

# **Modelowanie epidemii COVID-19**

**Modelling COVID-19 using Agent Based Models - presenting  
MOCOS approach**

**Marcin Bodych 2021-11-04**

**~225%**

**Delta variant infectivity compared to Chinese variant**

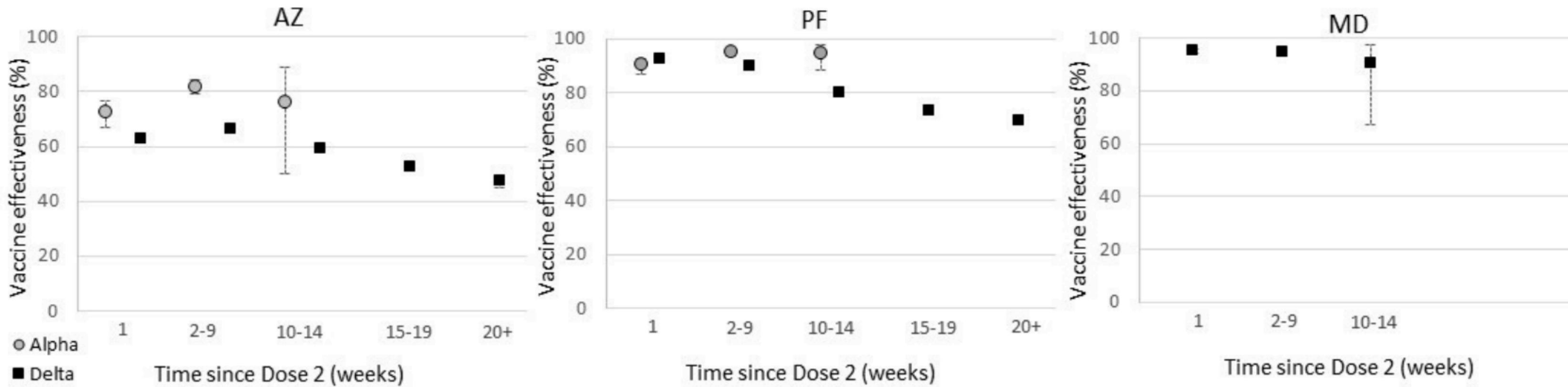
~60%

**Astra Zeneca vaccine effectiveness against Delta symptomatic disease 10-14 weeks after second dose**

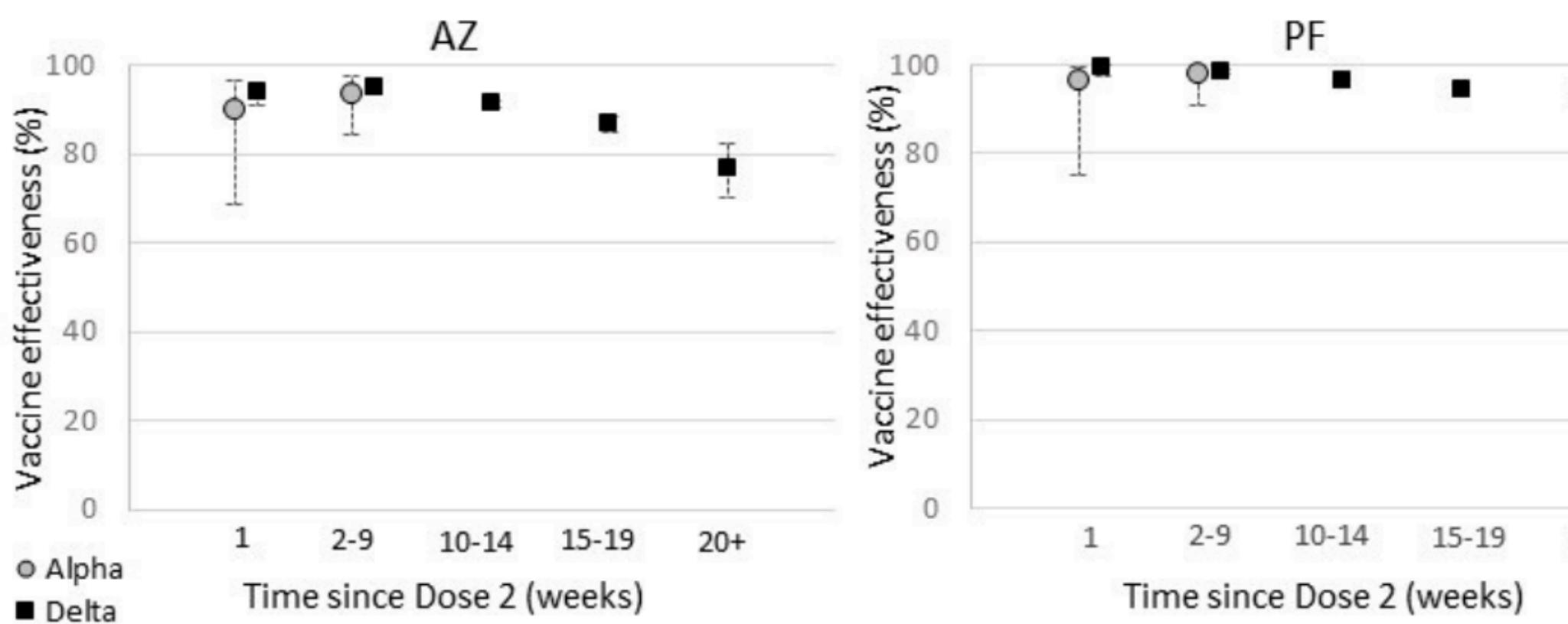
~90%

**BioNTech-Pfizer vaccine effectiveness against death 20+ weeks after second dose**

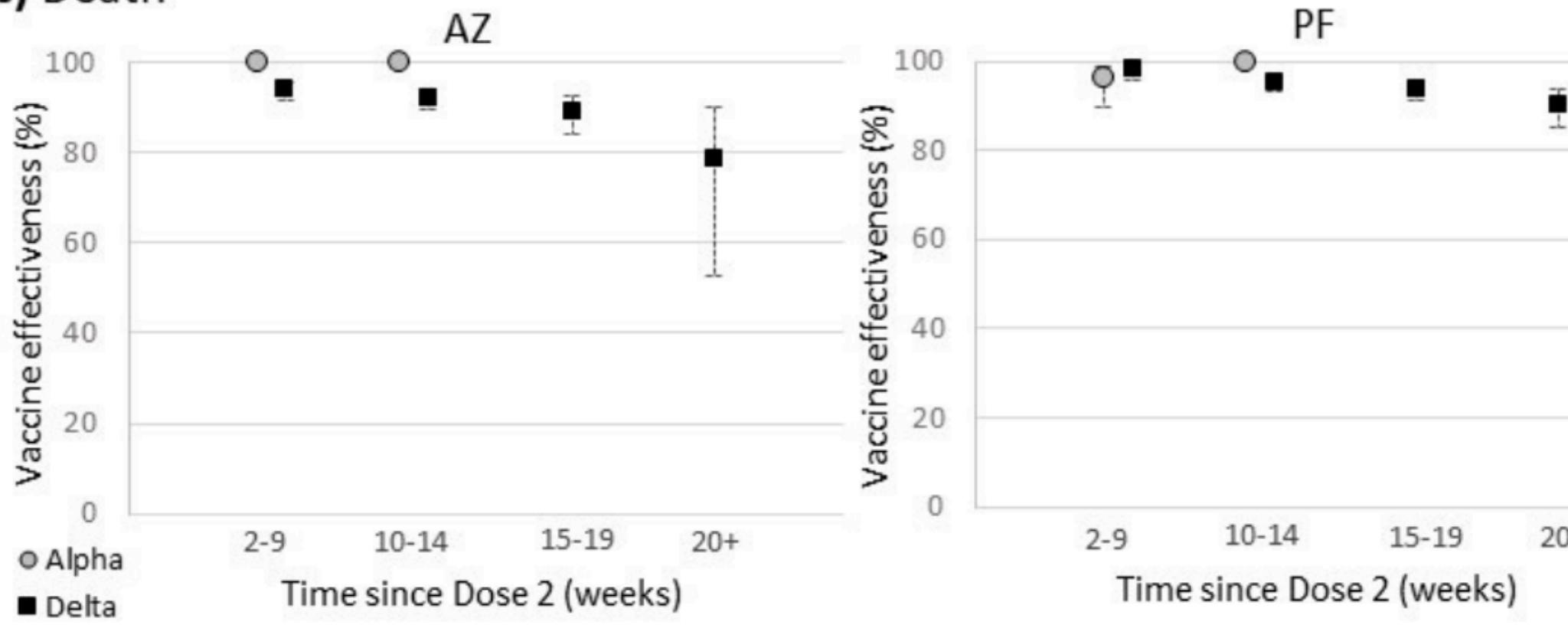
### a) Symptomatic disease



### b) Hospitalisation



### c) Death



[Link to the article](#)

# Agenda

## Table of Contents

- Overview of the model
- Implementation details
- Infrastructure usage
- Summary

## Key takeaway

Agent based models combined with good understanding of the dynamics can provide insightful guidance for even better understanding of the complex systems, even though only partial information is directly observable.

# Agenda

## Table of Contents

- Overview of the model
- Implementation details
- Infrastructure usage
- Summary

## Key takeaway

Agent based models combined with good understanding of the dynamics can provide insightful guidance for even better understanding of the complex systems, even though only partial information is directly observable.

Finally:



# MOCOS model in a pill

- Event driven microsimulation model of COVID 19 epidemics
- Continuous-time distributions of disease progression (developing symptoms, self-quarantine)
- Disease progression on the synthetic population based on available statistical data
- Random and varied structure of intra-household and out-household contacts
- Assessing social intervention measures: vaccination, quarantine, targeted testing, contact tracing, contact reduction, contact tracking app
- Extendable and adjustable to specific scenarios

# Input parameters

- Population with basic features:  
age, gender, household assignment
- Potential use of more features:  
social competence, workplace, high-risk profession,  
comorbidity diseases, smoking status, region  
assignment, geolocation
- There is no hard limit on population size
  - VM with 32GB RAM can simulate pandemics on Poland – size population)
- Configuration JSON specifying kernel intensities,  
initial conditions, detection delays, path to the  
generated population file
- Disease progression distributions

Example attributes:

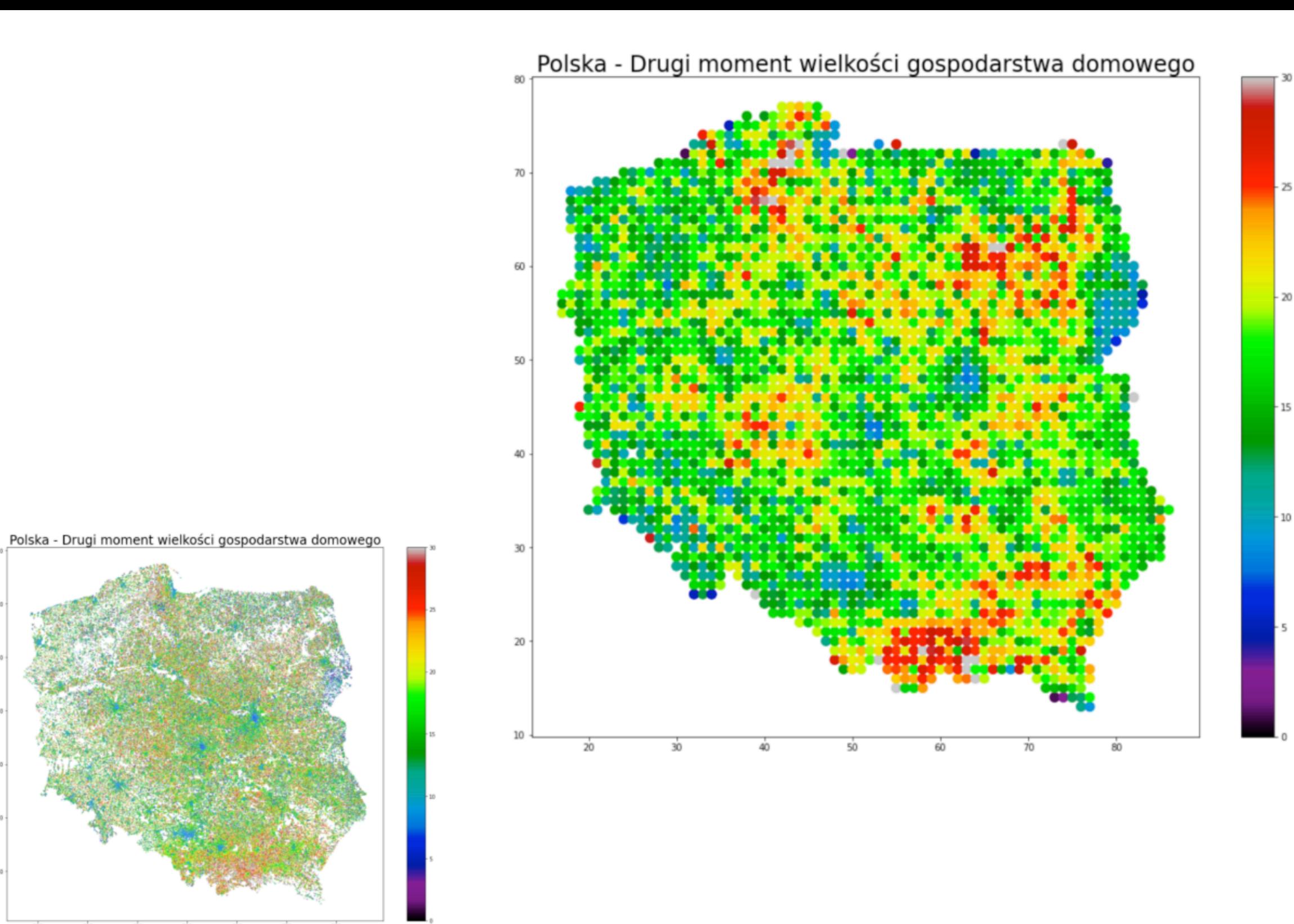
- ▶ Basic features:
  - ▶ Agent id: 102564
  - ▶ Age: 23 yo
  - ▶ Gender: male
  - ▶ Household: id 410256  
(Living with 4 other household members:  
47 yo female, 42 yo male, 19 yo female,  
15 yo female)
- ▶ (Optional) Other features:
  - ▶ Social competence: friendly (factor 3.32)
  - ▶ Workplace: id 5 (coworking with 20 other  
people with ids: [...])
  - ▶ High-risk profession: No
  - ▶ Comorbidity diseases: No
  - ▶ Smoking status: No
  - ▶ Region assignment: Pomerania
  - ▶ Geolocation: 54.372158, 18.638306



# Example configuration JSON

```
{  
    "num_trajectories" : 100,  
    "population_path" : "../poland_v3.jld2",  
    "detection_mild_proba" : 0.075,  
    "contact_tracking": {  
        "probability": 0.1,  
        "backward_detection_delay" : 1.75,  
        "forward_detection_delay" : 1.75,  
        "testing_time" : 0.25  
    },  
    "transmission_probabilities": {  
        "household": 0.01,  
        "constant": 0.135,  
        "hospital": 0.0,  
        "age_coupling_param": 0.405,  
        "age_coupling_data_path": "../contact_mat_pol.jld2"  
    },  
    "imported_cases" : [  
        {  
            "function" : "InstantOusideCases",  
            "params" : {  
                "num_infections" : 100,  
                "import_time": 0.0,  
                "strain" : "Delta"  
            }  
        }  
    ],  
    "initial_conditions": {  
        "cardinalities": {  
            "infectious": 0  
        },  
        "immunization": {  
            "order_data": "../poland_v3-immunization-order.jld2",  
            "level": 0.45  
        }  
    },  
    "phone_tracking": {  
        "usage" : 0.03,  
        "detection_delay" : 0.25,  
        "usage_by_household" : true  
    },  
    "stop_simulation_threshold": 40000000,  
    "stop_simulation_time": 400  
}
```

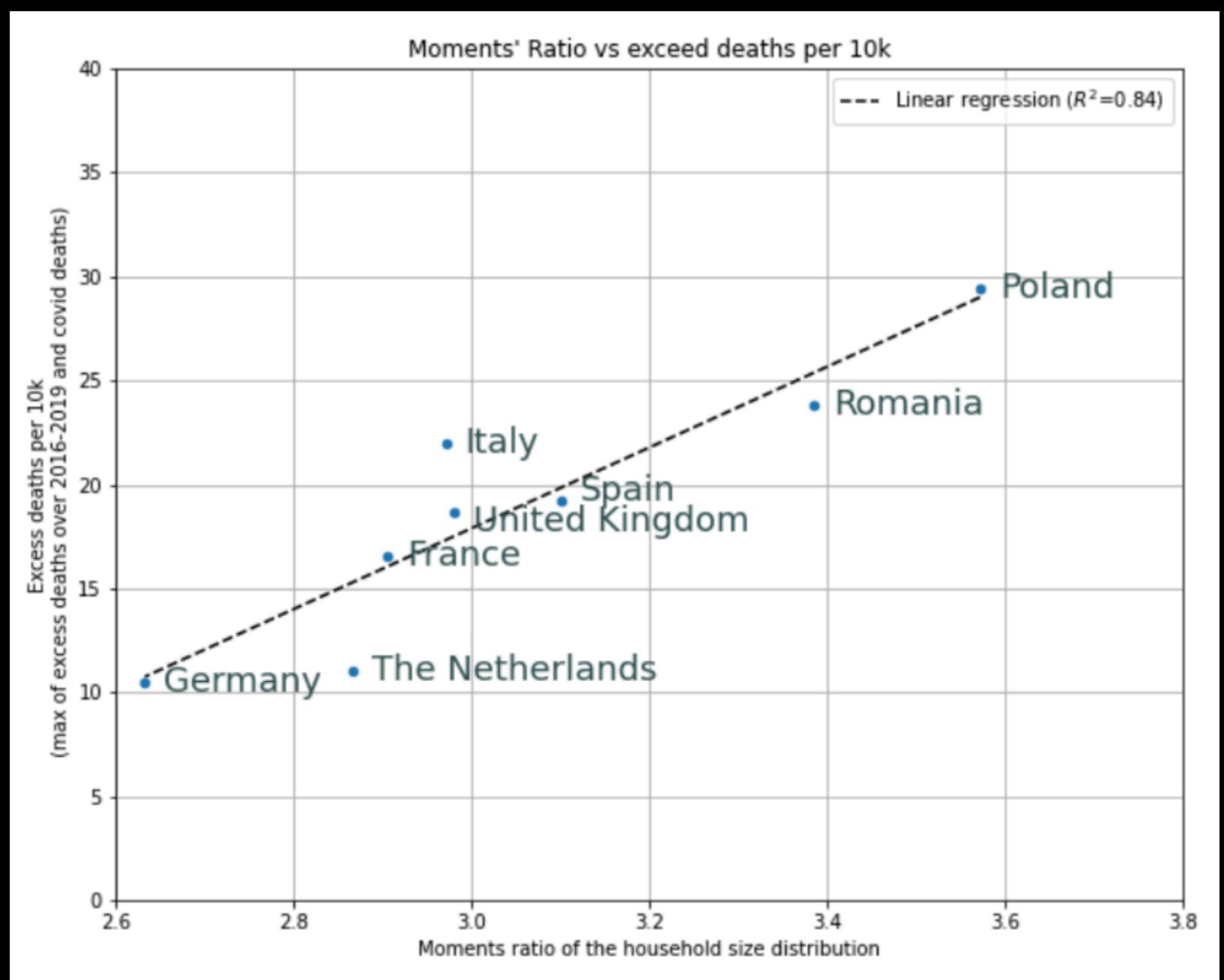
# Household structure plays significant role in epidemics spreading



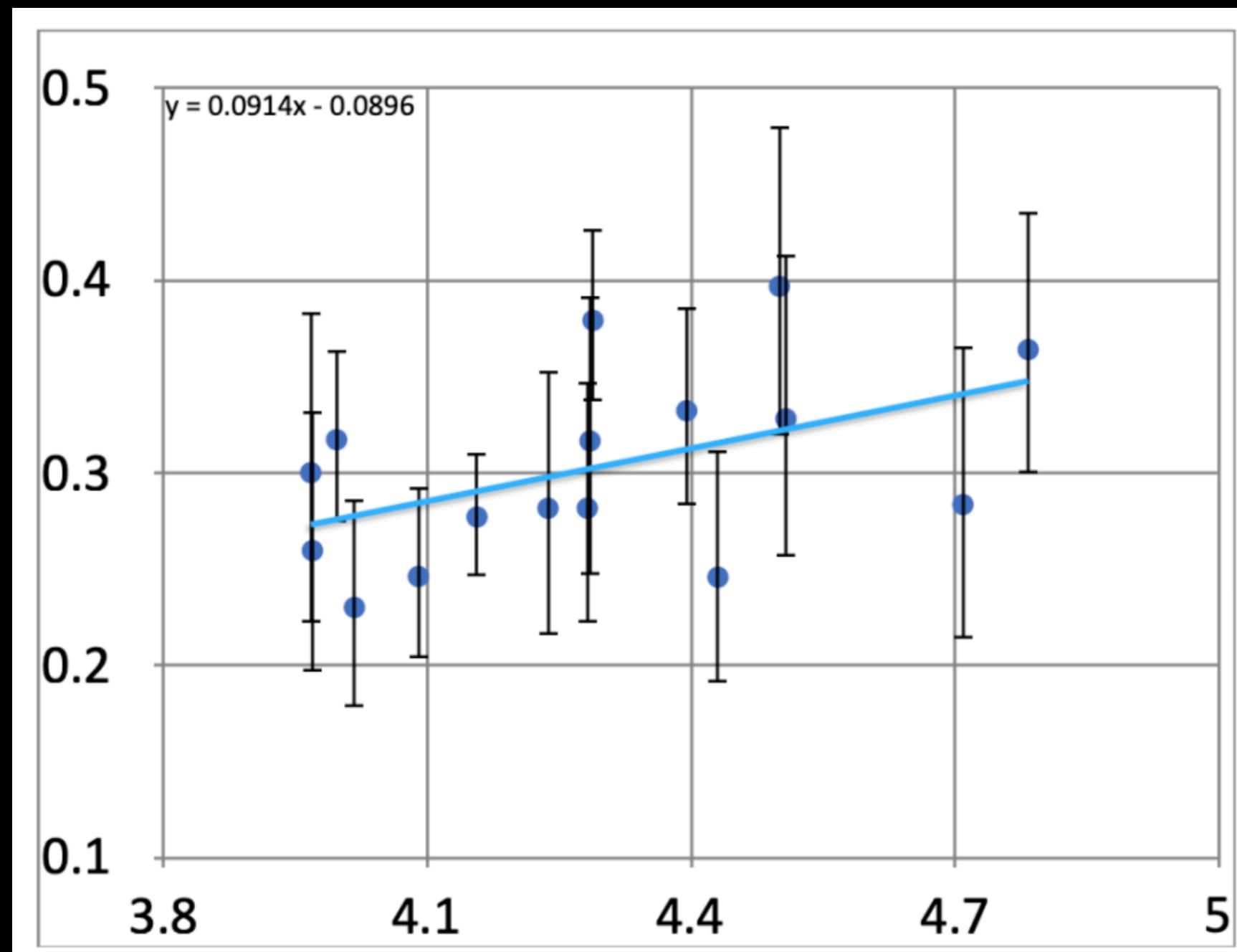
the household  
reproduction number

$$c \left( 1 + a_H \left( \frac{m_2}{E} - 1 \right) \right);$$

## Empirical correlation: excess deaths versus moment-ratio for large countries in Europe (>16mln)



## Empirical correlation: Seroprevalence for various voivodeships (April/May 2021) versus moment-ratio



# The priority queue controls order of executing events



Event-driven simulation



Queue stores time-ordered events



Priority queue can be updated anytime

# More technically about events (kudos to Tomasz O)

## Event structure

- Time (4 bytes)
- subject\_id (4 bytes)
- source\_id (4 bytes)
- event\_kind (1 byte)
- event\_sub\_kind (1 byte)

## Types of events

- Contact (infection contact through one of kernels)
- Disease progression
  - Become infectious
  - Mild Symptoms
  - Severe Symptoms
  - Recovery or Death
- Home treatment
- Hospitalization
- Detection
- Quarantine extension
- Tracking

# „Become infectious” break-down of the event

- Scheduled when Contact event happened according to sampled incubation time from progression vector
- Enqueue possible transmissions through each of contact kernel
- Enqueue possible healthcare detection
- Enqueue further events regarding disease progression depending on randomly attached severity (age-dependent) after calculating if the vaccination was efficient

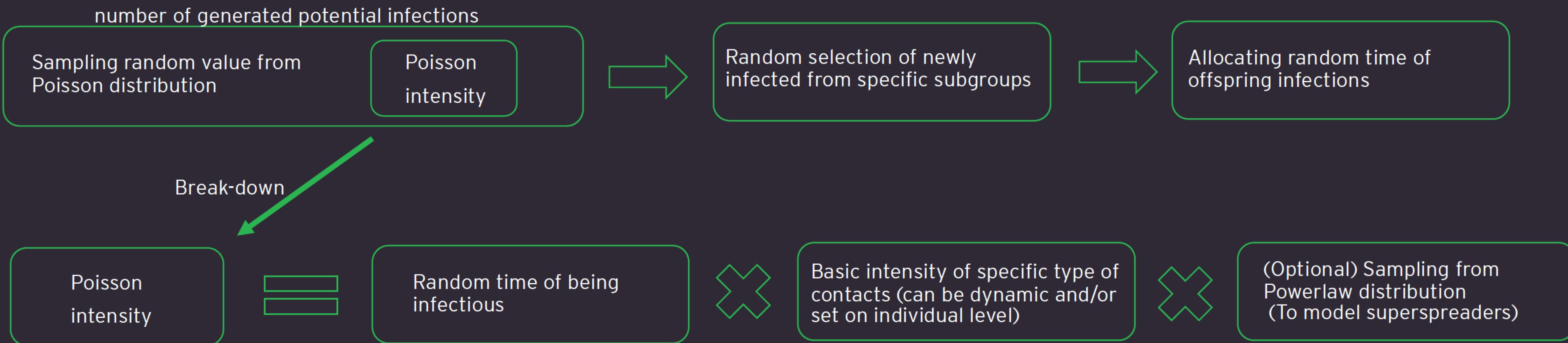
# Precomputing disease progression times to optimize the simulation

(Delay times will be realized under the condition that the infection will happen for a specific person)

- For each person in the sample population, following random delay times are calculated in advance:
  - T0 – incubation time
  - T1 – developing mild symptoms/contact with healthcare
  - T2 – developing severe symptoms/hospitalization
  - $T^*$  - death time realized with death probability correlated with age
  - $T_R$  – recovery time

# Kernels as a way of describing local spread of the epidemics

We model potential infections as easy-to-interpretable stochastical kernels that are generating induced infection-contact networks during the simulation



# Overview of kernels

We model potential infections as easy-to-interpretable stochastical kernels that are generating induced infection-contact networks during the simulation

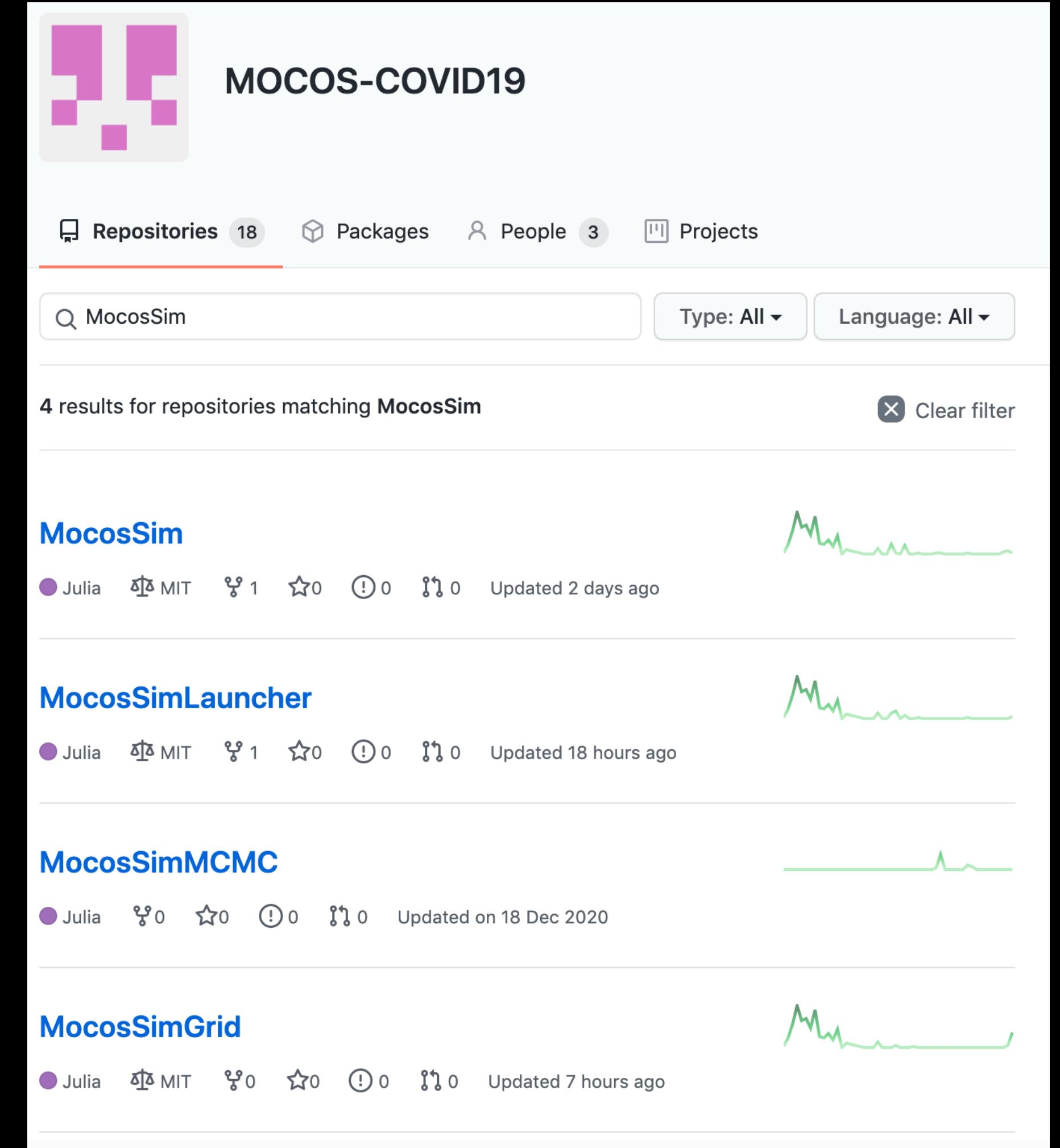
- Household kernel (modelling in-household contacts)
- Constant kernel (modelling sporadic contacts)
- Friendship kernel (modelling contacts within a group of friends)
- Hospital kernel (modelling contacts in hospitals)
- Age dependent kernel (modelling age restricted contacts – e.g. contacts in schools)
- TODO: Workplace kernel  
(depending on individual social features like “workplace id” or alternatively on the binary variable “actively working?” and categorical or numeric feature “size of workplace”)
- TODO: Geographical kernel  
(modelling contacts in specific groups within geographical boundaries)

# Current implementation is written in Julia example of data structures and algorithms used in the project 1/2

- FixedPointNumbers

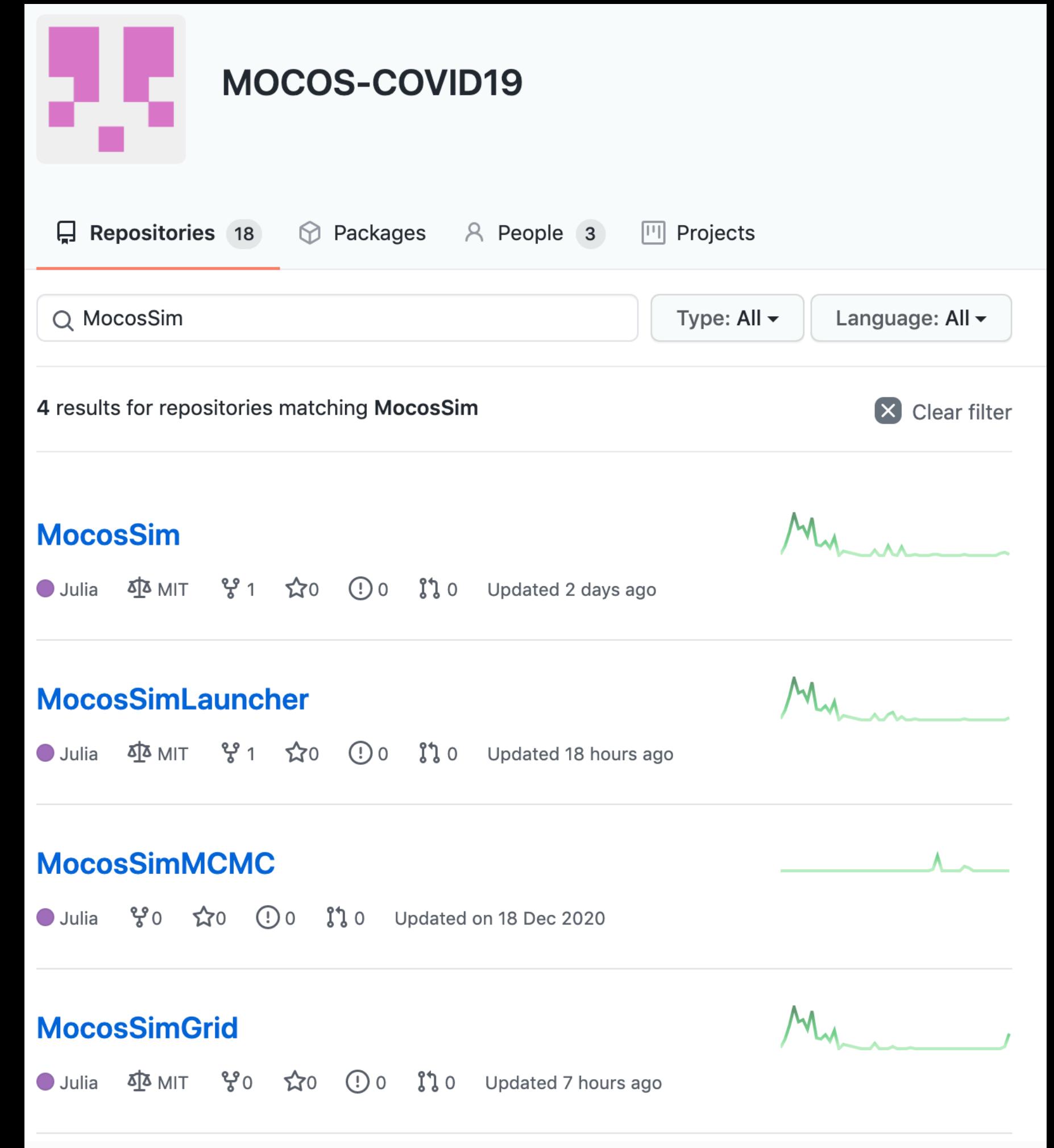
Represents a fractional, or non-integral, number. In contrast with the more widely known floating-point numbers, with fixed-point numbers the decimal point doesn't "float": fixed-point numbers are effectively integers that are interpreted as being scaled by a constant factor. Consequently, they have a fixed number of digits (bits) after the decimal (radix) point.  
Example: Fixed{Int32, 16} to store time events.

Main gain: controlled precision, more efficient memory consumption



# Current implementation is written in Julia example of data structures and algorithms used in the project 2/2

- RobinForest (custom implementation) allowing to quickly retrieve forest of infections
- BinaryHeap as a PriorityQueue efficiently storing all events sorted according to time and priority of the event
- AliasSampler – efficient sampling with many discrete outcomes. It is used for sampling people through the friendship kernel



# Infrastructure: Wroclaw Supercomputing and Networking Center (WCSS)

- Cluster BEM: 860 TFLOPS, 74.6GB memory, 1600 Intel Xeon processors with ~22k cores
- Running tasks through global queue manager PBS
- WCSS team has created a dedicated priority queue “-q covid-19” to be used exclusively for our team

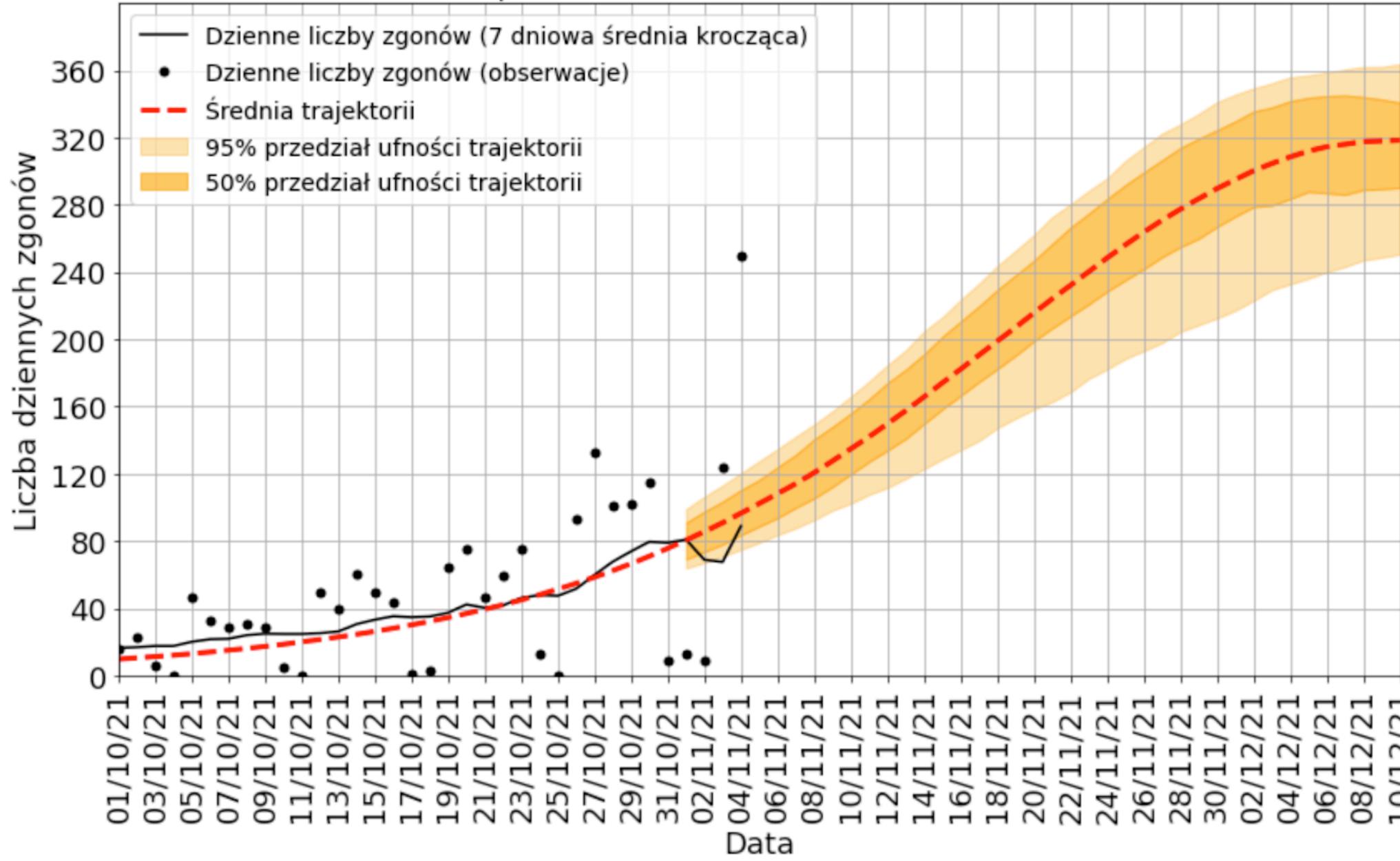
# Infrastructure: Wrocław Supercomputing and Networking Center (WCSS)

- Resources of typical task: generating 100 trajectories for single set of parameters
  - Simulation for Wrocław – 1CPU, 4GB memory, processing time up to 15min, artifacts up to 500MB
  - Simulation for Poland – 1CPU, 32GB memory, processing time up to 16h, artifacts up to 32GB
- Typical experiment: 861 tasks – 2d “grid” evaluating each combination of two parameters, e.g. probability of contact tracking & probability of detecting mild cases
  - Full artifacts from a single typical experiment weigh up to 7TB – require postprocessing on-site

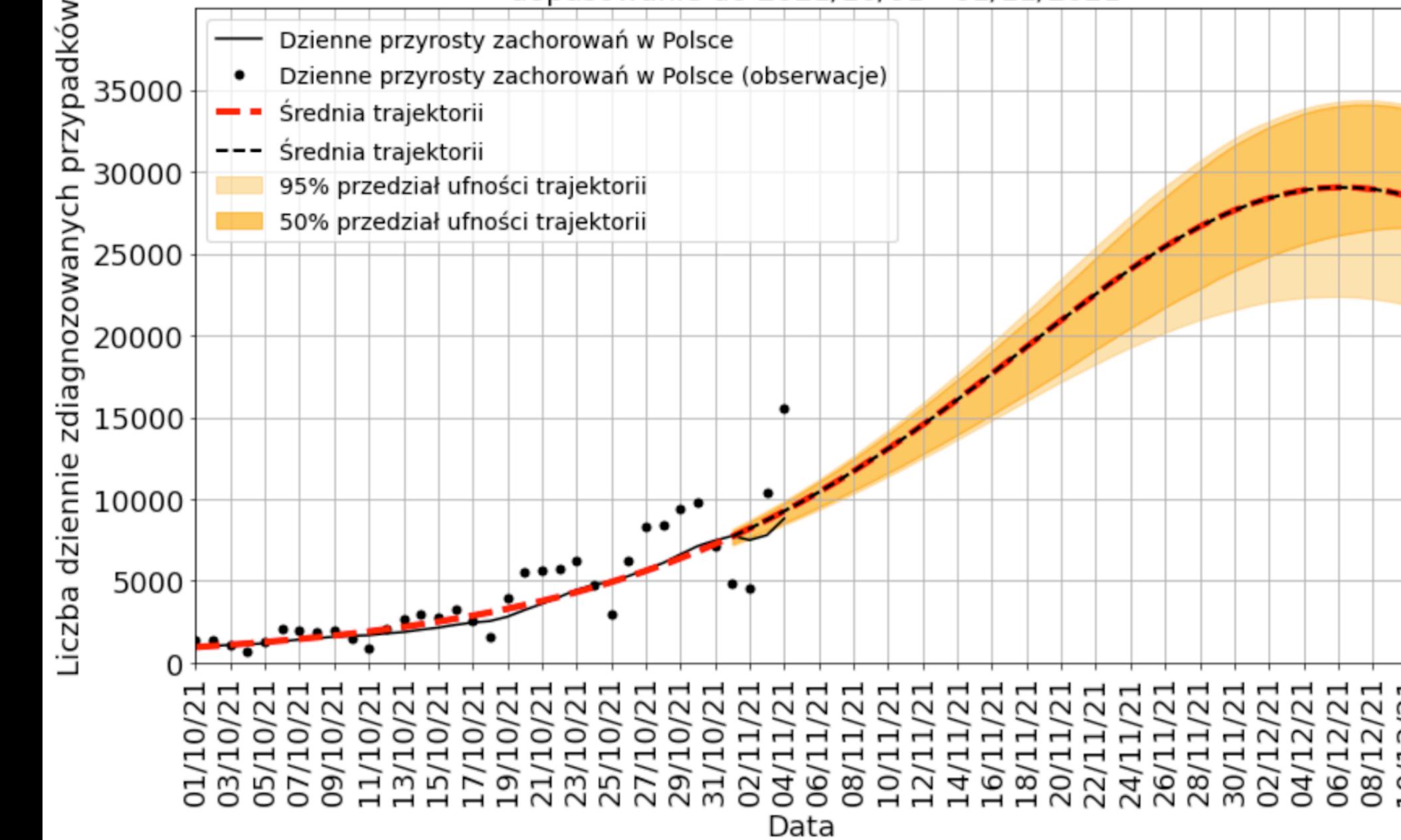
# Outputs from the model

- Trajectories: infections, detections, number of hospitalized, number of active cases, number of quarantined (Possibility to fit stochastic trajectories with observed data)
- Daily trajectories - grouping observables into days
- Age or gender specific trajectories or histograms,
- Disease progressions and outcome,
- Whether specific ICU limit was surpassed within a specific timeframe,
- Confidence intervals for number of reported cases or daily deaths
- The full forest of infections to recreate infection clusters along with infection ways (names of clusters)

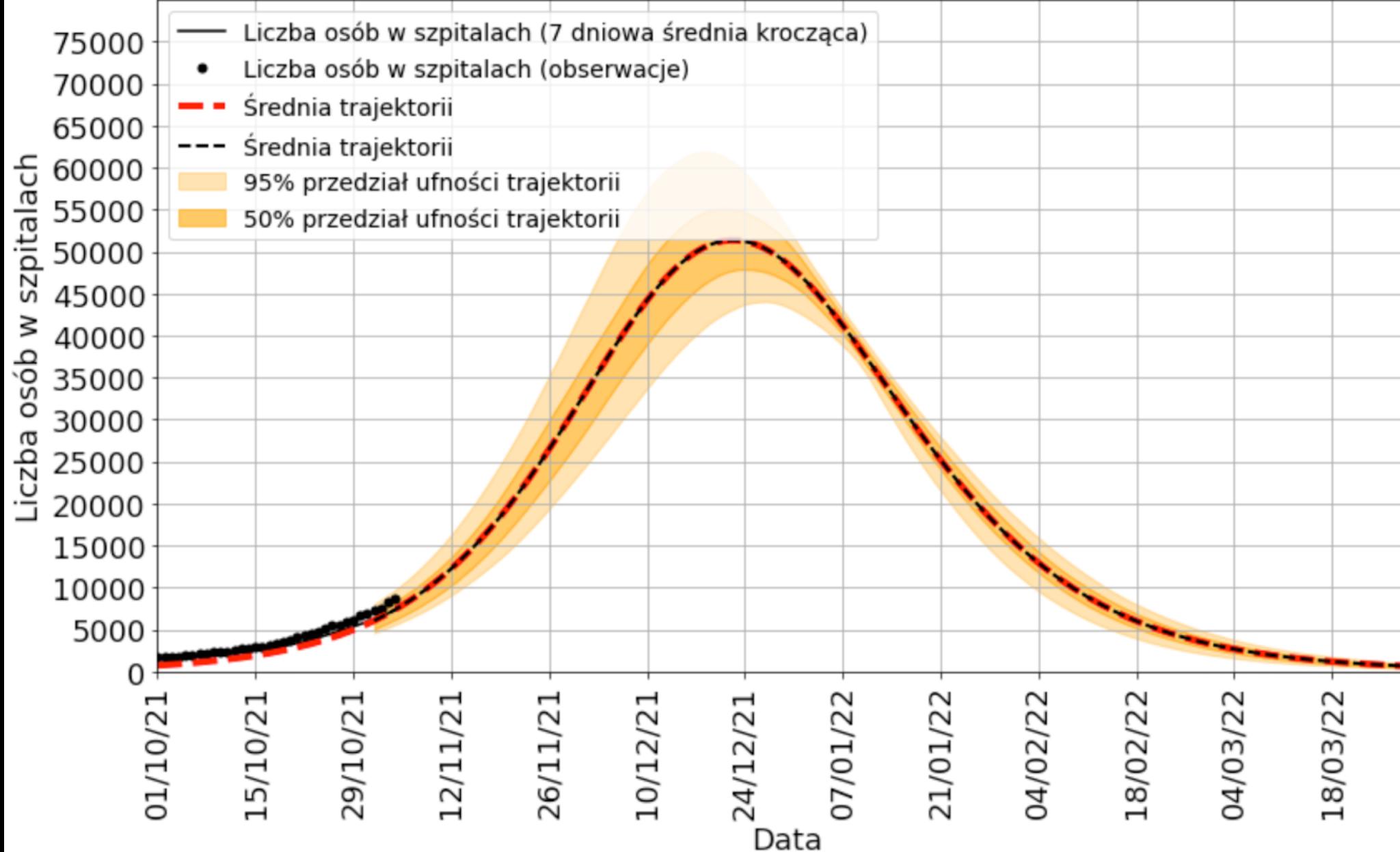
Symulacja rozwoju epidemii w Polsce po 2021/11/01  
- dopasowanie do 2021/10/01 - 01/11/2021



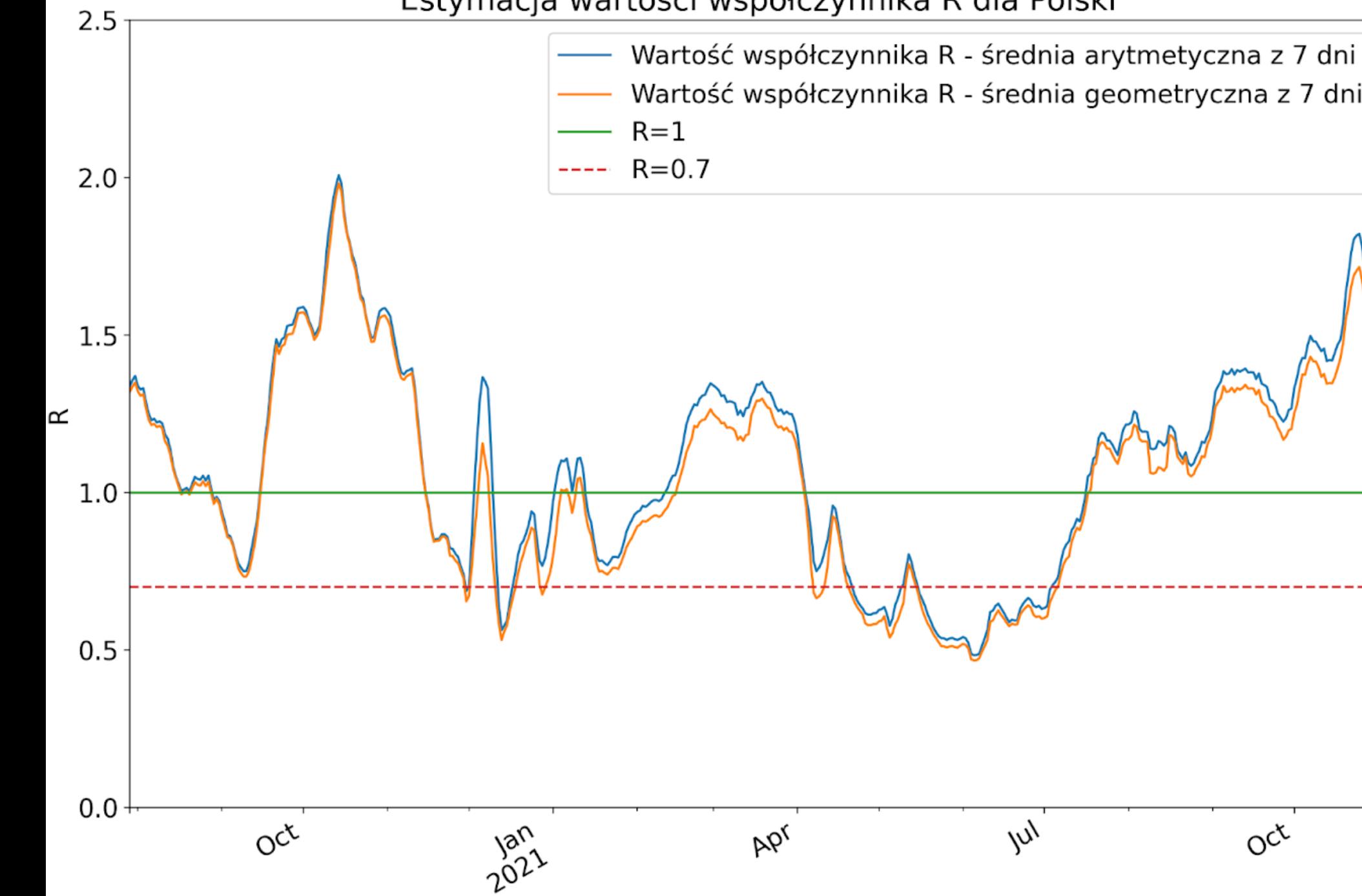
Symulacja rozwoju epidemii w Polsce po 2021/11/01  
- dopasowanie do 2021/10/01 - 01/11/2021

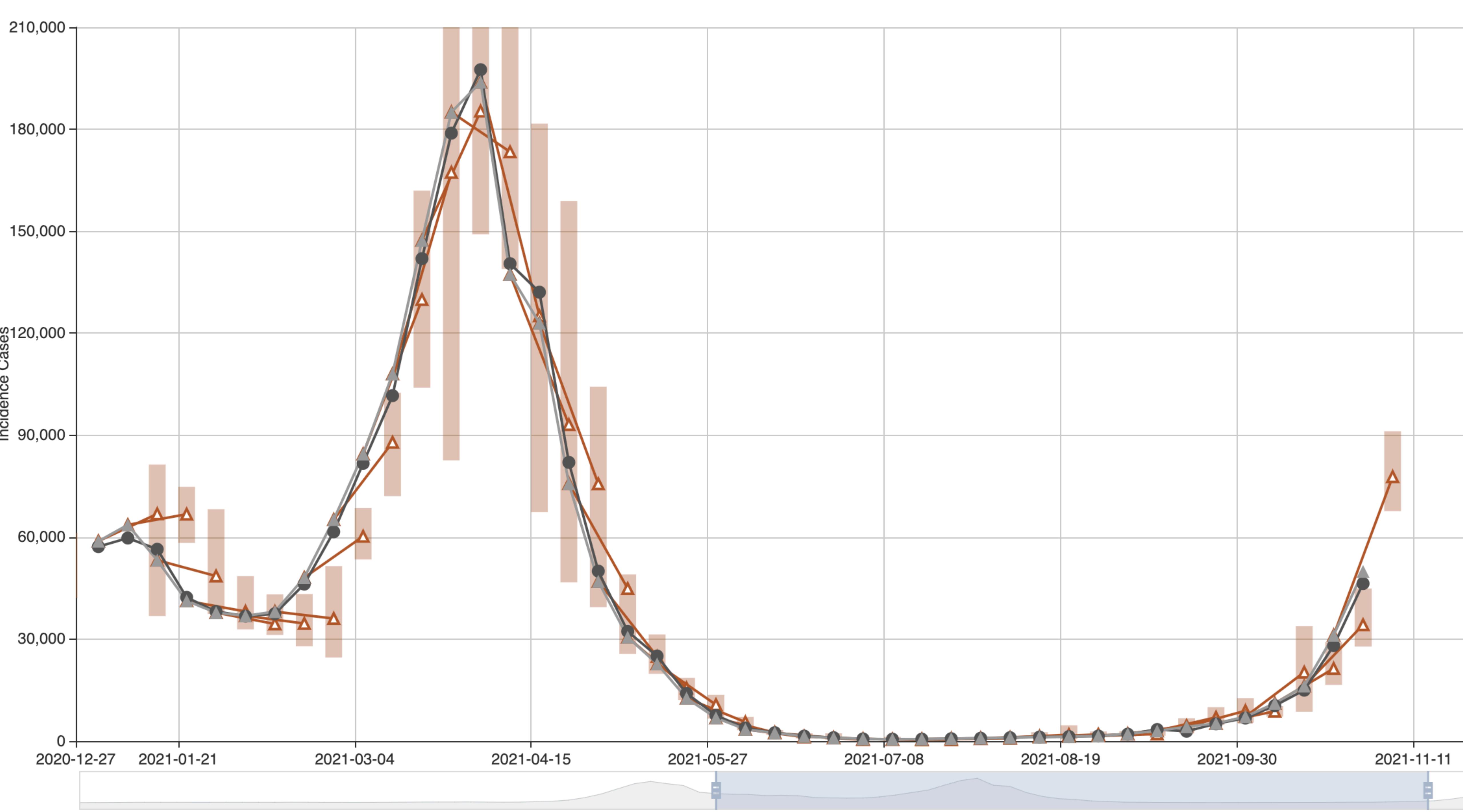


Symulacja rozwoju epidemii w Polsce po 2021/11/01  
- dopasowanie do 2021/10/01 - 01/11/2021



Estymacja wartości współczynnika R dla Polski





# High level remarks

## What we have learned from the model?

- Quick contact tracing: shorten the time to find contacts of diagnosed
- Effective contact tracing: increase success rate for identifying contacts of the diagnosed (we recommend to use mobile data + AI)
- Wide access to testing: helps turn the epidemics sub-critical
- Social distancing and safety: educate the society and encourage to follow the rules and be responsible
- Epidemics lasts longest when oscillating around critical line ( $R=1$ ), therefore effective countermeasures like restrictions and extensive contact tracing should be implemented until the end of epidemics or until the vaccination will be delivered to sufficient number of people

# Python prototype code

## Pros

- Easy for prototyping
- Good support of libraries handling each of required parts of code
- Quick way to create working solution
- Easy to configure profiler can rapidly locate bottlenecks
- Bottlenecks can be improved by custom solutions or rewriting some parts of code to c++ (we applied it to generating random numbers)
- Good python knowledge in the DS community (it is easy to introduce new person to the project)

# Why rewriting Python model to Julia?

- We identified few issues with python model:
  - Implementation was optimized, but still not so quick – simulating single trajectory of overcritical epidemics in Wroclaw lasts about one hour
  - Huge gain was obtained in the meantime by delegating random numbers generation to C++ code
  - Memory consumption was enormous for overcritical Poland runs (up to 50GB)
- Why julia?
  - Julia is the best for statistical applications (probability involved)
  - We have a team member with very deep understanding of julia (he used this language for the sake of his doctoral thesis)
- Full support to realistic settings of contact tracing
- Runtime improved at least 10x
- After further refactor handling artifacts was much lighter

Explain risk of:

- Severe condition
- Death

Observation:

Age  32

Gender  male

Cardiovascular Disease

Cancer

Kidney Disease

Diabetes

Other Diseases   
Chronic diseases, such as liver  
diseases, immunodeficiencies  
(including HIV), chronic lung  
diseases.

Gender: male, Age: 32, Cardiovascular Disease, Other Diseases

[crs19.pl](#)

After diagnosis of Covid-19 disease, the conditional probability of



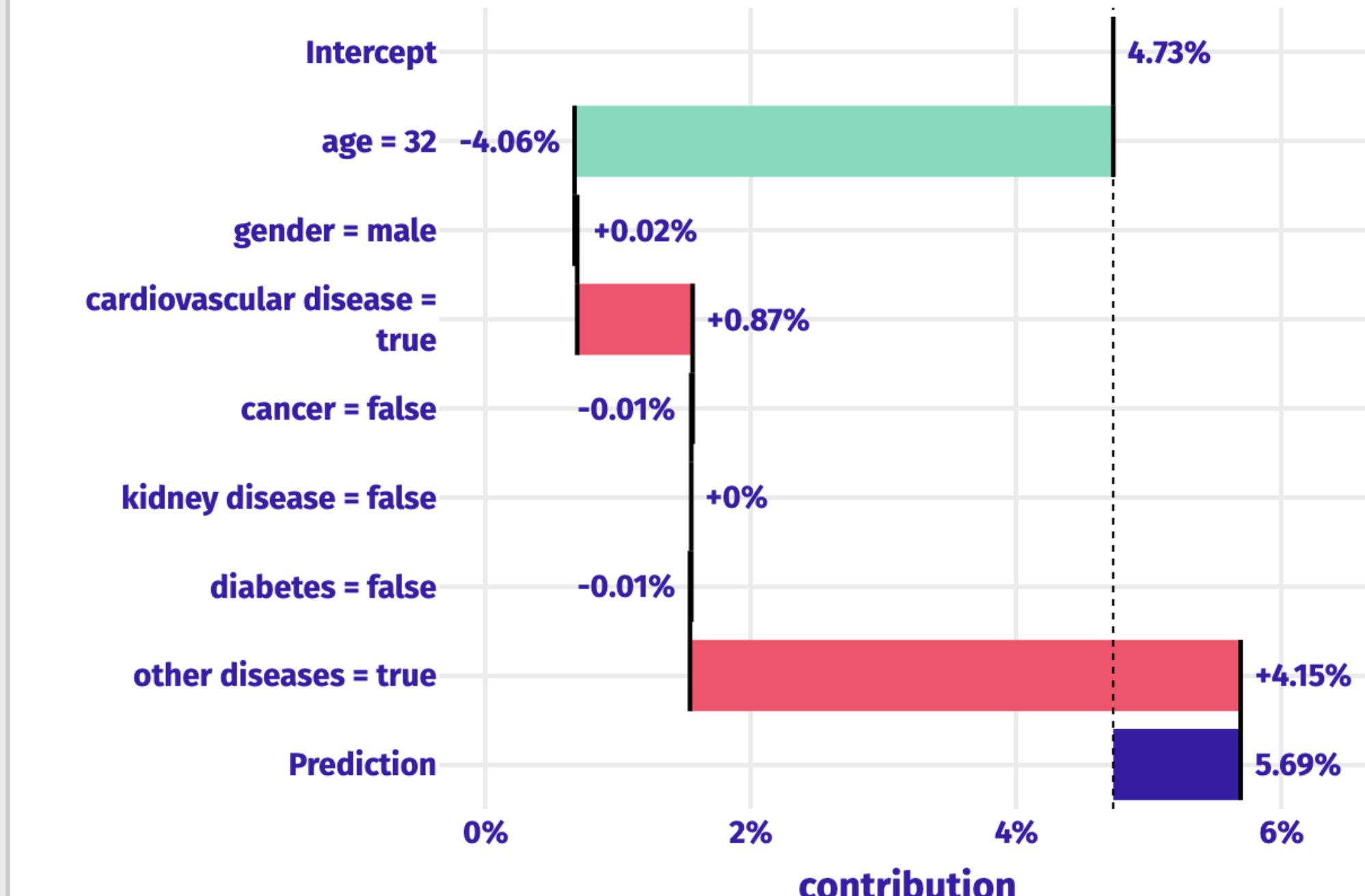
severe condition is  
**5.69%**



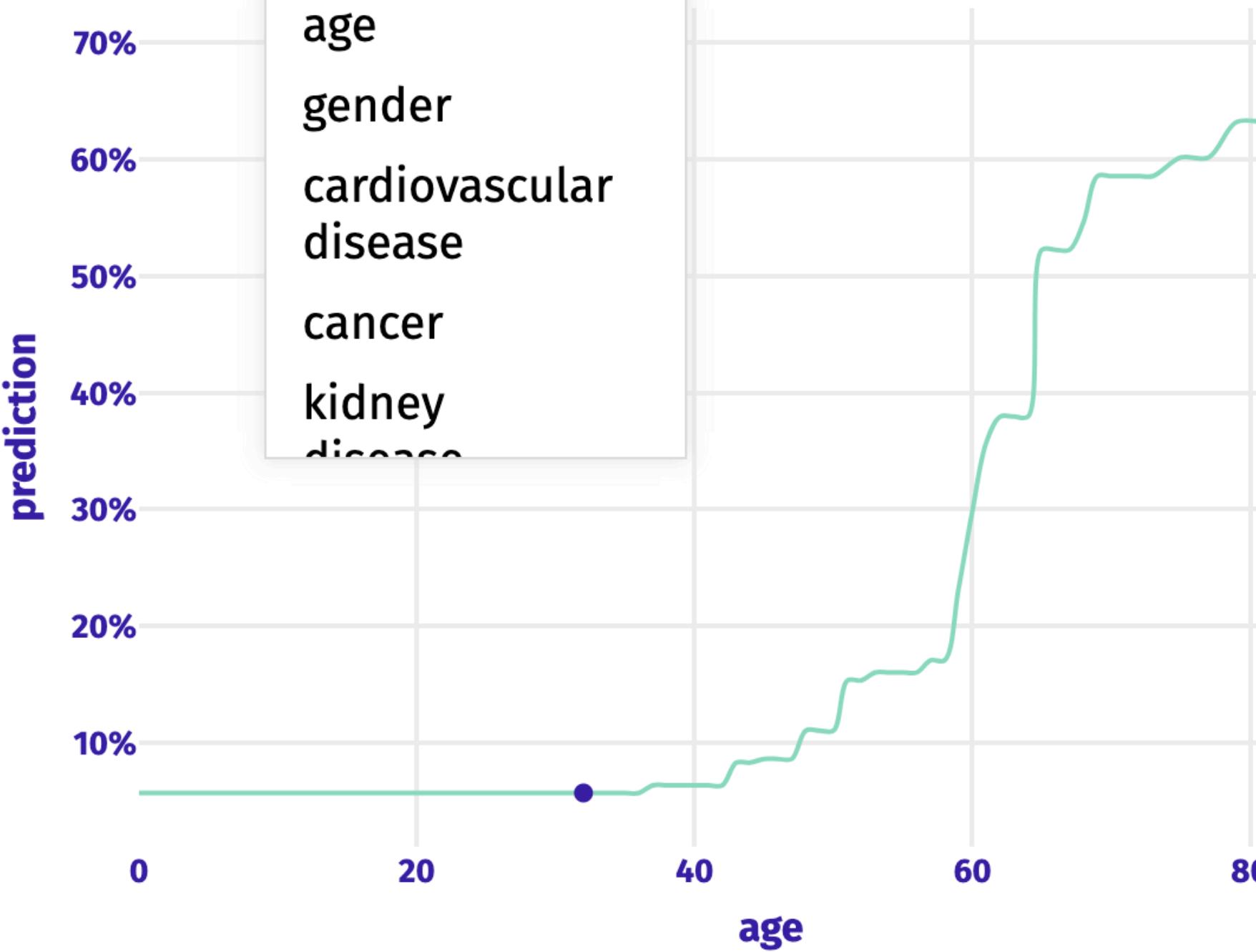
death is  
**0.99%**

## Explain severe condition prediction

### Break Down of your prediction

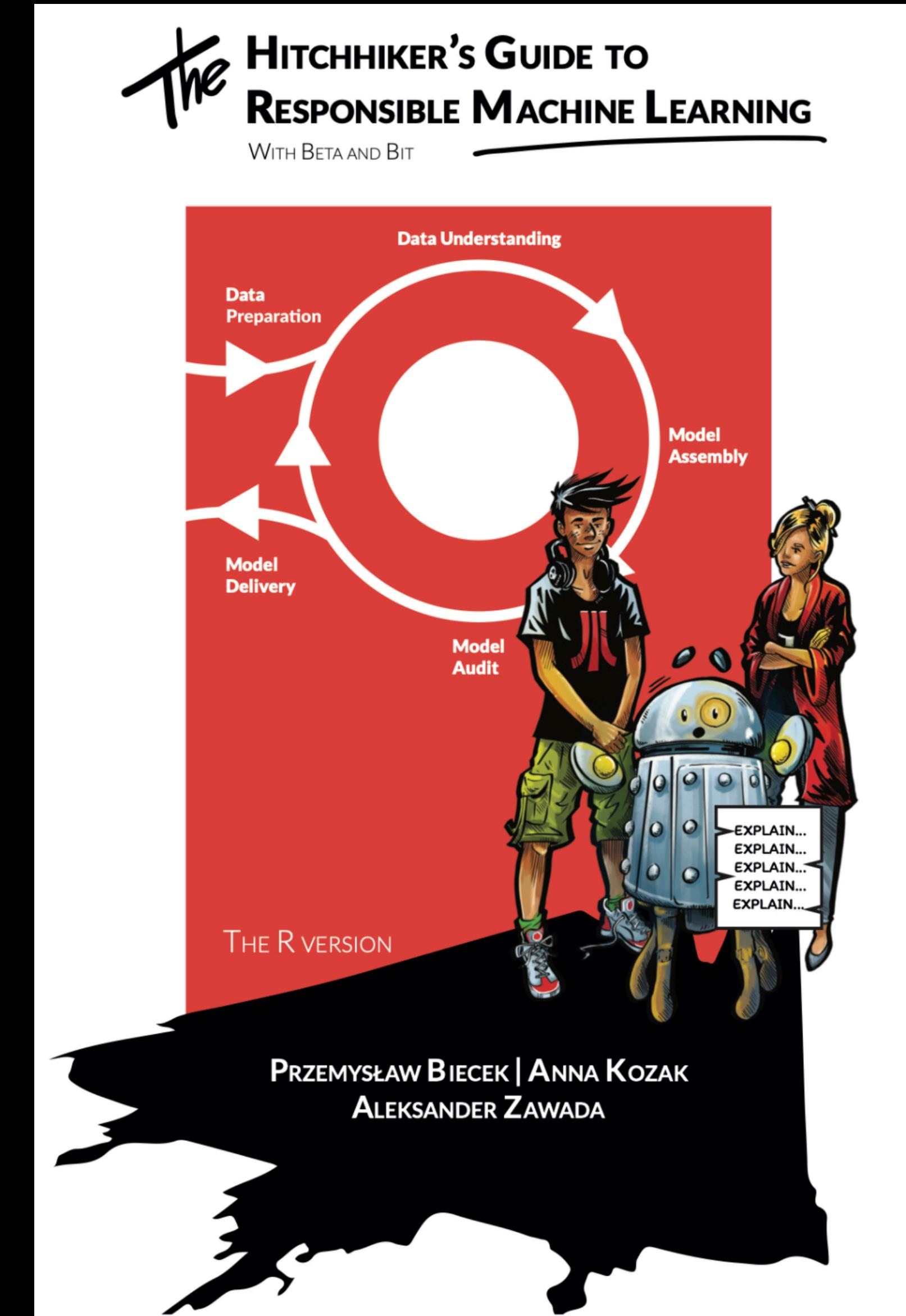


What if  age will change



# The Hitchhiker's Guide to Responsible Machine Learning

<https://betaandbit.github.io/RML/>



# MOCOS group

- Simulations and illustrations based on the model and data analysis: Marcin Bodych, Tomasz Ożański, Marek Bawiec, Ewa Szczurek
- MOCOS Wrocław: Barbara Adamik, Marek Bawiec, Viktor Bezborodov, Marcin Bodych, Maciej Filiński, Radosław Idzikowski, Tyll Krüger, Agata Migalska, Tomasz Ożański, Barbara Pabjan, Ewa Skubalska-Rafajłowicz, Ewaryst Rafajłowicz, Wojciech Rafajłowicz, Piotr Sobczyk, Piotr Szymański
- Correspondence (head of MOCOS group): Tyll Krüger (e-mail: [tyll.krueger@pwr.edu.pl](mailto:tyll.krueger@pwr.edu.pl))
- MOCOS Warszawa: Przemysław Biecek, Tomasz Cąkała, Anna Gambin, Krzysztof Gogolewski, Barbara Poszewiecka, Daniel Rabczenko, Magdalena Rosińska, Michał Startek, Ewa Szczurek
- MOCOS Germany: Wolfgang Bock, Jan Pablo Burgard, Thomas Götze
- MOCOS Philippines: Johniel Babiera, Jinky B. Bornales, Randy L. Caga-anan, Dubhe Joy S. Carmen, Gabriel C. Engcong, Yashika Jayathunga, John Alfred M. Liwanag, Jennifer Joyce, M. Montemayor, Ruth P. Serquiña, Dante Dinawanao

# References

- Mocos webpage <https://mocos.pl>
- Github repository of the MOCOS <https://github.com/MOCOS-COVID19>
- Covid-19 risk calculator by Mi2 data lab <https://crs19.pl/>
- European Covid-19 Forecast Hub <https://covid19forecasthub.eu/visualisation>
- Github repository of the German-Poland forecast hub <https://github.com/KITmetricslab/covid19-forecast-hub-de>
- German-Poland forecast website <https://kitmetricslab.github.io/forecasthub/forecast>

# Few papers where we were co-authors

- A pre-registered short-term forecasting study of COVID-19 in Germany and Poland during the second wave
- Estimation of the Severeness Rate, Death Rate, Household Attack Rate and the Total Number of COVID-19 Cases Based on 16115 Polish Surveillance Records
- Mitigation and herd immunity strategy for COVID-19 is likely to fail
- Assessing the risk of COVID-19 epidemic resurgence in relation to the delta variant and to vaccination passes