Julia Brettschneider

Data Science and Covid-19

1 Hawkes process modelling of the pandemic

2 Simulations about Covid management schemes in schools

3 Microscopic imaging of blood samples in (long) Covid

https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/brettschneider/

Applied statistician/data scientists

Collaborators: engineers, life scientists, clinicians…

Domains: genomics, microscopy, detectors, cancer, screening, finance, OR…

Methodological topics: data quality, spatial statistics, decision theory, concepts of probability and risk

Short biography

- Reader (since 2021), Associate Professor (2010-2021), Assistant Professor (2007-2010), Dept of Statistics, University of **Warwick, UK**
- **Turing fellow** since 2017
- Assistant Professor, Dept of Math/Stats & Dept of Community Health/Epidemiology & Cancer Research Institute, **Queen's University, CN**
- Visiting Assistant Professor and Research Statistician, Dept of Statistics at **University of California at Berkeley, USA**
- Postdoctoral fellow in Computational Biology at **Eurandom, NL**
- **PhD (2001) in Mathematics**, thesis supervisor Prof H Föllmer, **Humboldt Uni Berlin, D**
- Masters in Mathematics (with Computer Sciences and Psychology), thesis supervisor Prof H Föllmer, **University Bonn, D**

Hawkes process modelling of the pandemic

4th year integrated Master's projects by Marianna Maflefthérou and by Adam Davison

Students of MORSE (Mathematics+Operational Research+Statistics+Economics), Department of Statistics, University of Warwick

Supervision Dr Julia Brettschneider

Hawkes Process I

Definition: Point Process

Point processes are a class of random process whose realisations are a set of points on some given space.

i.e. A sequence of random variables $t = \{t_1, t_2, \ldots, t_d\}$ taking values in a subset of $\mathbb{R}^d$.

Definition: Temporal Point Process

Temporal Point process is a point process over time: It describes the occurrence of random events over time.

i.e. a sequence of events $t = \{t_1, t_2, \ldots, t_d\}$ s.t. $0 \leq t_1 < t_2 < \ldots < t_d$

Hawkes process modelling of the pandemic

1. Data from Israel

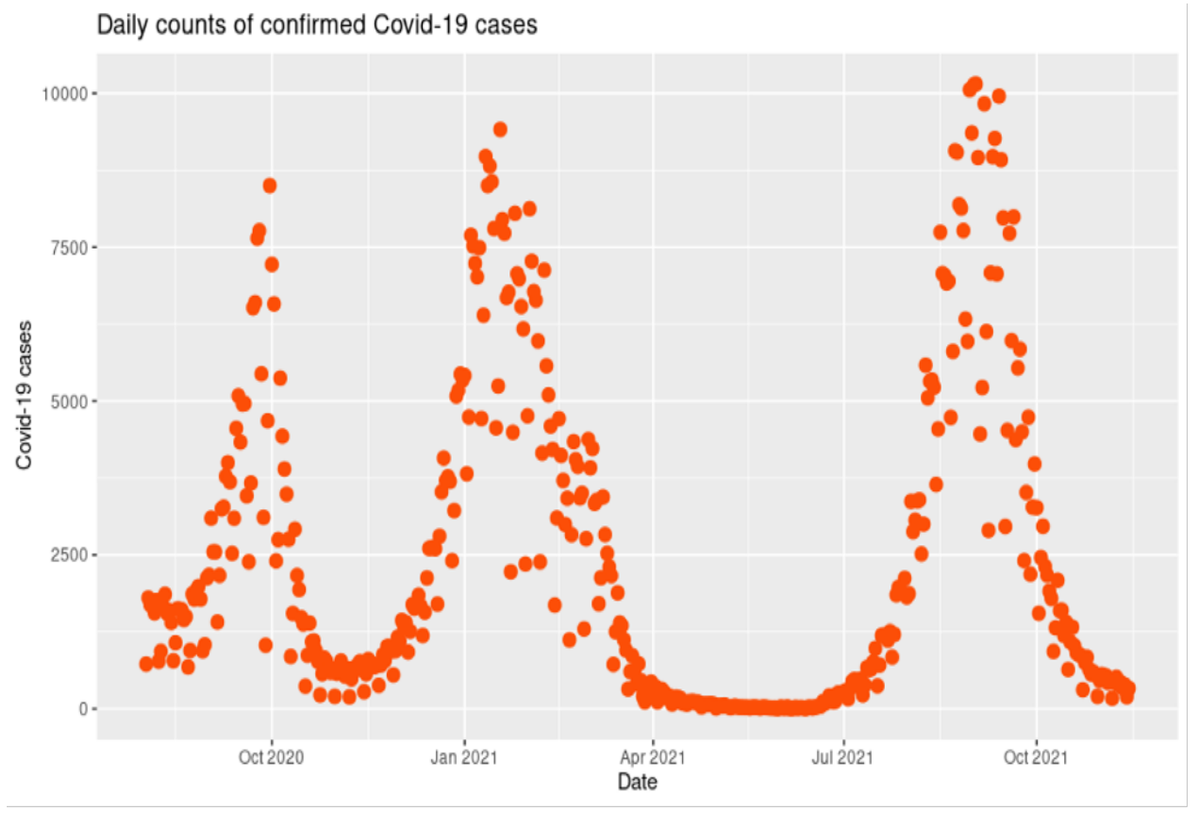
Vaccine for data deal with Pfizer/BioNTech

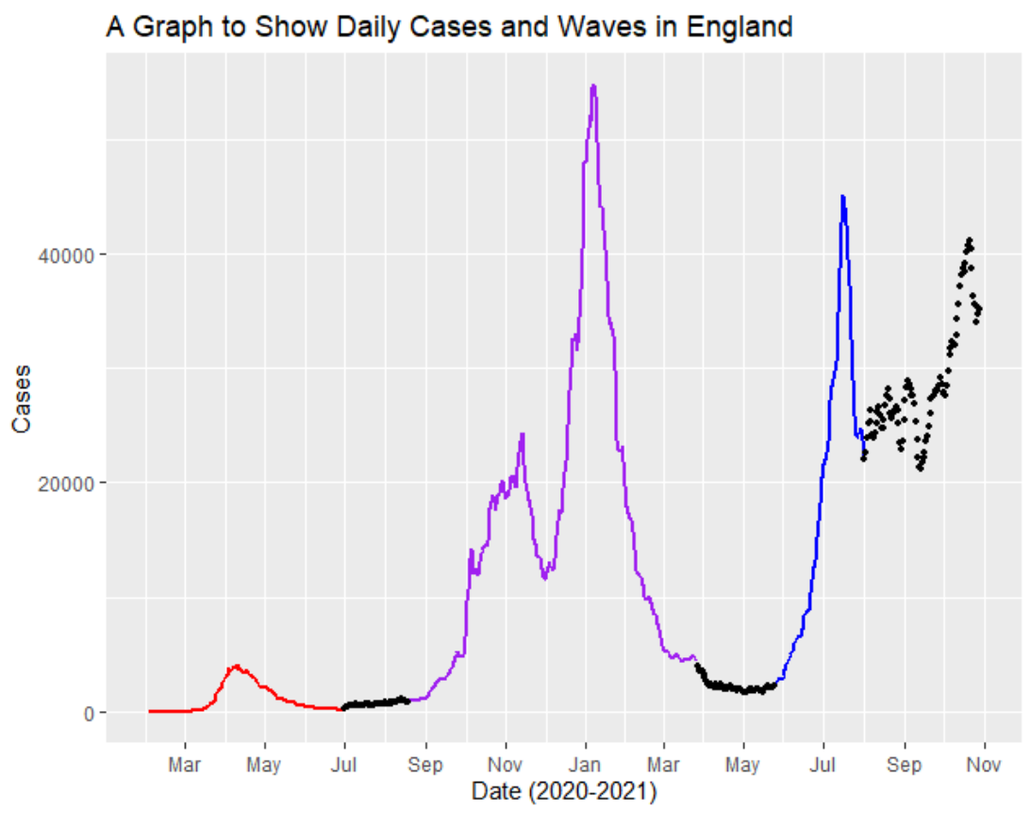
(Marianna Mavrolefthérou's project)

2. Data from England

ONS/Oxford survey

(Adam Davison's project)

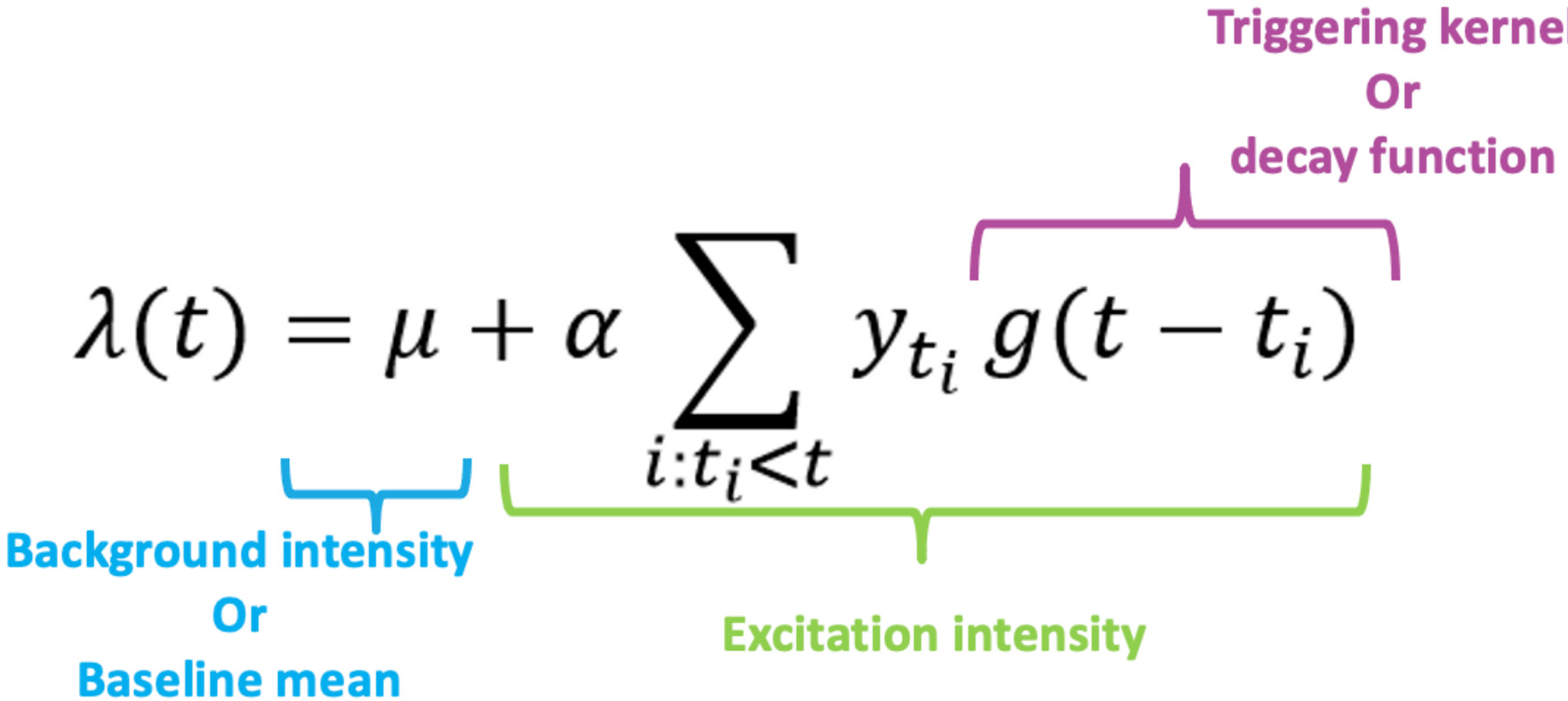




Hawkes process modelling of the pandemic

Hawkes Process

Self-exciting temporal point process described by the conditional intensity function:



Hawkes process modelling of the pandemic

Model Setup

Number of daily cases (or deaths)
[e.g. Rousseau et al 2020]

- It is a natural response to assume that ***Y(t)* follows *Poisson(λ(t))***

Since Poisson distribution is governed by one parameter λ, which is the expected number of times an event occurs in an interval of time or space.

Hence, we set:

$$P(Y(t) = y \mid \lambda(t)) = \frac{\lambda(t)^y e^{\lambda(t)}}{y!}$$

- Further to this:

we choose the triggering kernel **g()** to be the **geometric** excitation kernel:

$$g(t - t_i|\beta) = \beta(1 - \beta)^{t - t_i - 1}$$

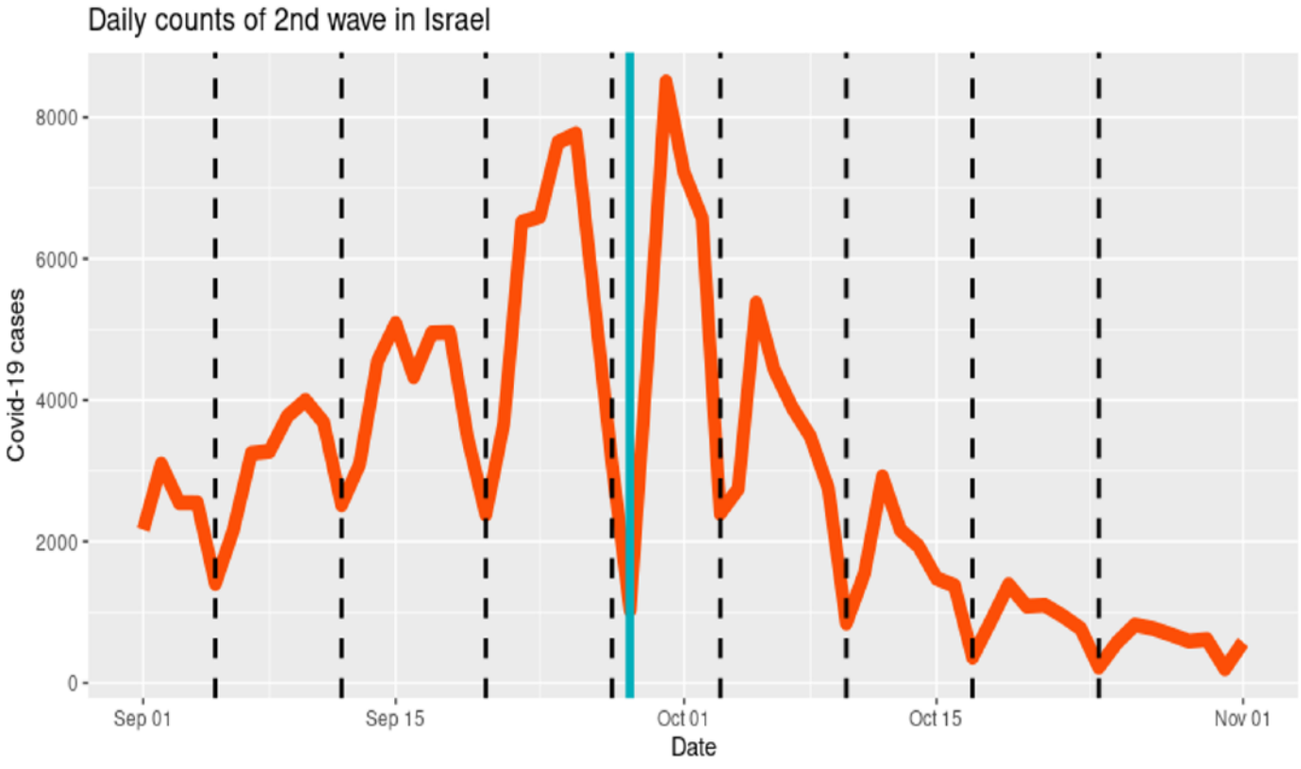
Since it can be shown to be the generalization of the exponential distribution in discrete time.

Hawkes process modelling of the pandemic

Model fitting in practice:

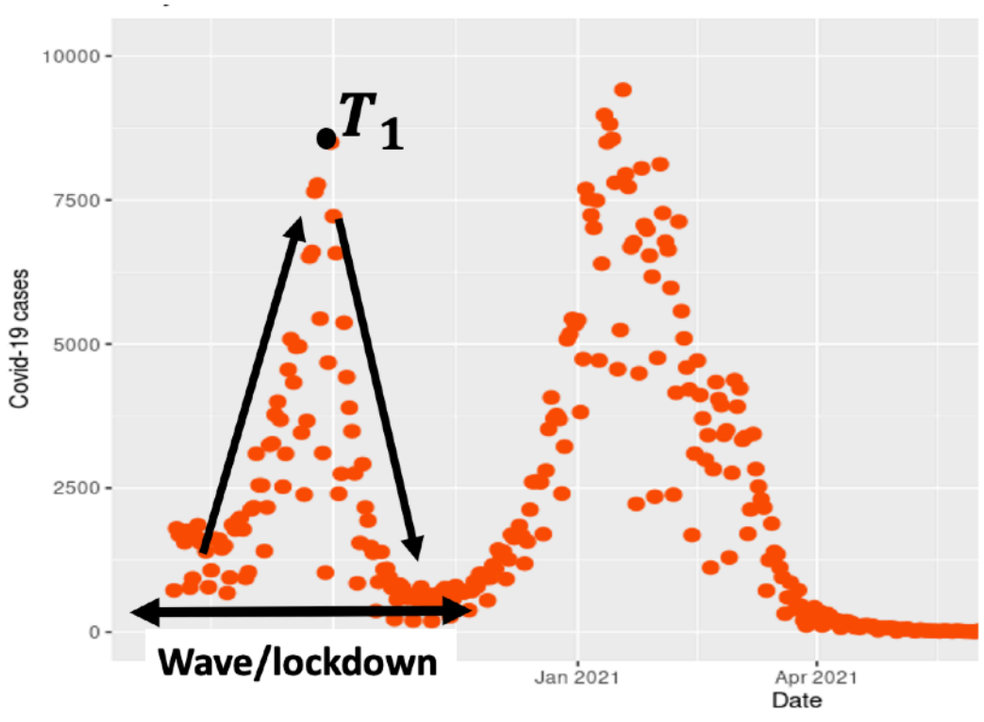
1. Data preprocessing

Small counts, smoothing



2. Waves

Piecewise fitting, piques



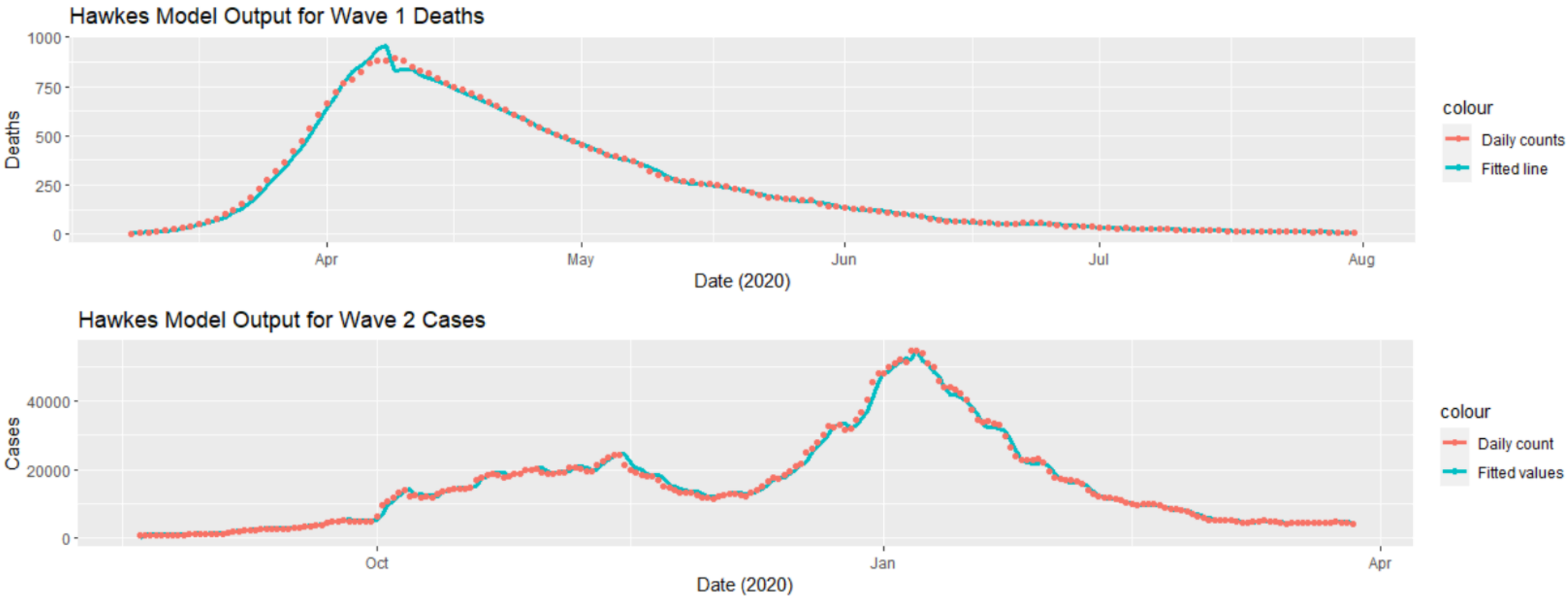
3. Likelihood function

4. Inference

- ML with Nelder-Mead
- Bayesian posterior MCMC

Hawkes process modelling of the pandemic

English data:



Hawkes process modelling of the pandemic

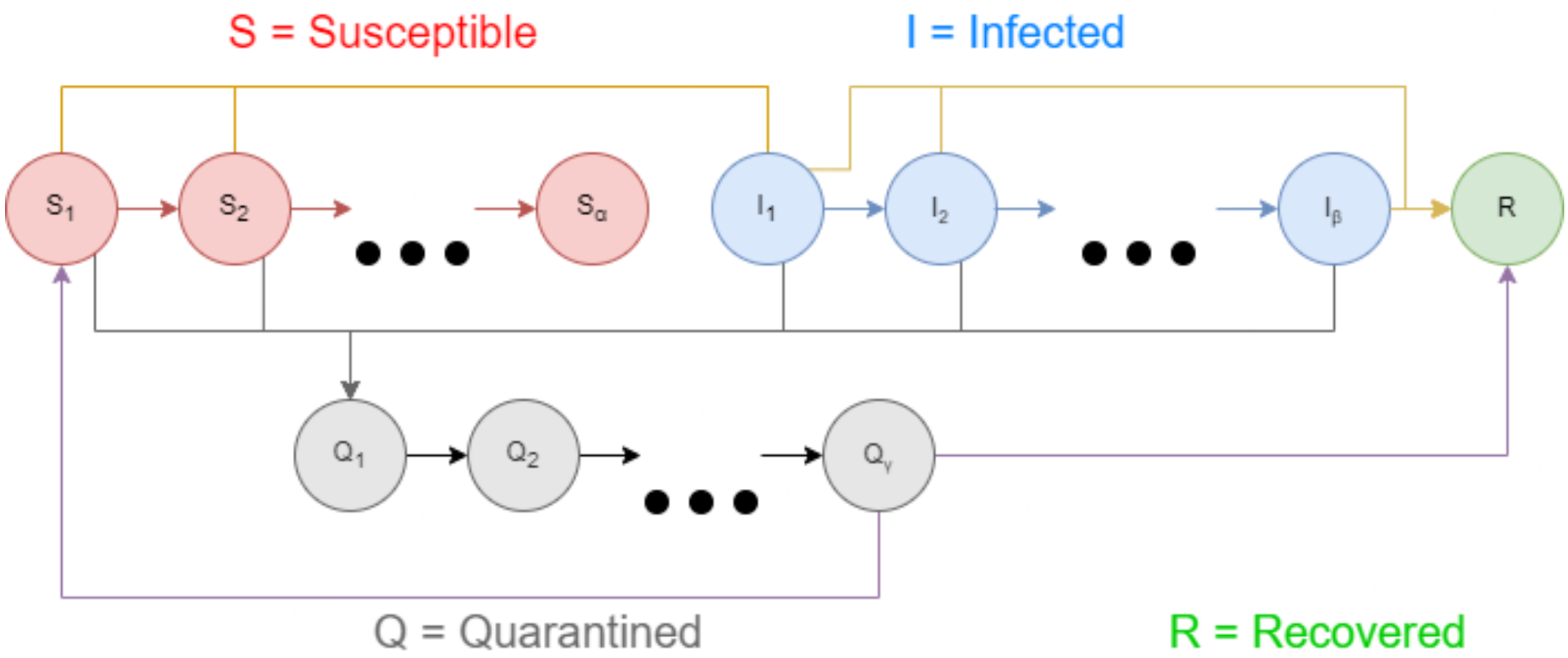
Further and future work:

- Models including age
- Including vaccination rates
- Testing coverage/reliability
- Behavioural indicators (e.g. google searches or mobility)
- Lockdown effects
- Events (e.g. football, holidays) and interaction with other factors
- Mixture population
- Regional models

Simulations about Covid management schemes in schools

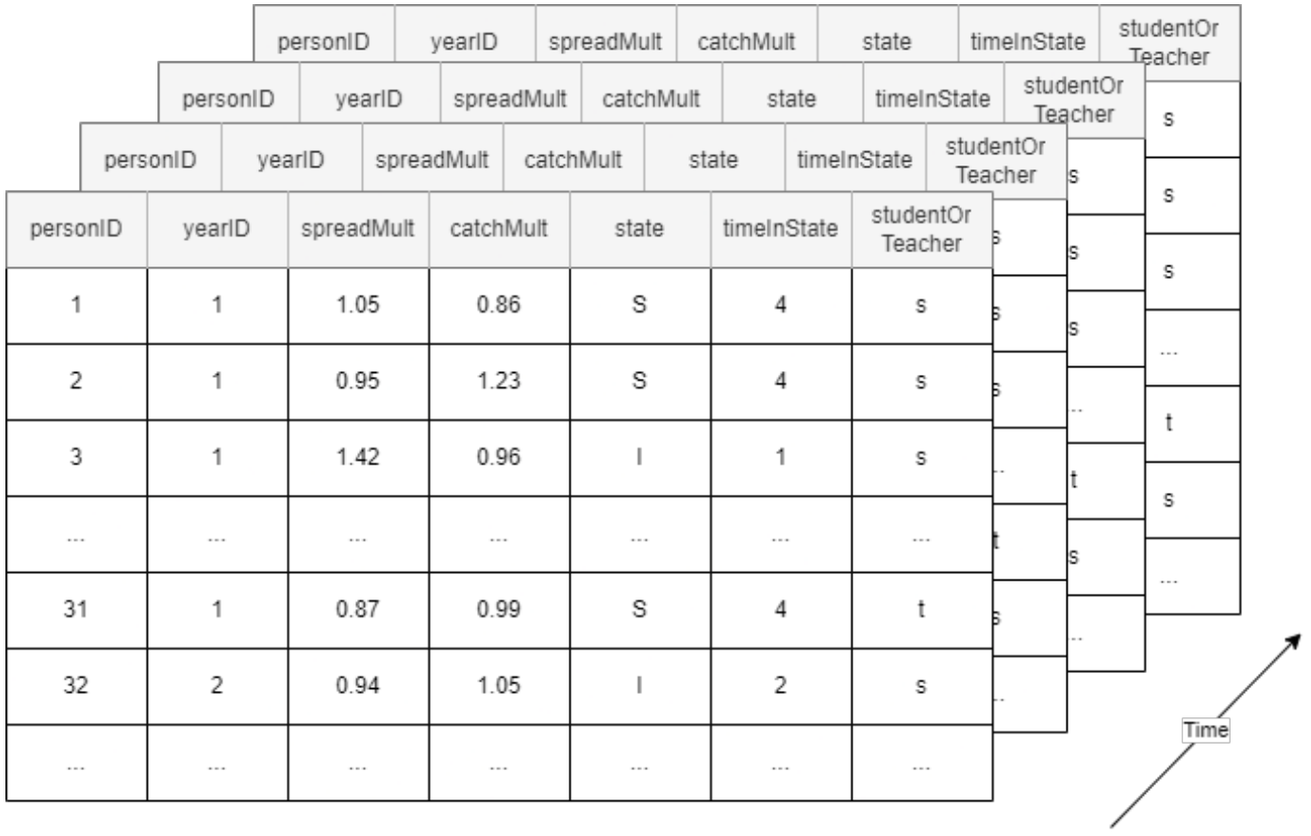
3rd year project by Jonathan Cordell
Data Science student, Department of Statistics, University of Warwick
Supervision Dr Julia Brettschneider

State Flow



Simulations about Covid management schemes in schools

Data Structure



4.3.2 Next state

f(

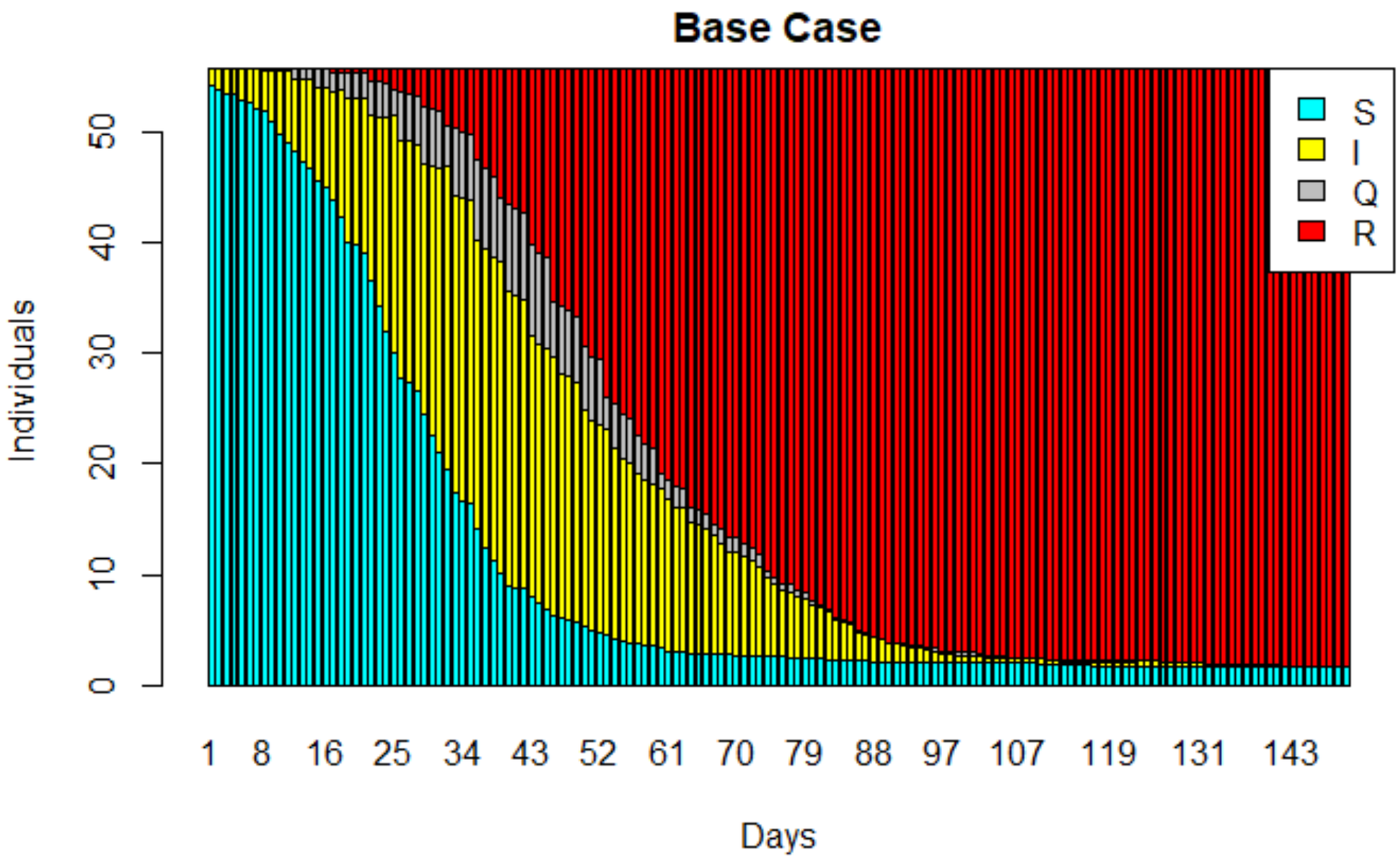
personID	yearID	spreadMult	catchMult	state	timeInState	studentOr Teacher
1	1	1.05	0.86	S	1	s
pID	1	0.95	1.23	S	1	s
...	...	...	...	...	...	...

,pID)

=

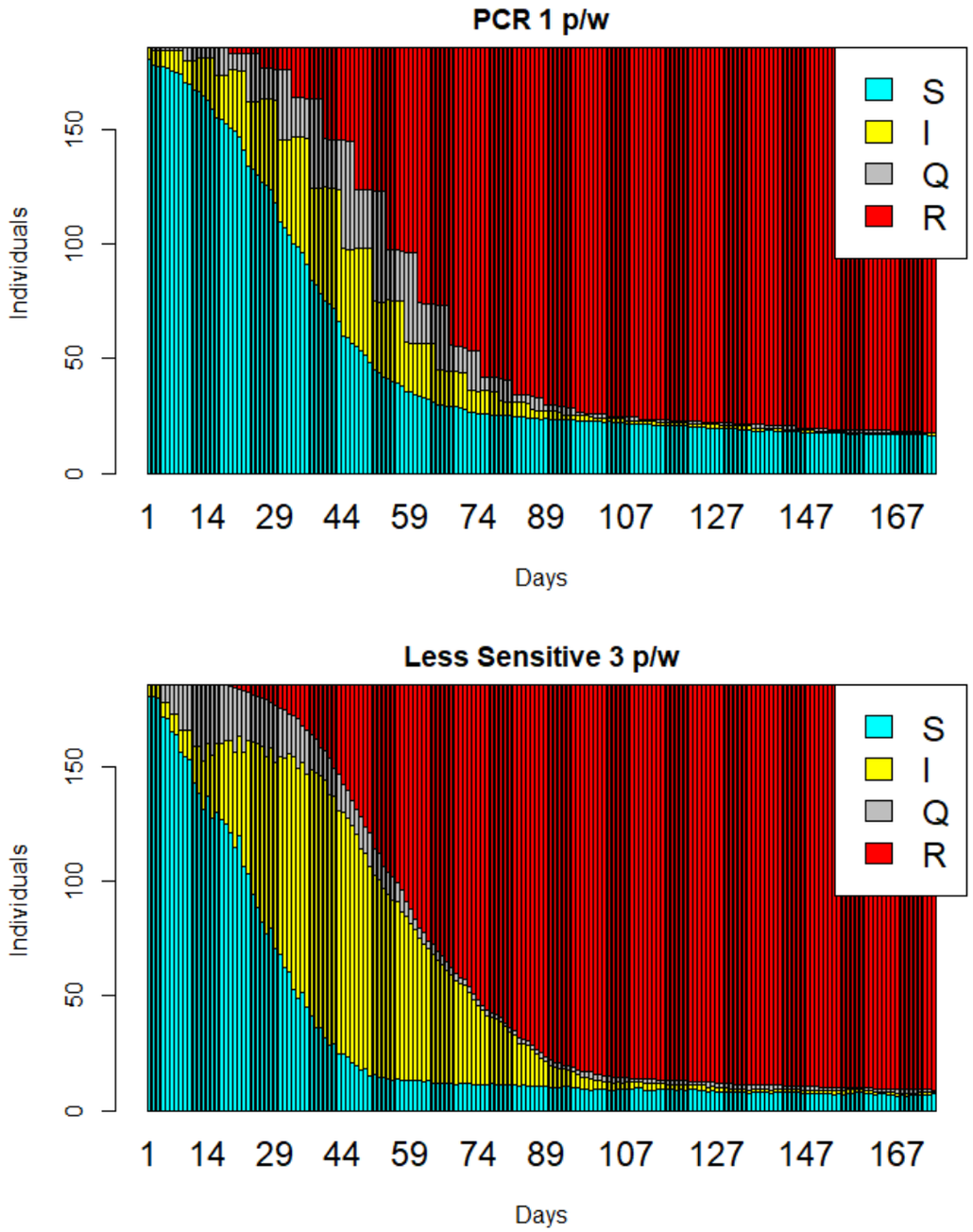
personID	yearID	spreadMult	catchMult	state	timeInState	studentOr Teacher
pID	1	0.95	1.23	S	2	s

Simulations about Covid management schemes in schools



Simulations about Covid management schemes in schools

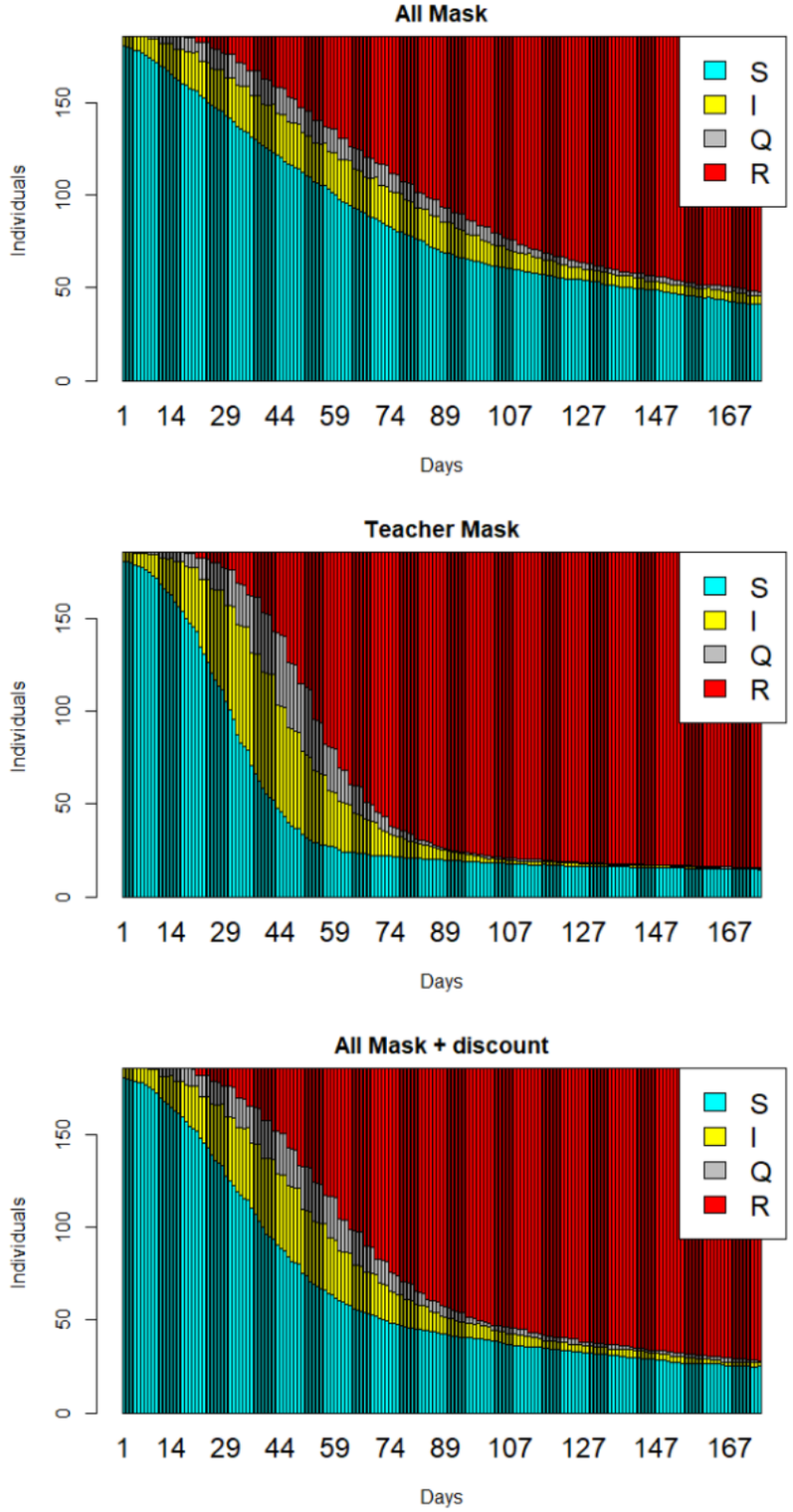
Scenario 1: Testing



name	PCR 1 p/w	Less Sensitive 3 p/w
totalQ	1633.8	1206.3
avgQ	8.78	6.50
finalI	169.2	177.8
infectionEnd	119	NA
overThreshold	16.2	2.9
meanVar	2.0255380	0.9610914
classMax	8.5	5.6
percVaccInfected	0.723913	0.826087
percUnVaccInfeced	0.9142857	0.9742857
totalTests	1756.3	4604.1

Simulations about Covid management schemes in schools

Masks



name	Teacher Mask	All Mask	All Mask + discount
totalQ	1301.7	1145.1	1287.6
avgQ	7.01	6.16	6.91
finalI	171.2	143.4	160.4
infectionEnd	120.9	NA	NA
overThreshold	13.6	0.0	1.2
meanVar	1.3405309	0.6720547	0.9838840
classMax	5.8	4.3	5.3
percVaccInfected	0.7434783	0.4847826	0.6260870
percUnVaccInfeced	0.9414286	0.6628571	0.8250000
totalTests	3114.1	5074.2	4037.7

Simulations about Covid management schemes in schools

Further and future work:

- Vaccinations
- Fitting/comparing with real data
- Case study based on data from a secondary school

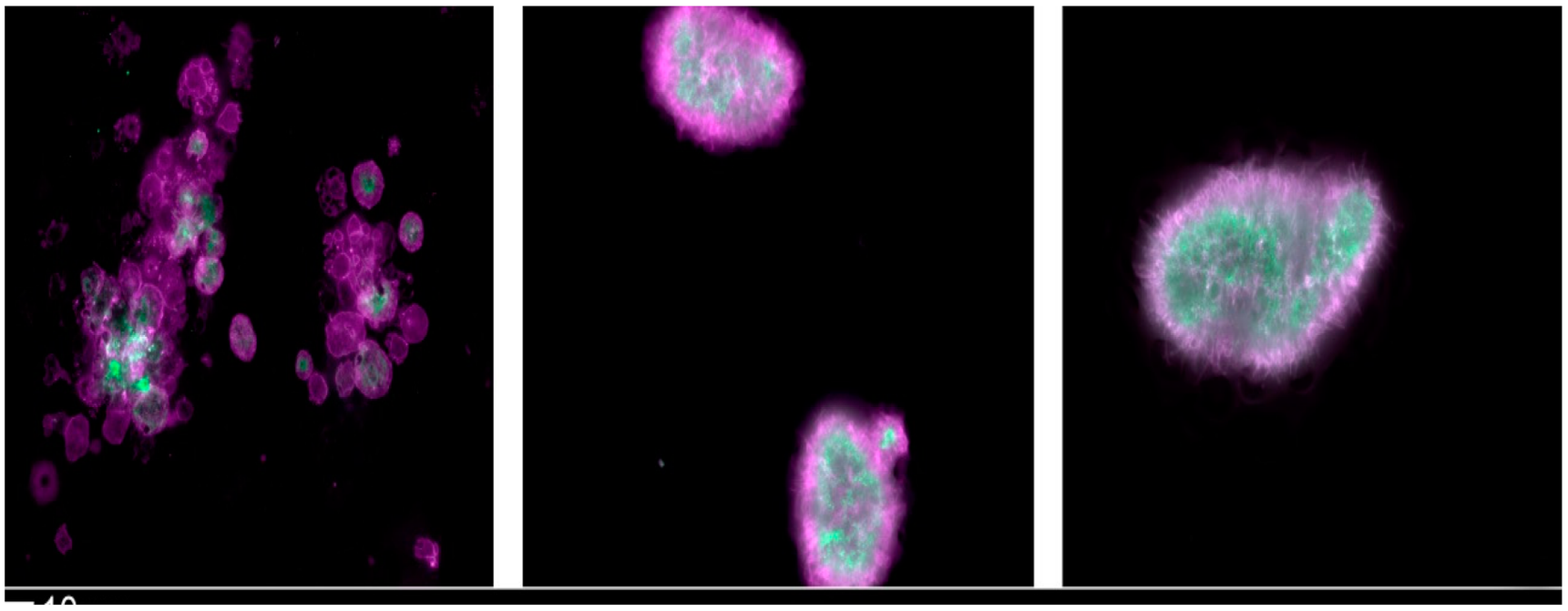
3 Microscopic imaging of blood samples in (long) Covid

4th year integrated Master's projects by Marilena Diamantidou and Grace Barnes

Students of MORSE (Mathematics+Operational Research+Statistics+Economics), Department of Statistics, University of Warwick

Supervision Dr Julia Brettschneider

Data from experimental work by Pretorius et al., 2021



3 Microscopic imaging of blood samples in (long) Covid

Challenges:

- Diagnostic criteria
- Treatment
- Progression
- Timing of intervention
- Automation of diagnosis

3 Microscopic imaging of blood samples in (long) Covid

Challenges:

- Diagnostic criteria
- Treatment
- Progression
- Timing of intervention
- Automation of diagnosis

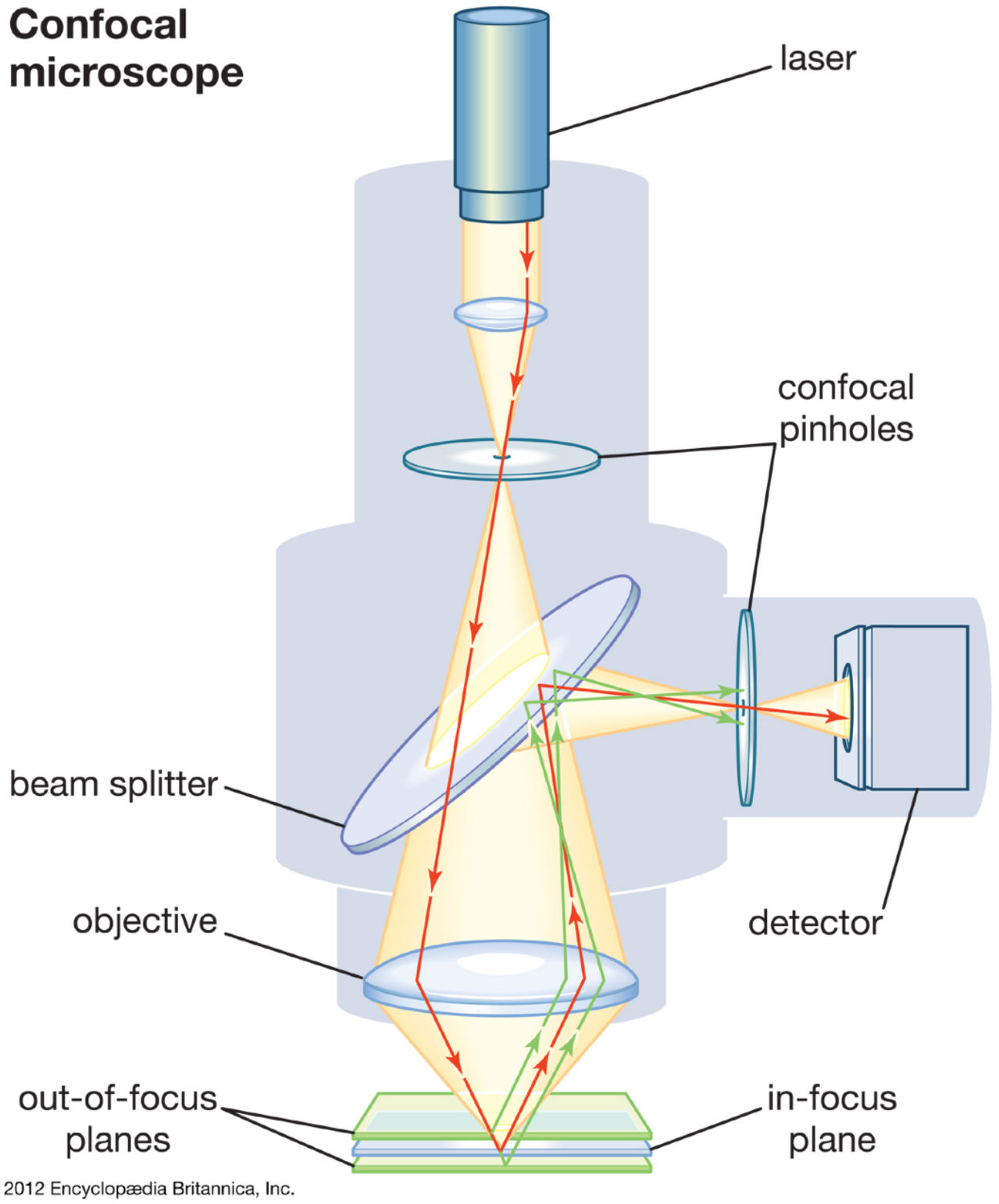
Confocal microscope

Wide-field microscopy:

- All of specimen excited at the same time
- Large unfocused background

Confocal microscope:

- Field of view limited by geometric optics
- Pinhole in front of the detector to eliminate out-of-focus signal
- Long exposure required
- Scanning arrangement to build up image of larger region
- Better resolution

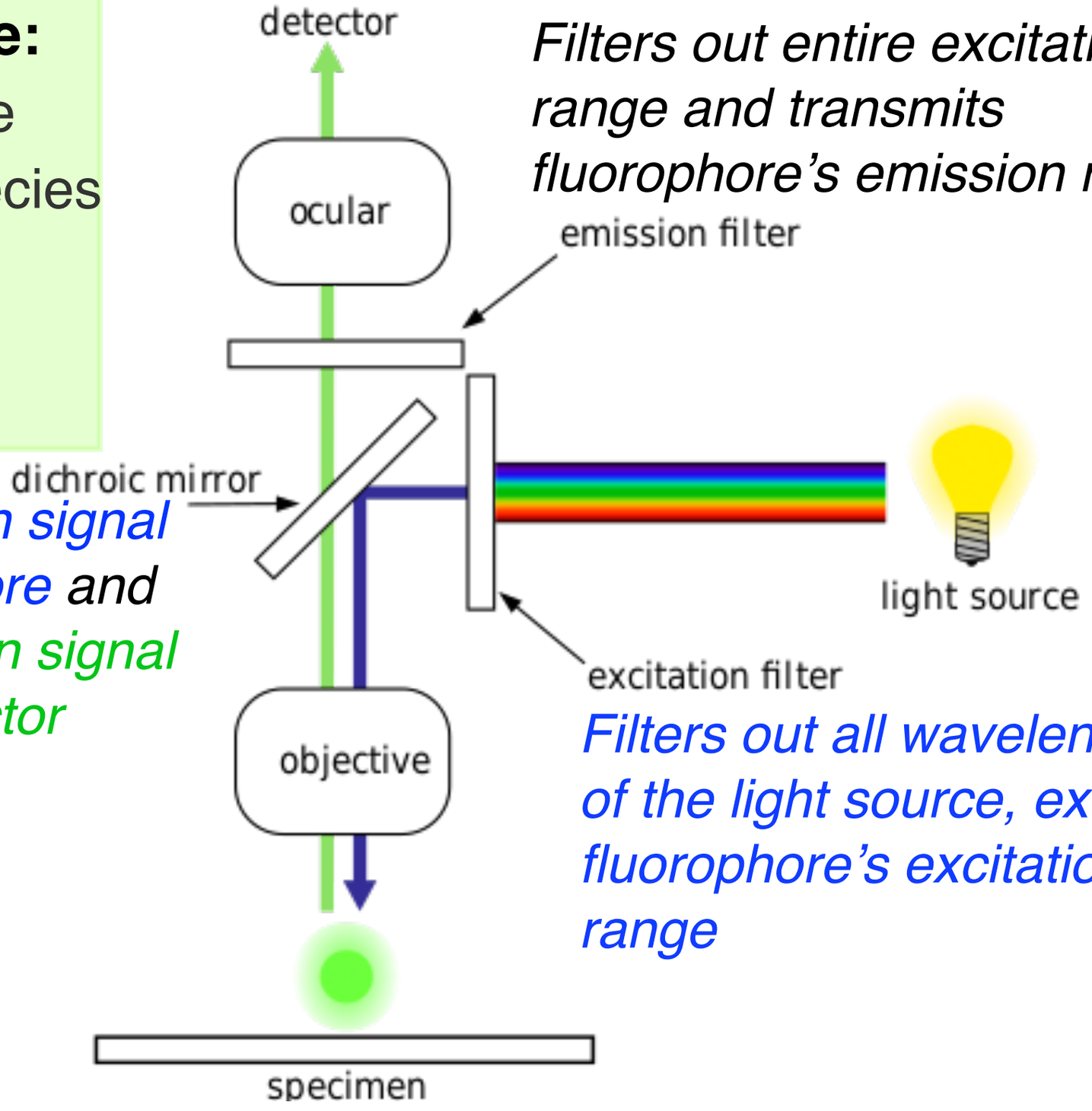


https://www.britannica.com/technology/microscope/Confocal-microscopes

Fluorescent microscope

Fluorescent microscope:

- high intensity light source
- excites a fluorescent species in a sample
- Sample emits different wavelength



Filters out entire excitation range and transmits fluorophore's emission range

Reflects excitation signal towards fluorophore and *transmits emission signal towards the detector*

Filters out all wavelengths of the light source, except fluorophore's excitation range

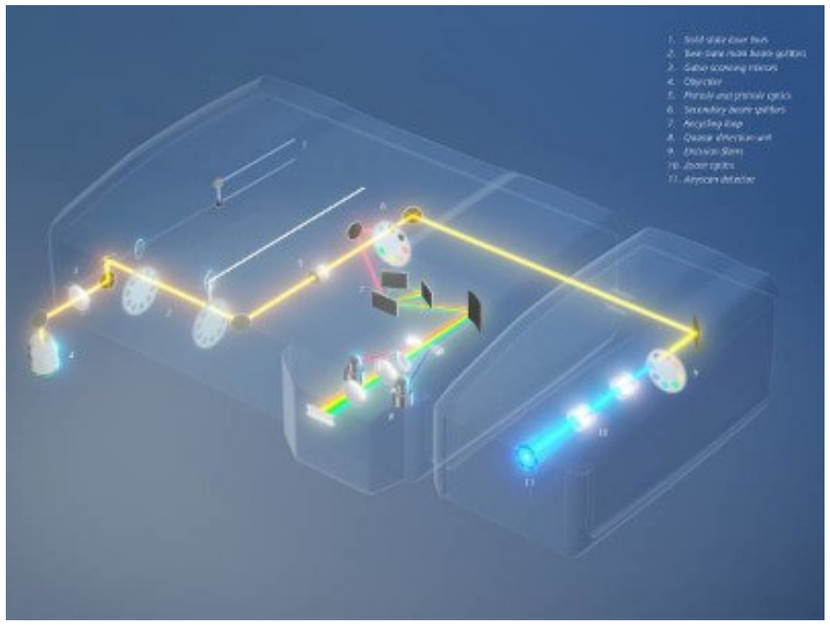
Labelled with fluorescent protein (e.g. GFP)

https://www.wikiwand.com/en/Fluorescence_microscope

Confocal fluorescent laser microscope

Fluorescent confocal microscope:

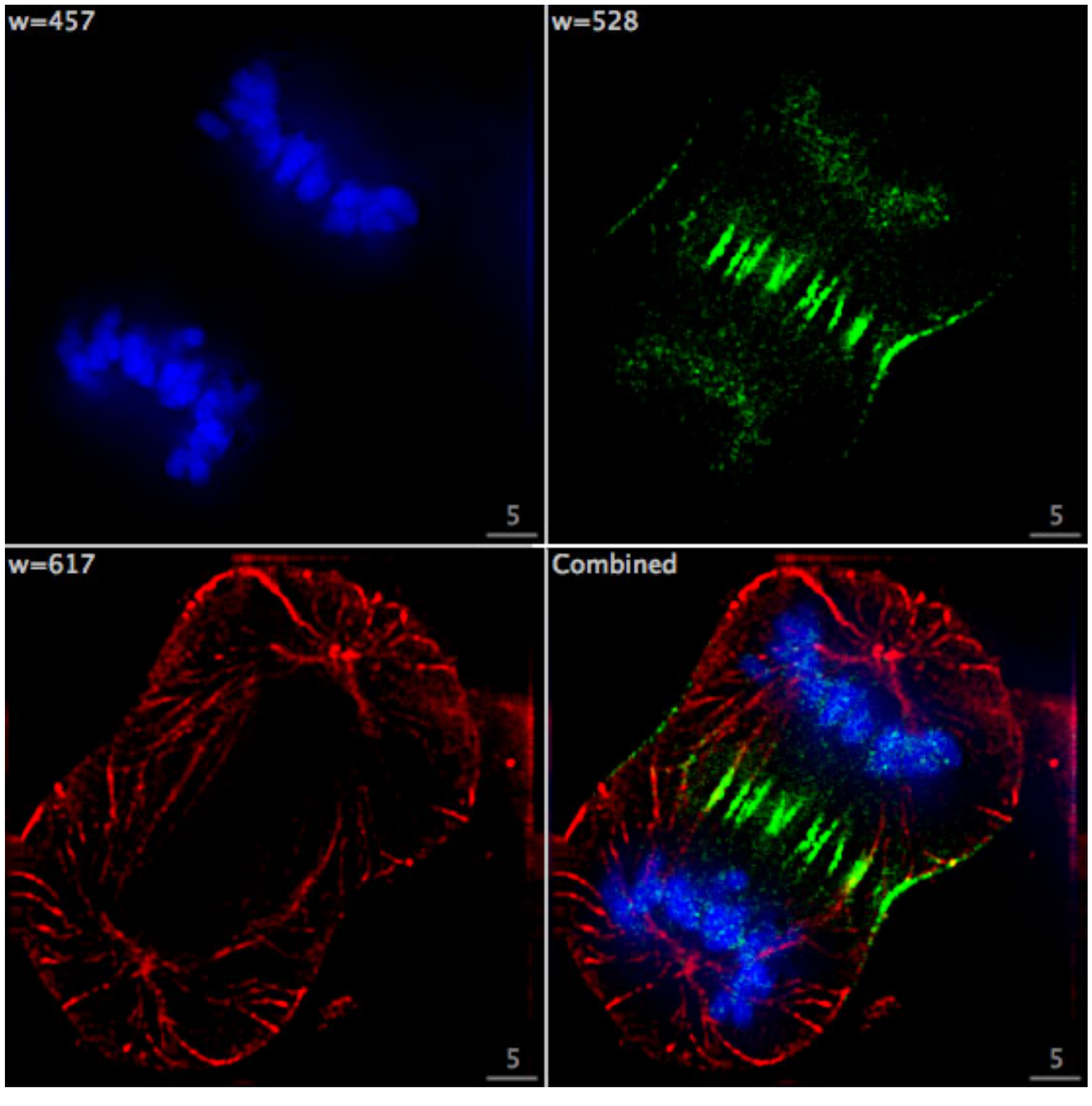
- Combination of two ideas in microscopy technology
- High resolution images
- Life cells
- 2D or 3D through scanning schemes
- Multi-channel through use of range of fluorescent proteins



https://www.biocompare.com/25608-Microscopes-and-Cell-Imaging-Systems/14617250-ZEISS-LSM-980-Confocal-Laser-Scanning-Microscope/?pda=25608|14617250_0_1|2254289,2254327|1|&dfp=true

https://en.wikipedia.org/wiki/Green_fluorescent_protein#/media/File:Fluorescence_from_Fluorescent_Proteins.jpg

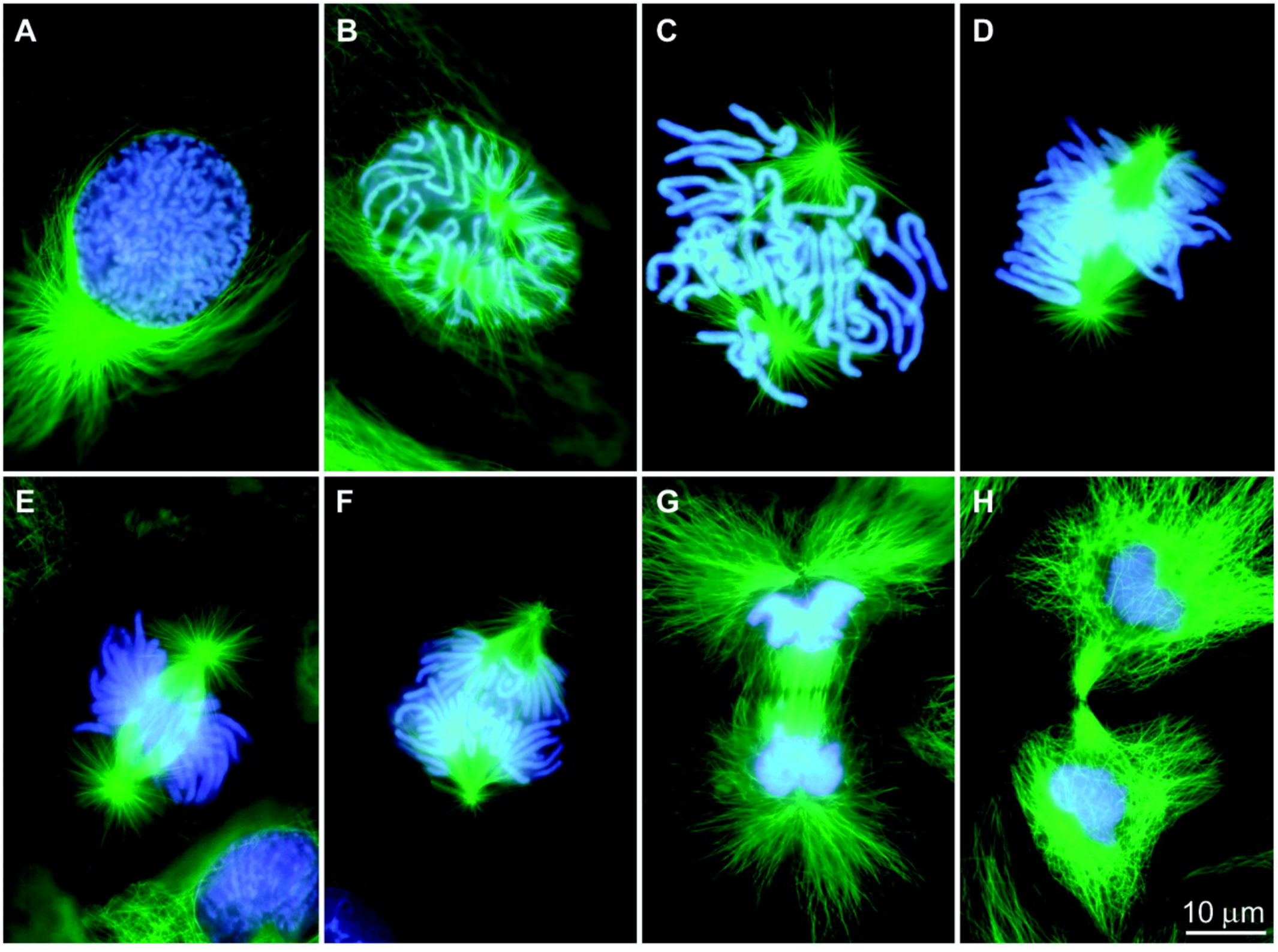
Example:
3 components in dividing human cancer cells



Scanning scheme for fluorescent imaging:

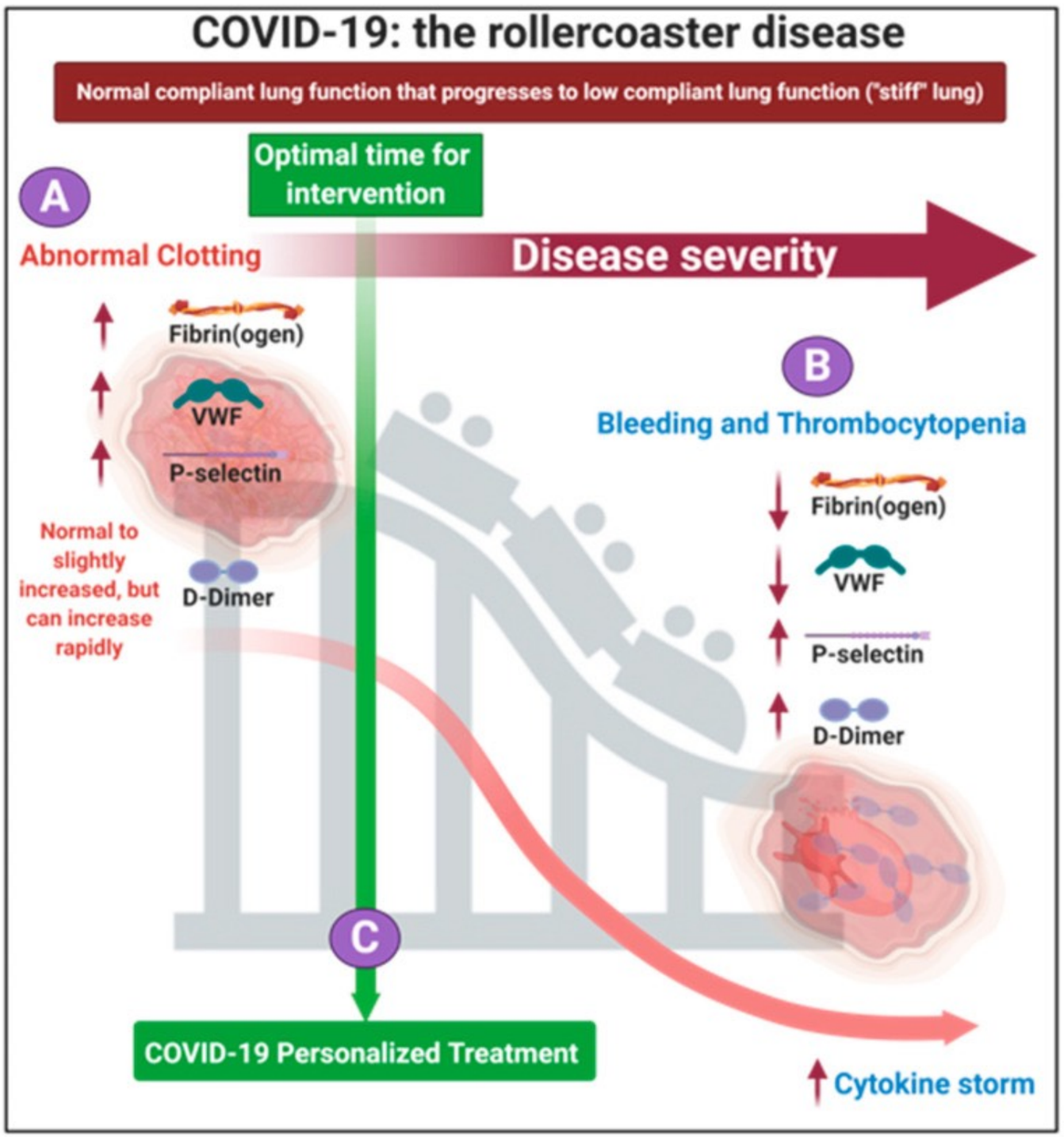
- Blue: Chromosomes (DNA)
- Green: INCENP (protein)
- Red: microtubules
- Fluorophores imaged separately using different excitation and emission filters
- Images captured sequentially
- Overlaid

Microtubules formation during mitosis



https://science.sciencemag.org/content/300/5616/91/tab-figures-data

3 Microscopic imaging of blood samples in (long) Covid

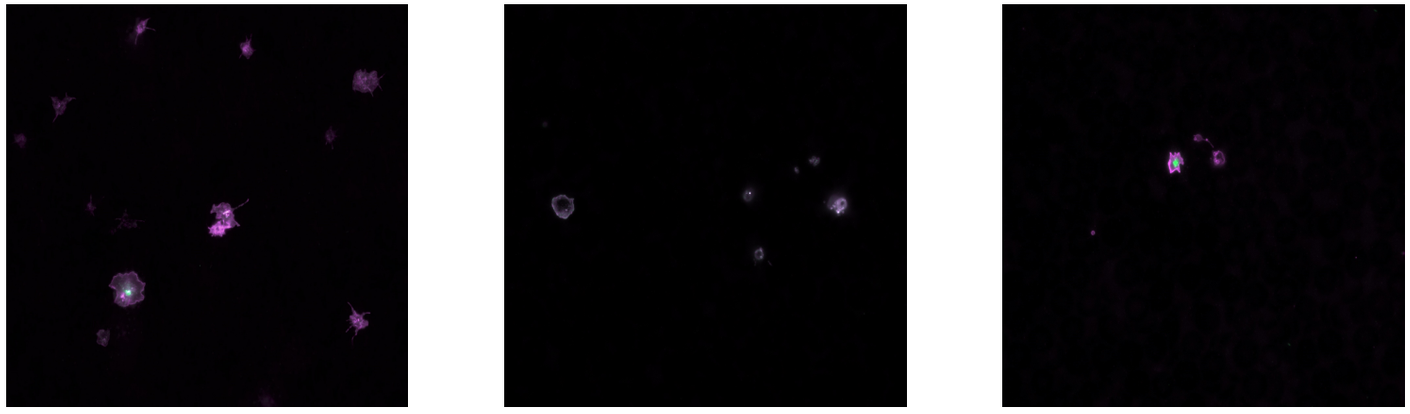


Challenges:

- Diagnostic criteria
- Treatment
- Progression
- Timing of intervention
- Automation of diagnosis

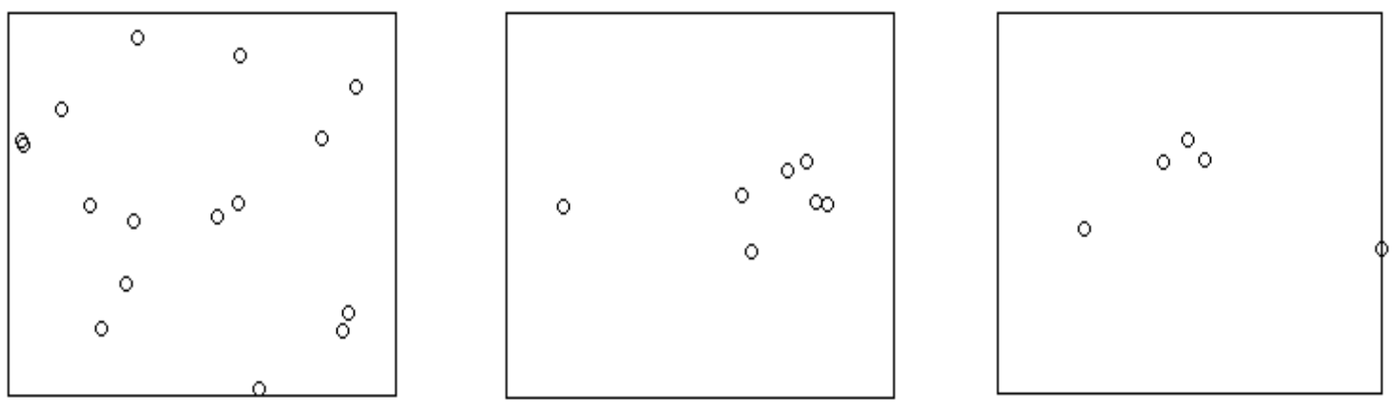
3 Microscopic imaging of blood samples in (long) Covid

Spatial point processe models for platelets in Covid blood samples



(a) (b) (c)

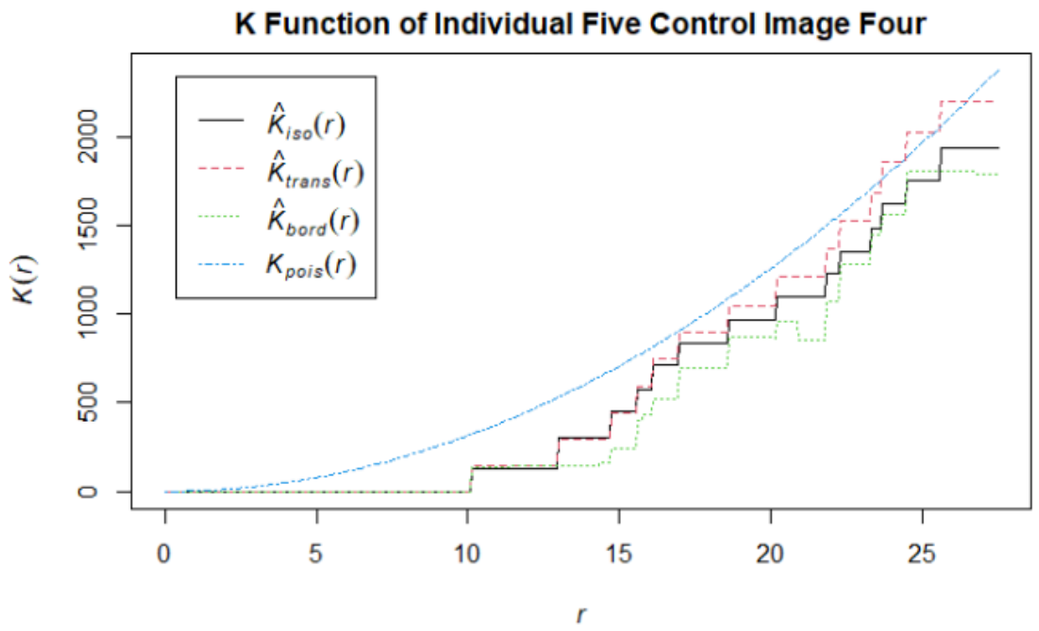
Original microscopy images of (a) Covid-19 case subject 7, image 4, (b) Control subject 2, image 4, and (c) Control subject 2, image 7.



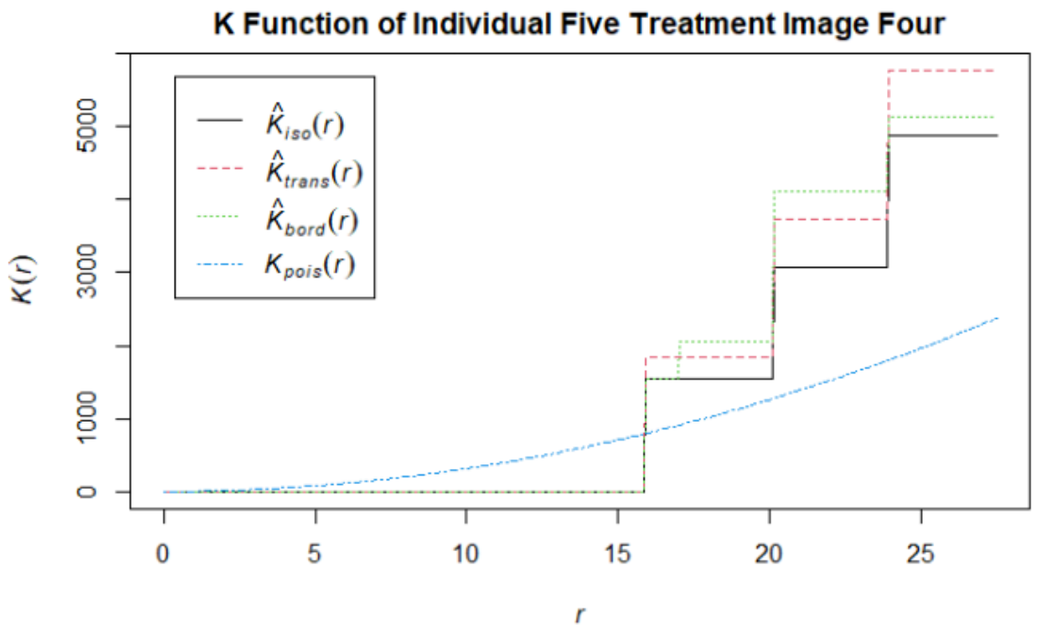
(a) (b) (c)

Spatial point patterns of (a) Covid-19 case subject 7, image 4, (b) Control subject 2, image 4, and (c) Control subject 2, image 7.

3 Microscopic imaging of blood samples in (long) Covid



(a) Control K function



(b) Treatment K function

Controls showing patterns corresponding to being completely spatially at random

3 Microscopic imaging of blood samples in (long) Covid

Morphological measures

Circularity:

ratio of the surface area of an object to the square of its perimeter (Cox [1927], Leach [2013]). Mathematically,

$$M_{Ff} = \frac{4\pi A}{\mathrm{p}^2}$$

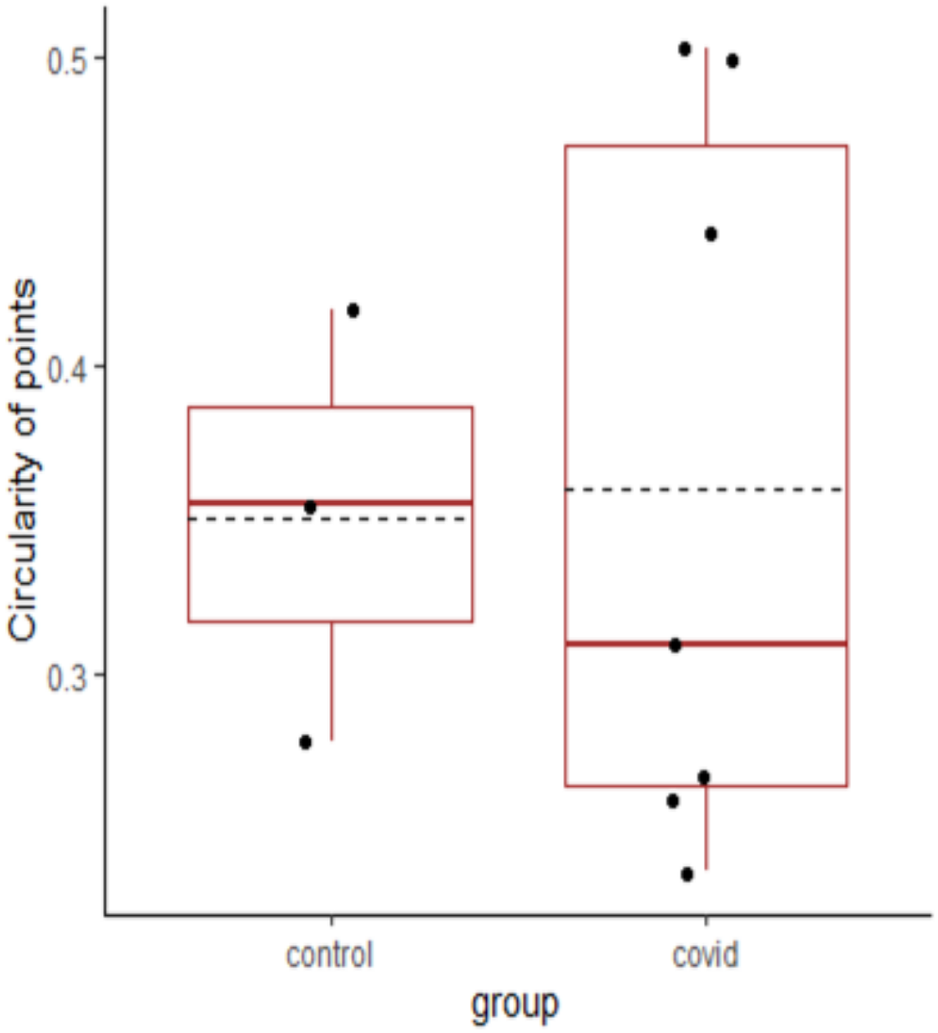
A : surface area, and P : perimeter. Circularity is

Roundness:

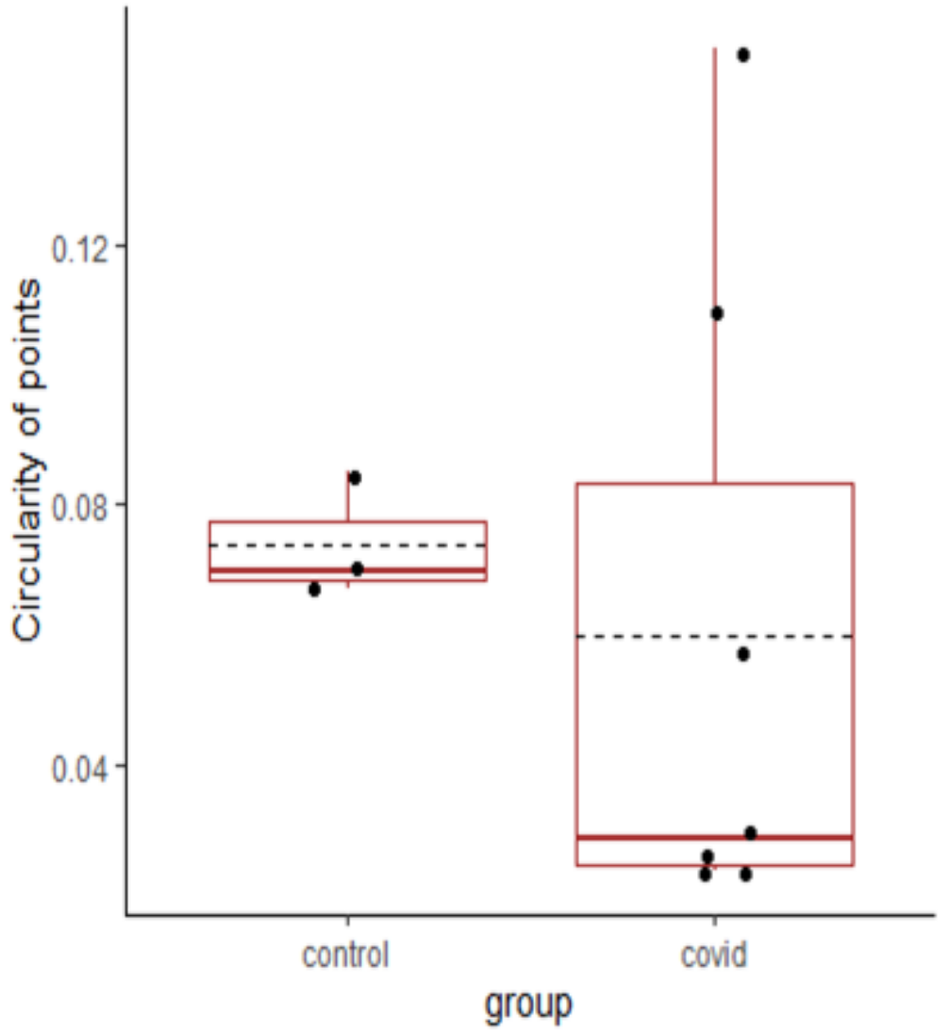
$$M_{RN} = \frac{4A}{\pi D_{\max}^2},$$

$D_{\max}$: maximum diameter, A : surface area and where the diameters are calculated using euclidean distances. Roundness is essentially a ratio

3 Microscopic imaging of blood samples in (long) Covid



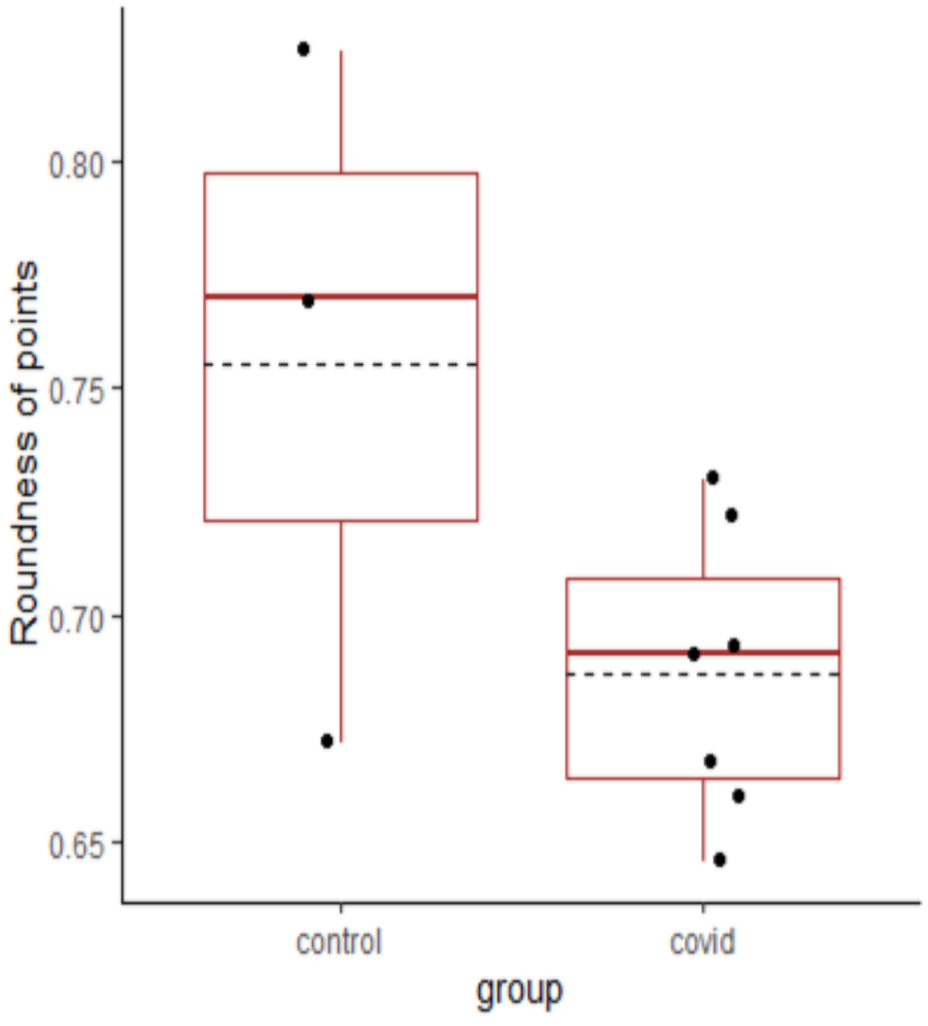
(a)



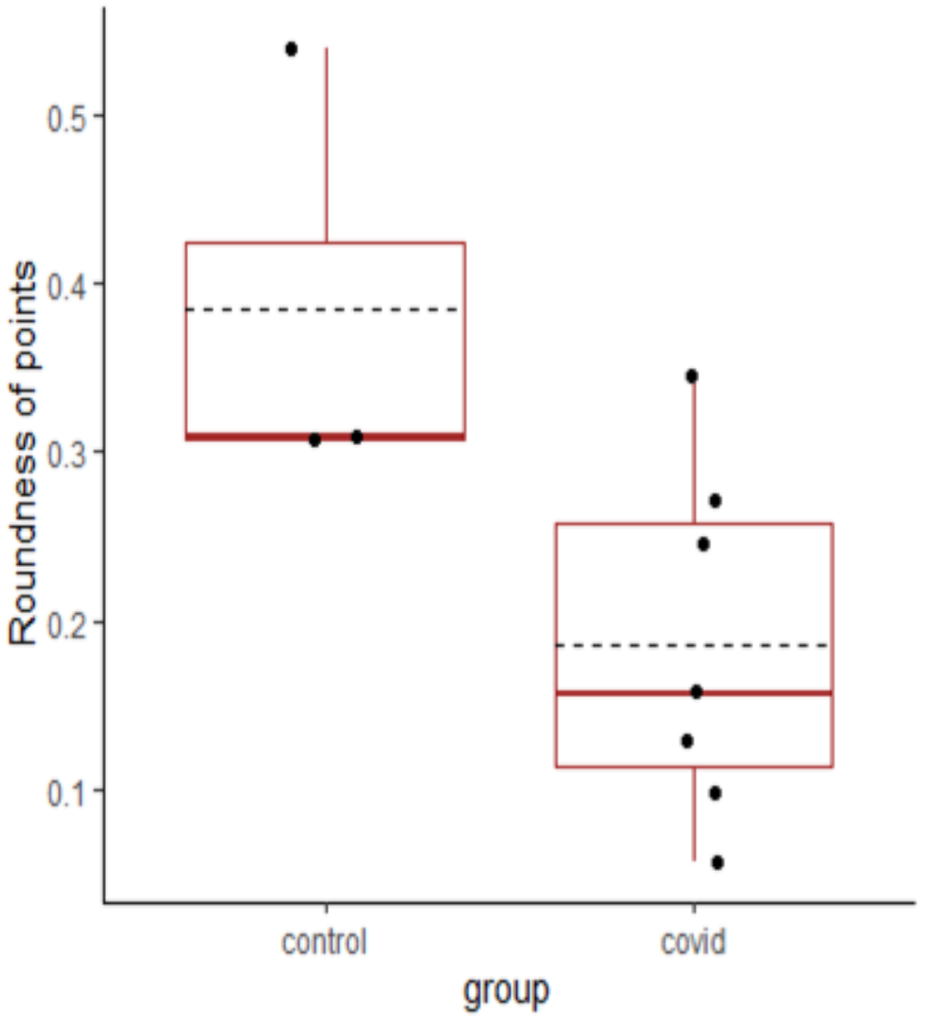
(b)

Boxplots of the circularity index, for observations divided into controls and treatments. Figure (a) uses as scores the average values of circularity indexes, and Figure (b) uses as scores the overall minimum circularity. Medians are indicated by solid horizontal lines and means by dashed horizontal lines.

3 Microscopic imaging of blood samples in (long) Covid



(a)



(b)

Boxplots of the roundness index, for observations divided into controls and treatments. Figure (a) uses as scores the average values of roundness indexes, and Figure (b) uses as scores the overall minimum roundness. Medians are indicated by solid horizontal lines and means by dashed horizontal lines.