \frac{SI}{N_{\mathrm{pop}}}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \lambda \frac{SI}{N_{\mathrm{pop}}} - \mu I$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \mu I \,.$$

UNIVERSITÄT BONN

A model capable of identifying the fraction of EURO 2020 related infections
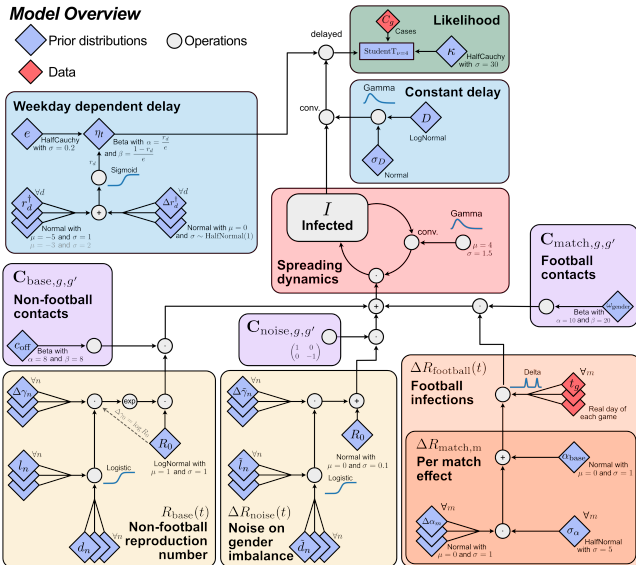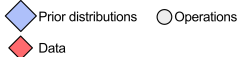
The core idea: Parametrize the "Physics"

Specific sensitivity comes from

- ▶ More men do watch soccer than females
- ▶ There might be a time structure of cases correlated to fan activity

Write the dynamics of infected I and susceptible E as a discrete convolution

$$I_g(t) = \frac{S_g(t)}{N} \sum_{g'=1}^{2} \mathbf{R}_{\text{eff},g,g'}(t) \sum_{\tau=0}^{10} I_{g'}(t-1-\tau)\, G(\tau),$$

$$S_g(t) = S_g(t-1) - E_g(t-1),$$

$$G(\tau) = \text{Gamma}(\tau; \mu = 4, \sigma = 1.5).$$

And parametrize the asymmetric dynamics between the genders

$$\mathbf{R}_{\text{eff},g,g'}(t) = R_{\text{base}}(t) C_{\text{base},g,g'} + \Delta R_{\text{football}}(t) C_{\text{match},g,g'} + \Delta R_{\text{noise}}(t) C_{\text{noise},g,g'},$$

$$\mathbf{C}_{\text{base}} = \begin{pmatrix} 1 - c_{\text{off}} & c_{\text{off}} \\ c_{\text{off}} & 1 - c_{\text{off}} \end{pmatrix},$$

with $c_{\text{off}} \sim \text{Beta}(\alpha = 8, \beta = 8)$.

$$\mathbf{C}_{\text{match, unnorm.}} = \begin{pmatrix} (1 - \omega_{\text{gender}})^2 & \omega_{\text{gender}}(1 - \omega_{\text{gender}}) \\ \omega_{\text{gender}}(1 - \omega_{\text{gender}}) & \omega_{\text{gender}}^2 \end{pmatrix}$$

$$\mathbf{C}_{\text{match}} = \frac{\mathbf{C}_{\text{match, unnorm.}}}{\left\| \mathbf{C}_{\text{match, unnorm.}} \cdot \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}^T \right\|_2}$$

$$\omega_{\text{gender}} \sim \text{Beta}(\alpha = 10, \beta = 20).$$

UNIVERSITÄT BONN

Of course this is futile, because . . .

The UEFA of course knew better all the time:

At this stage of the tournament, how do you judge the success of the COVID-19 mitigation measures that have been put in place for the EURO?

Dr Bahtijarević: We do not see any big increase in the number of cases across the EURO 2020 host venues, or in most of the countries of the participating nations. This demonstrates that our measures are effective and being implemented correctly.
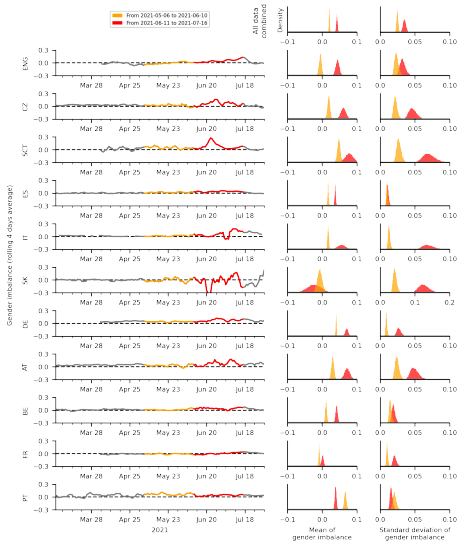
Dr Bahtijarević: UEFA has put in place a number of mitigation measures and, so far, we are satisfied with the overall outcome. We are constantly monitoring how these measures are working, and whenever we see areas to improve, we react together with the competent local authorities to tackle them.

https:

//www.uefa.com/returntoplay/news/026a-129e7a75d53b-81e501a50eb2-1000--covid-measures-at-uefa-euro-2020/

UNIVERSITÄT BONN

Infections from EURO 2020 fan activity

Is there something at all?

Infections from EURO 2020 fan activity
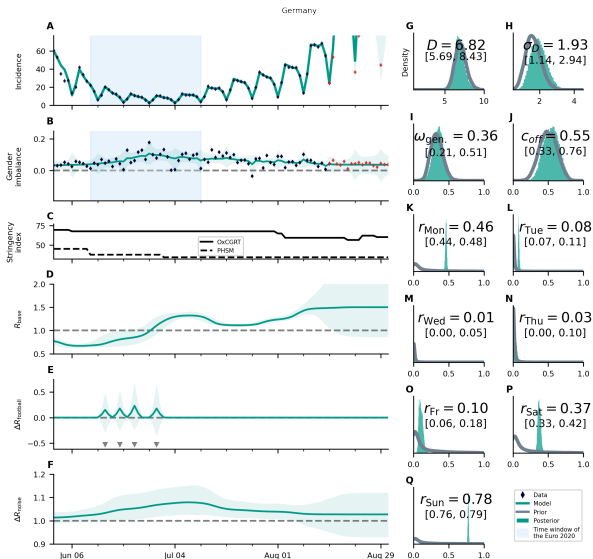
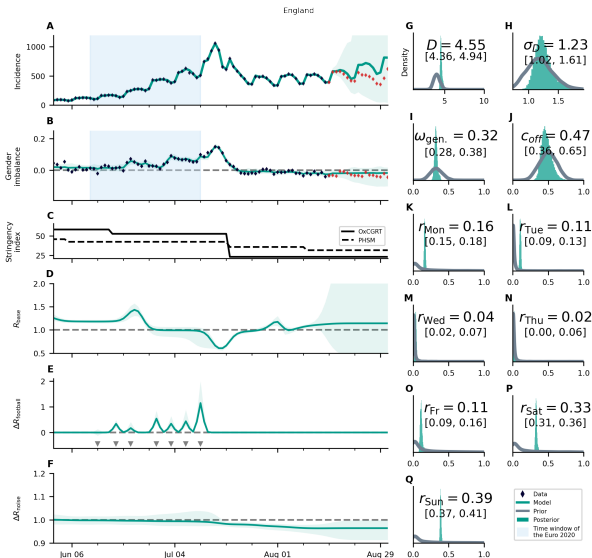# Infections from EURO 2020 fan activity
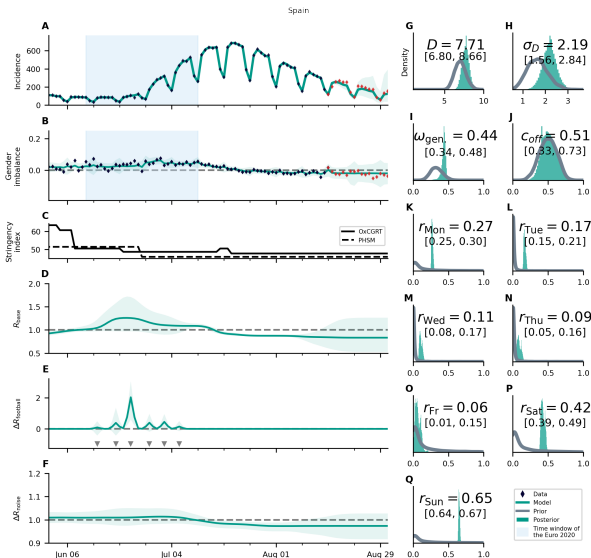
Infections from EURO 2020 fan activity
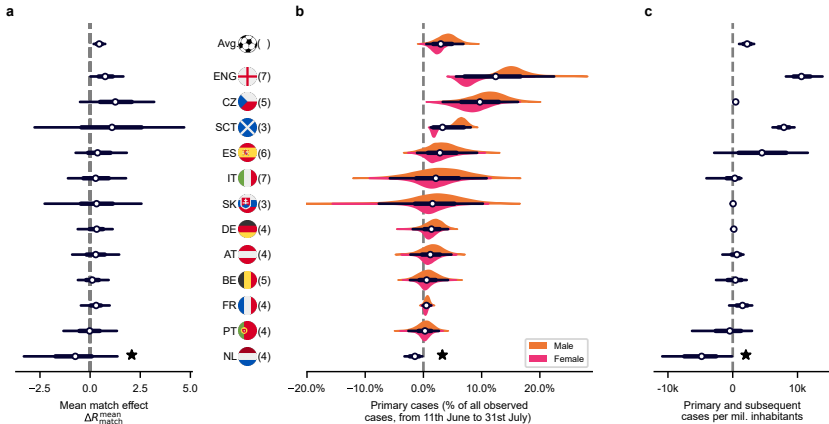
Does it get more striking?
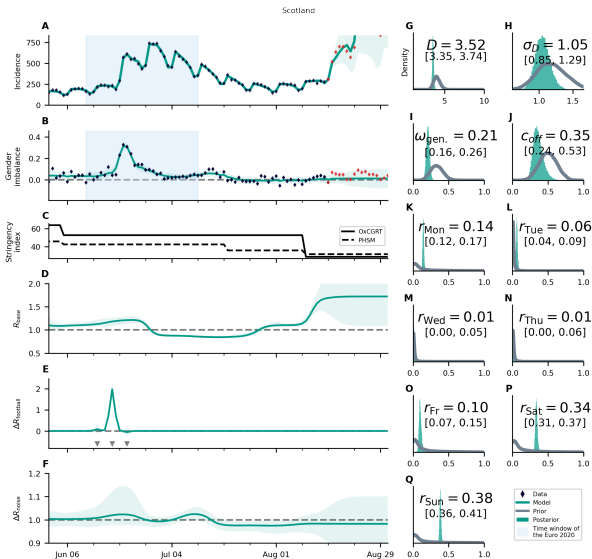
Infections from EURO 2020 fan activity
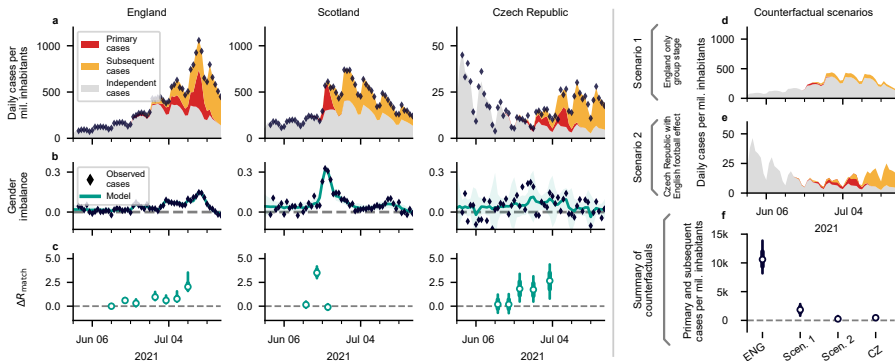
Does it get even more striking?

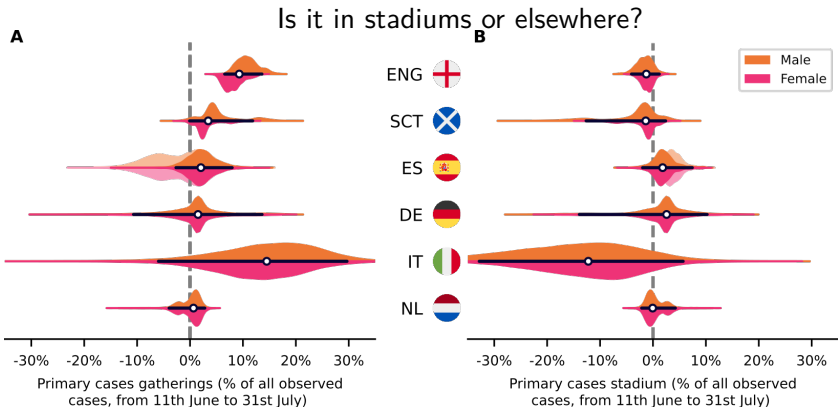# Infections from EURO 2020 fan activity

Let's look at all results

Infections from EURO 2020 fan activity
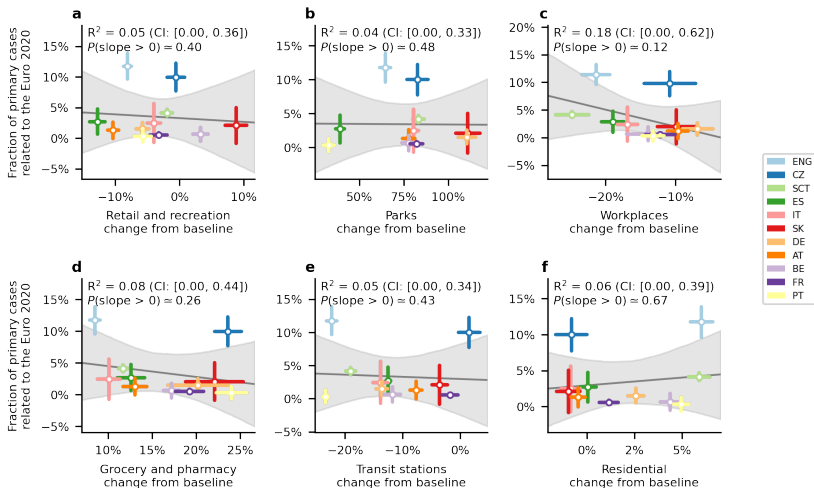
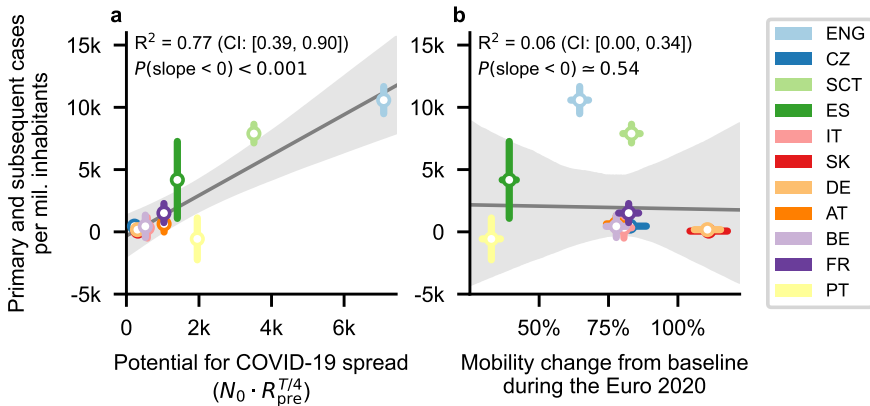## Let's look at all results

Understanding the EURO 2020 Infections Model

Is it in stadiums or elsewhere?

Understanding the EURO 2020 Infections Model

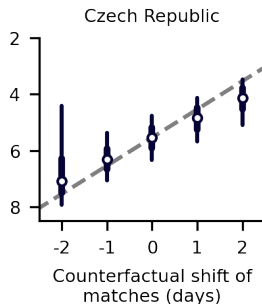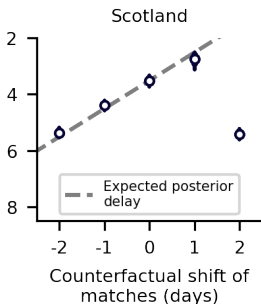## Why are the results so different for different countries?
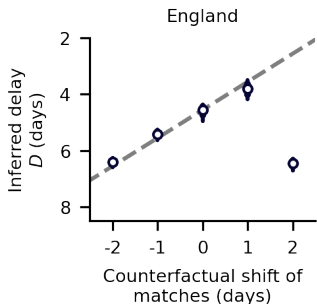
Understanding the EURO 2020 Infections Model

Why are the results really so different for different countries?

Understanding the EURO 2020 Infections Model

Is that really believable?

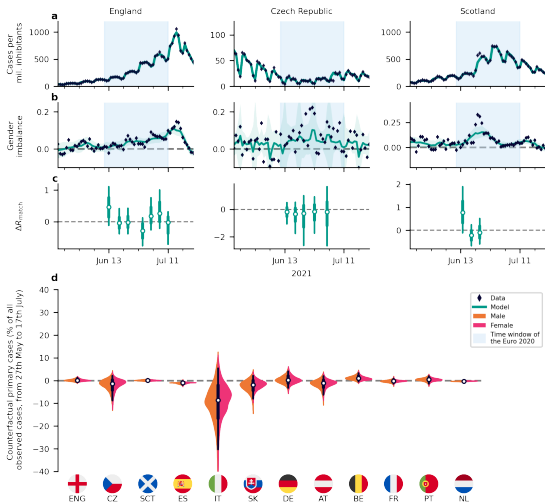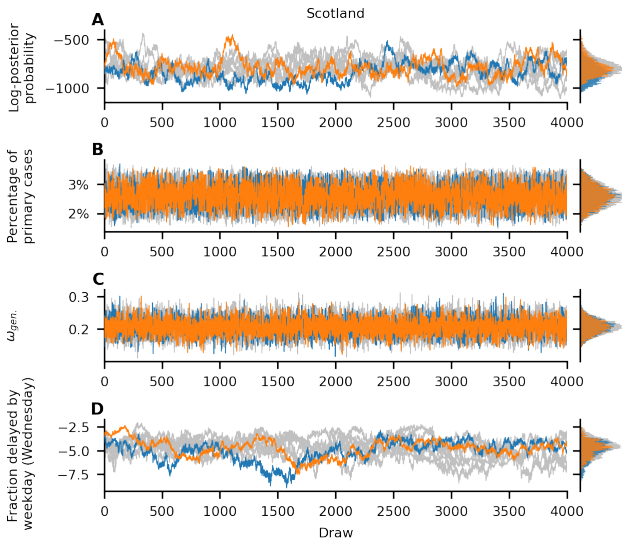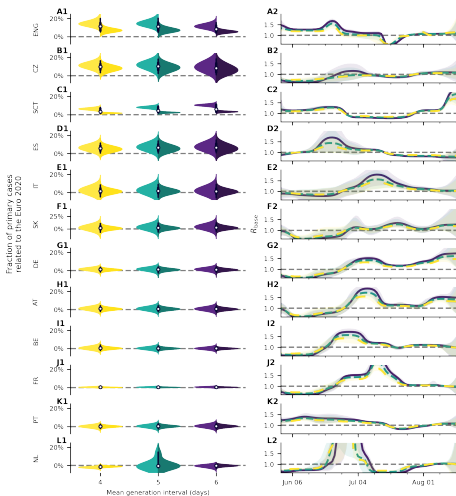# Understanding the EURO 2020 Infections Model

Understanding the EURO 2020 Infections Model

Is that really believable?

Understanding the EURO 2020 Infections Model

What??? Bayesian????? From YOU????????

UNIVERSITÄT BONN

Understanding the EURO 2020 Infections Model

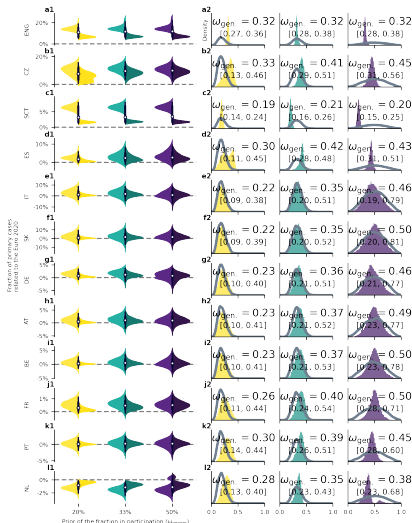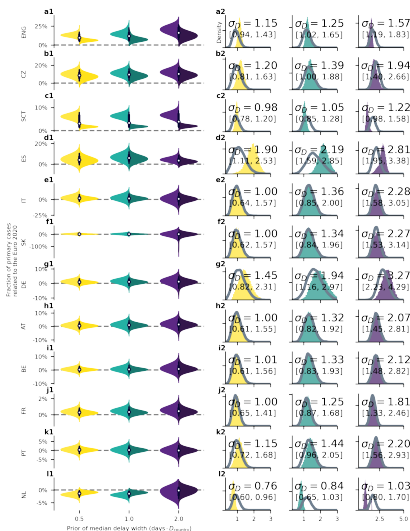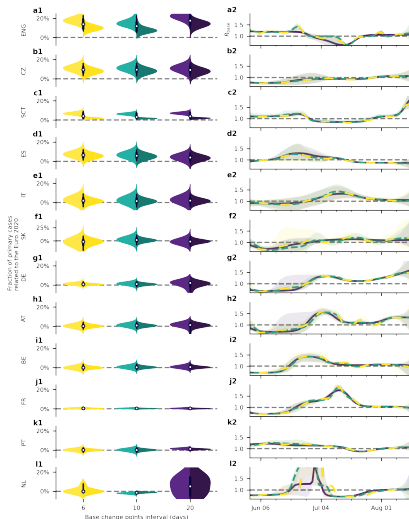## What??? Bayesian????? From YOU????????

Understanding the EURO 2020 Infections Model

What??? Bayesian????? From YOU????????

Understanding the EURO 2020 Infections Model

What??? Bayesian????? From YOU????????

UNIVERSITÄT BONN

Is a small effect on R significant?
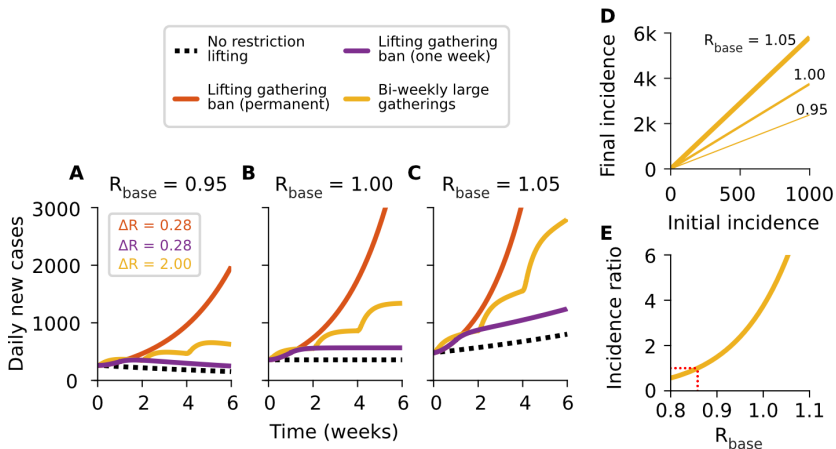
Conclusions

▶ We hope we can be helpful

▶ We can learn from social experiments

▶ The situation still doesn't look like politics informed by science has the safe standing we hoped for, but we as scientists can only try to be helpful

▶ A lot of complex projects are being thought of, and two more rather complex models are being worked on

▶ Contact me if you are intersted to invest time and to work with us

UNIVERSITÄT BONN

APPENDIX

Parametrization of Ignorance

To account for effects not related to the soccer games, e.g. non-pharmaceutical interventions, vaccinations, seasonality or variants, we introduce a slowly changing reproduction number $R_{\text{base}}(t)$, which is identical for both genders and should map all other not specifically modelled gender independent effects.

$$R_{\text{base}}(t) = R_0 \exp\left(\sum_n \gamma_n(t)\right) \tag{19}$$

$$R_0 \sim \text{LogNormal}\left(\mu = 1, \sigma = 1\right) \tag{20}$$

This base reproduction number is modelled as a superposition of logistic change points $\gamma(t)$ every 10 days, which are parameterized by the transient length of the change points l, the date of the change point d and the effectivity of the change point $\Delta\gamma^*$. The subscripts n denotes the discrete enumeration of the change points.

$$\gamma_n(t) = \frac{1}{1 + e^{-4/l_n \cdot (t - d_n)}} \cdot \Delta\gamma_n \tag{21}$$

$$\Delta\gamma_n \sim \mathcal{N}\left(\Delta\gamma_{n-1}, \sigma_{\Delta\gamma}\right) \quad \forall n \quad \text{with } \Delta\gamma_0 = \log R_0 \tag{22}$$

$$\sigma_{\Delta\gamma} \sim \text{HalfCauchy}\left(0.5\right) \tag{23}$$

$$l_n = \log\left(1 + \exp(l_n^\dagger)\right), \tag{24}$$

$$l_n^\dagger \sim \mathcal{N}\left(4, 1\right) \quad \forall n \quad \text{(unit is days)} \tag{25}$$

$$d_n = 27^{\text{th}} \text{ May 2021} + 10 \cdot n + \Delta d_n \quad \text{for } n = 0, \ldots, 9 \tag{26}$$

$$\Delta d_n \sim \mathcal{N}\left(0, 3.5\right) \quad \forall n \quad \text{(unit is days)} \tag{27}$$

A "mechanistic" Compartment Model

$$\frac{d}{dt}S_{a,r,p \in \{1,2,4\}} = -\mu_{a,r,p}S_{a,r,p} \qquad\qquad + \delta_{a,r,p}^{vacc.}$$

$$\frac{d}{dt}S_{a,r,p=3} = -\mu_{a,r,p=3}S_{a,r,p=3} \qquad + \sum_{g,p \in \{1,2\}} \left(\chi_{a,p,g} \cdot \tau^{rec.} \cdot R_{a,r,p,g}\right) \quad + \delta_{a,r,p=3}^{vacc.}$$

$$\frac{d}{dt}S_{a,r,p=5} = -\mu_{a,r,p=5}S_{a,r,p=5} \qquad + \sum_{g,p \in \{3,4,5\}} \left(\chi_{a,p,g} \cdot \tau^{rec.} \cdot R_{a,r,p,g}\right) \quad + \delta_{a,r,p=5}^{vacc.}$$

$$\frac{d}{dt}E_{a,r,p,g} = +\mu_{a,r,p,g}S_{a,r,p} \qquad - \tau^{lat.}E_{a,r,p,g}$$

$$\frac{d}{dt}I_{a,r,p,g} = +\tau^{lat.}E_{a,r,p,g} \qquad - \tau^{inf.}I_{a,r,p,g}$$

$$\frac{d}{dt}R_{a,r,p,g} = +\tau^{inf.}I_{a,r,p,g} \qquad - \tau^{rec.} \cdot R_{a,r,p,g}$$

$$\frac{d}{dt}D_{a,r,p,g} = \qquad\qquad\qquad + (1 - \chi_{a,p,g}) \cdot \tau^{rec.} \cdot R_{a,r,p,g} \quad .$$

UNIVERSITÄT **BONN**

A "mechanistic" Compartment Model

$$\mu_{a,r,p,g} = \sum_{a_s,r_s,p_s} \left(\mu_{r,r_s}^M \cdot \mu_{a,a_s,r}^C \cdot \mu_{a,r}^R \cdot \frac{I_{a_s,r_s,p_s,g}}{N_{a_s,r_s}} \cdot \mu_{g,a,a_s,p,p_s}^T \right) + \mu_g^F \cdot \mu_{g,p}^{T'} \cdot$$

UNIVERSITÄT BONN

A "mechanistic" Compartment Model

$$\mu_{a,a_s}^{C,home} = r^{home} \cdot t^{home} \cdot C_{a,a_s}^{home}$$

$$\mu_{a,a_s,r}^{C,school} = r^{school} \cdot t^{school} \cdot \alpha_r^{school} \cdot C_{a,a_s}^{school}$$

$$\mu_{a,a_s,r}^{C,work} = r^{work} \cdot t^{work} \cdot \alpha_r^{work} \cdot a_{effect}^{school-work} \cdot C_{a,a_s}^{work}$$

$$\mu_{a,a_s}^{C,other} = r^{other} \cdot t^{other} \cdot C_{a,a_s}^{other} \cdot$$

UNIVERSITÄT BONN

A "mechanistic" Compartment Model

$$\delta^{vacc.}_{a,r,p=1} = - \delta^{first}_{a,r}$$

$$\delta^{vacc.}_{a,r,p=2} = + \delta^{first}_{a,r} \quad - \tau^{partvacc.} \cdot S_{a,r,p=2}$$

$$\delta^{vacc.}_{a,r,p=3} = - \delta^{second}_{a,r} + \tau^{partvacc.} \cdot S_{a,r,p=2}$$

$$\delta^{vacc.}_{a,r,p=4} = + \delta^{second}_{a,r} - \tau^{fullvacc.} \cdot S_{a,r,p=4}$$

$$\delta^{vacc.}_{a,r,p=5} = \qquad\qquad + \tau^{fullvacc.} \cdot S_{a,r,p=4} \cdot$$

UNIVERSITÄT BONN

A "mechanistic" Compartment Model

Contact matrix home

A "mechanistic" Compartment Model

A "mechanistic" Compartment Model

A "mechanistic" Compartment Model

UNIVERSITÄT BONN
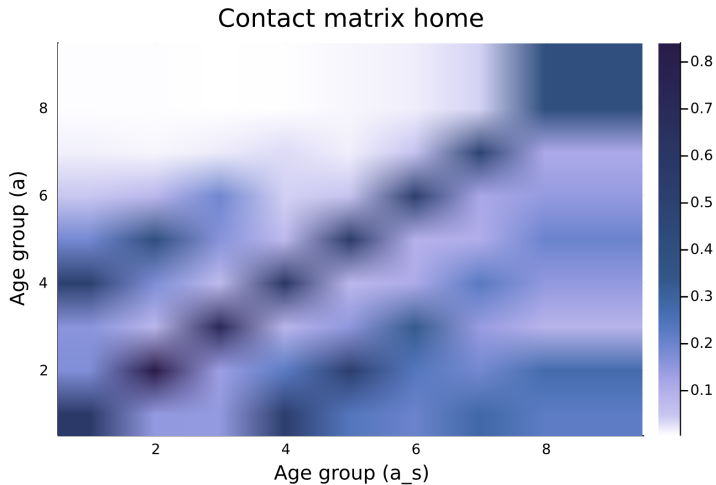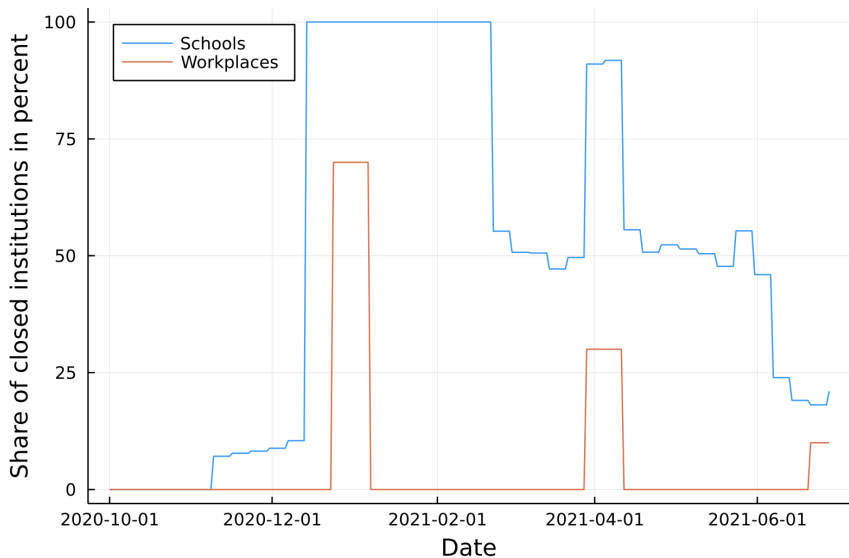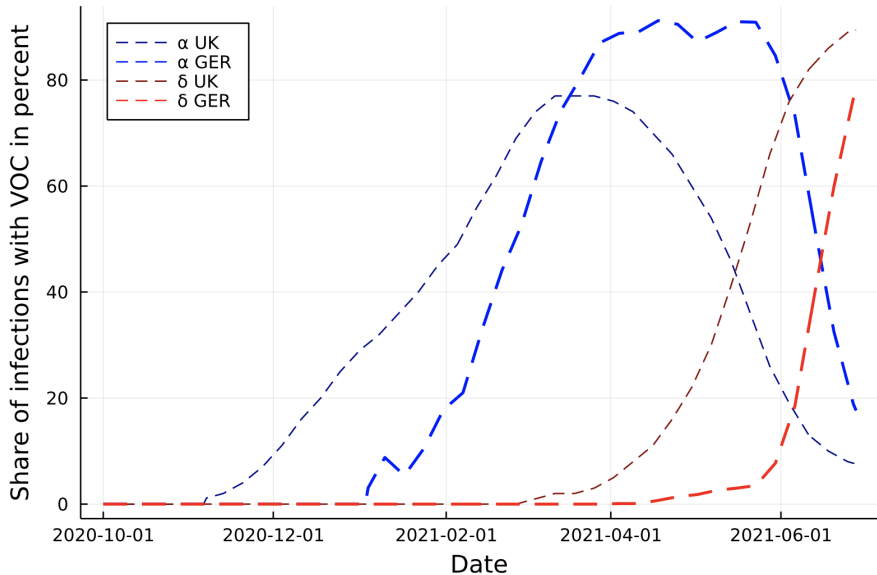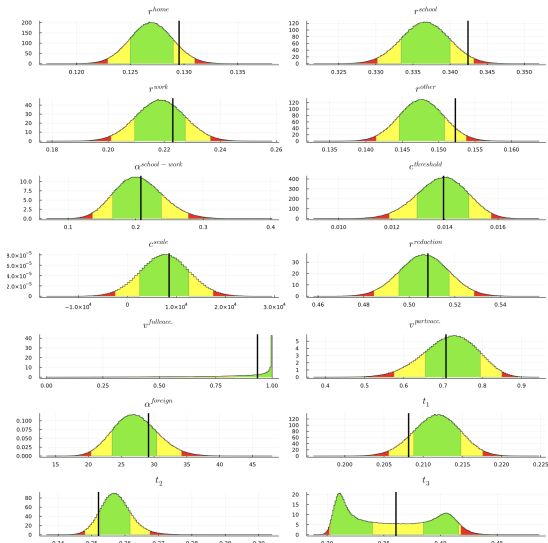
A "mechanistic" Compartment Model
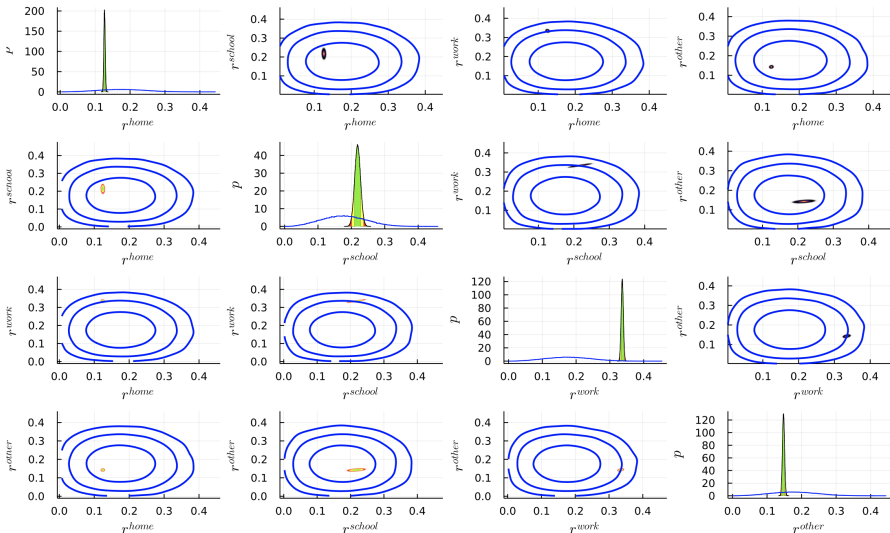
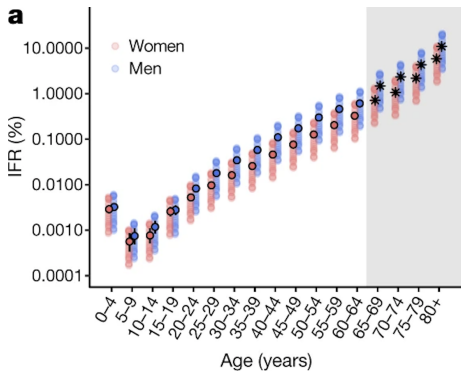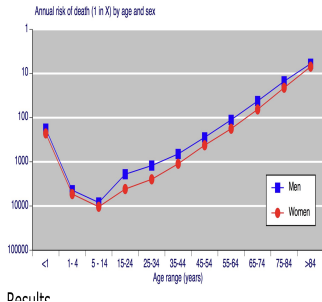# Different ways of looking at the Data

a

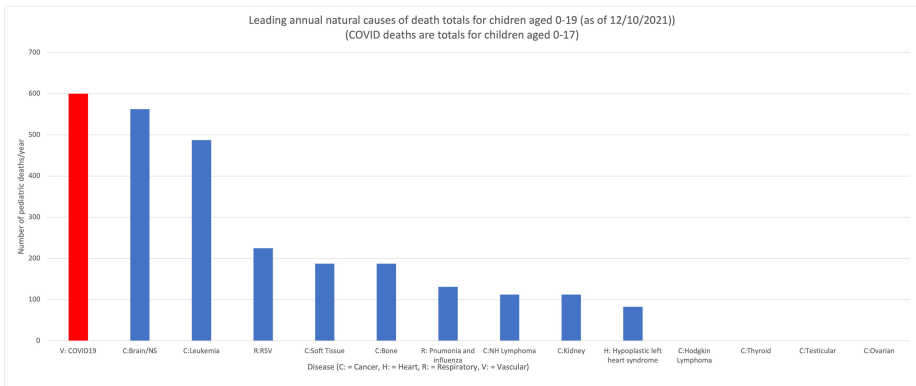Figure 1: Annual risk of death by age and sex in the UK

Luckily, childrens lifes are far less endangered by COVID than that of the elderly, So don't need to vaccinate children, right?

Well, apart from illness, what is the relevant metric?

O'Driscoll *et al*. "Age-specific mortality and immunity patterns of SARS-CoV-2". Nature 590, 140–145 (2021). `https://doi.org/10.1038/s41586-020-2918-0`

Different ways of looking at the Data

Leading annual natural causes of death totals for children aged 0-19 (as of 12/10/2021))
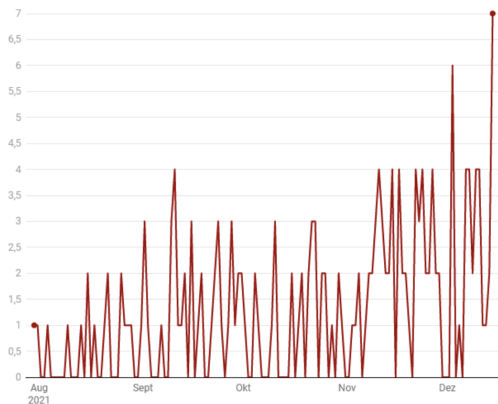(COVID deaths are totals for children aged 0-17)

COVID is the leading non-accidental cause of death of children in the US

UNIVERSITÄT BONN

Different ways of looking at the Data

Anzahl gemeldeter ITS-Erstaufnahmen von COVID-19-Fällen

Deutschland, Kinder-Intensivstationen

Stand: 15.12.2021 12:23
Quelle: DIVI-Intensivregister • Daten herunterladen • Erstellt mit Datawrapper

In Germany, ICU admissions of children are rising – expected to get worse with Omicron

UNIVERSITÄT BONN

Different ways of looking at the Data

Different ways of looking at the Data

STUDIE NIMMT BEDENKEN

Schulen sind kein Treiber der Pandemie

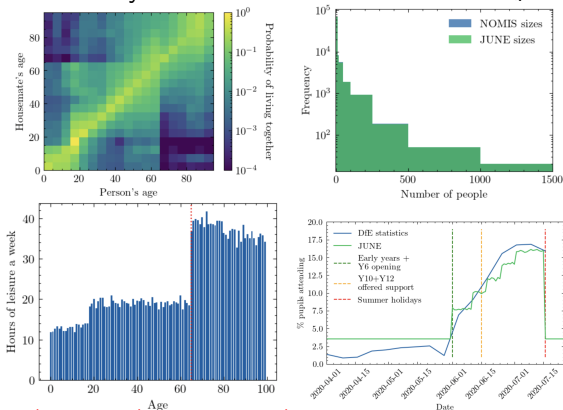Von Heike Schmoll, Berlin
15.12.2021, 16:22

Das Infektionsrisiko in Schulen ist deutlich niedriger als im häuslichen Umfeld. Das belegt eine Studie im Auftrag der Kultusministerkonferenz. Schließungen dürfe es also nur im absoluten Notfall geben.

Infektionsrisiko der SchülerInnen in Deutschland

Das Infektionsrisiko von SchülerInnen sowohl aus den Landkreis- und Bundeslanddaten der KMK als auch aus den im Arbeitspaket 3 erhobenen Daten lag eher selten (z. B. im Rahmen der 2. Welle) oberhalb des Bevölkerungsrisikos. Im Rahmen der 3. Welle ist das Infektionsrisiko von SchülerInnen im Vergleich zur zweiten Welle deutlich zurückgegangen. Dies kehrt sich ab Sommer 2021 um, vermutlich weil die Impfabdeckung bei SchülerInnen niedriger und die Testerfassung besser war als bei der übrigen Bevölkerung (S.52-59).

Independent from PUNCH4NFDI:
From SHERPA to Agent Based Models

▶ Matthias Schott (Mainz) and Friedemann Neuhaus applied the JUNE framework written by Frank Kraus *et al.* to Germany

https://github.com/fneuhaus/JUNE_germany/

https://github.com/IDAS-Durham/JUNE

https://royalsocietypublishing.org/doi/full/10.1098/rsos.210506

UNIVERSITÄT BONN

A few general proposals from the above comparison

Improving the data and improving the data management could include:

▶ A central data base, definition of compartments and pre-processing
Analogy from High Energy Physics: ATLAS and CMS each extract
about 60 TB/s from their detectors, store $\mathcal{O}(GB/s)$, (re)process
$\mathcal{O}(100\ GB/s)$ and handle $\mathcal{O}(100's\ PB)$ on the WLCG

  ▶ Analysts do not work with raw data, but with Analysis Object Data

▶ A central calibration and uncertainty parametrization
Analogy from HEP: Everything from test beam measuremnts to
alignment, online/offline data quality control and (re) processing

  ▶ We get physics quasi in real time
  ▶ Need common control measurements and auxiliary data handled in the
    same consistent compartmentalization

UNIVERSITÄT BONN
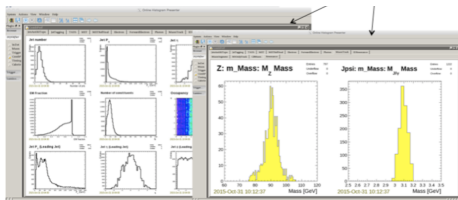
A few general proposals from the above comparison

(cont'd)

- ▶ Provide a formalized channel of interaction between researchers working with different data, e.g.
  Analogy from HEP: old example LEP EW WG, current example LHC Higgs WG, . . .
    - ▶ Deriving real-world effective contact networks from cell phone data needs to be in the hands of only a few, but:
    - ▶ The analysis needs to be curtailed to different contact situations, different transmission chains, etc. We need to optimize direct interaction on all these issues
- ▶ Obvious international cooperation and standardization
- ▶ A common analysis infrastructure
  Analogy from HEP: WLCG, or maybe a better goal: Make real use of the EOSC

UNIVERSITÄT BONN