June

open-source individual-based epidemics model

Frank Krauss

Institute for Data Science
Institute for Particle Physics Phenomenology
Durham University

DESY Data Science Colloquium, 22.4.2021

Outline

1. introduction: modelling epidemics
2. a first individual-based model
3. introducing JUNE
4. results for the first wave in England
5. a spin-off: Cox's Bazaar Refugee Operation
6. conclusions & outlook

June Almeida

discovered Corona Virus (1964)

introduction:

modelling epidemics

the begin of epidemiology

- start of epidemiology as a science in early 20's

  Kermack-McKendrick models, three papers from 1927-1935: W.O.Kermack & A.G.McKendrick, "*A contribution to the mathematical theory of epidemics*", Proc. R. Soc. Lond. A115 (1927) 700–721 reprinted in:
  Kermack, W. and McKendrick, A. "*Contributions to the mathematical theory of epidemics – I-III*", Bulletin of Mathematical Biology. 53 (1–2) (1991) 33-118

- side-remark: about as old as quantum physics
- mathematical models developed in aftermath of Spanish Flu

50,000,000 dead worldwide (estimated IFR: 10%),
675,000 in US; 228,000 in UK; 426,600 in Germany

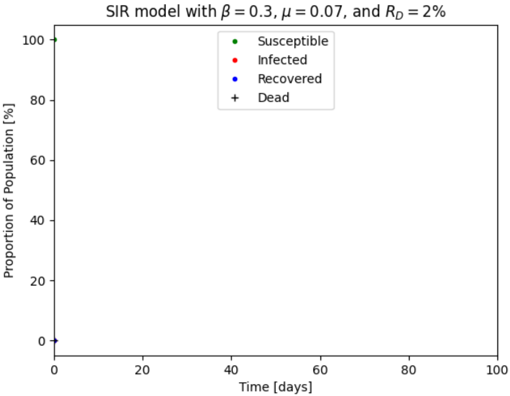
analytical models (SIR and friends)

- simple compartment models:
  distribute population N in boxes according to infection status
- coupled rate equations for moving from one compartment to next
- simplest example: **S**usceptible-**I**nfected- **R**covered)

$$\begin{aligned}
\frac{\partial S(t)}{\partial t} &= -\beta\frac{I(t)}{N}\cdot S(t)\\
\frac{\partial I(t)}{\partial t} &= +\beta\frac{I(t)}{N}\cdot S(t)-\mu I(t)\\
\frac{\partial R(t)}{\partial t} &= +\mu I(t)\\
N &= S(t)+I(t)+R(t)=\mathrm{const.}
\end{aligned}$$

- can introduce more compartments (“exposed”, “dead”, etc.)

example result for SIR model



improving SIR: social mixing

- add detail of population: age & gender
- this amounts to sub-boxes in one compartment, e.g. $S \to S = \sum_i S_i$
- modify interactions between boxes: social mixing matrices $\chi_{ij}^{(E)}$

  count number of daily contacts of people i with people j in social environment E (home, work, school, ...)

  - J.Mossong et al. (POLYMOD), PLoS Med 5 (2008), 3:e74
  - P.Klepac et al. (BBC pandemics project), https://www.medrxiv.org/content/10.1101/2020.02.16.20023754v2

- rate equations become, for example

$$\frac{\partial S_i(t)}{\partial t} = -\sum_E \beta^{(E)} \sum_j \chi_{ij}^{(E)} \frac{I_j(t)}{N_E} \cdot S_i(t)\,,$$

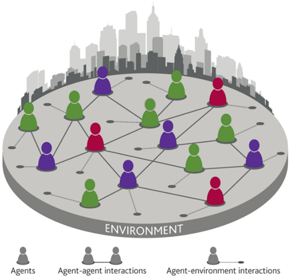
where $\beta^{(E)}$ parametrizes the infection probability in setting E

analytical models to simulation

- advantages of analytical models:
  - "just" differential equations, transparent dynamics
  - fast solutions/fits to existing developments
  - satisfactory capture of global trends
- obvious drawback of analytical models
  - population in "boxes" – insensitive to geography
  - deterministic models to describe statistical process
  - population densities only ("mean-field"): outliers/super-spreader events/variations not captured
- ideal for short-term forecasting (often driven by machine learning), not so good for understanding of inner dynamics
- solution: agent-based models

agent-based models: paradigm

- method for simulating complex systems
- individual agents represent actors:
  - individuals, organisations, etc.
  - interact with each other and environment according to fixed rules
- study impact of individual agent behaviour patterns on whole system



- core method of computational social sciences
  frequent preconception: useful for qualitative analyses only, impossible for quantitative real-world applications (too complex)

  - "mother of all ABM's":
    Schelling, T.C., "*Dynamic Models of Segregation*", Journal of Mathematical Sociology 1 (1971) 143–186
  - some reviews:
    Macy, M.W. and Willer, R., Annual Review of Sociology 28 (2002) 143-166;
    MacNamee B. and Cunningham P., International Journal of Intelligent Games and Simulation 2(2003) 186–221;
    Bruch E. and Atwell J. Social Methods Research 44 (2015) 186–221.

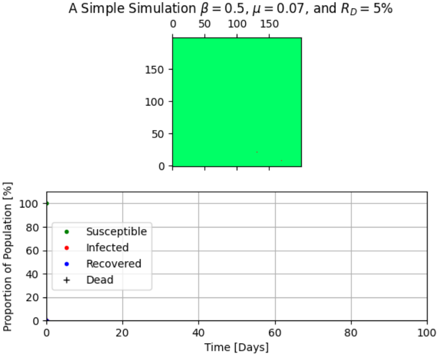
a simple example

(starting point of JUNE, I built it for my kids in March to explain how epidemics spread)

- agents fixed on grid (200×200)
- i infects susceptible s with

$$\mathcal{P}_{si} = \beta \exp\left[-(r_{si}^2 - r_0^2)/r_0^2\right]$$

  with relative spatial distance r_{si}
- time to recovery or death distributed according to Gaussian
- observation(s):
  - similar to SIR solution
  - no herd immunity below 100%



a first individual-based model

agent-based models for epidemics: construction principle

- track disease progression through each individual

  (thus populations become highly heterogeneous by health status during simulations),
- track contacts of individuals in social environments and geographies
- add explicit rules for disease transmission and impact
- agents are individuals: $\longrightarrow$ **individual-based model**
- need to construct:

population & social environment $+$ disease transmission & health impact

most famous example: Imperial College model

N.Ferguson et al., "*Strategies for containing an emerging influenza pandemic in Southeast Asia*", Nature 437 (2005) 209-214
N.Ferguson et al., "*Strategies for mitigating an influenza pandemic*", Nature 442 (2006) 448–452

- population model:
  - distribute population in age & sex according to national distributions
  - distribute population according to national population density
- social environments:
  - four relevant environments: household, company, school, other
  - household, company, and school sizes and composition distributed according to nation-wide distribution
  - "other" captures all contacts in the closer or wider community:
    - $\rightarrow$ venues such as: shops, gyms, pubs, ...
    - $\rightarrow$ activities such as visiting friends & relatives, travel, ...
    - $\rightarrow$ distribute them according to a $1/(r + r_0)$-law w.r.t. place of residence

- one underlying assumption:
  transmission in equal parts in household, work/school, and other

(I am not sure in how far they changed this)

- disease transmission probability in interval Δt:

$$\mathcal{P}_{si}(t,\ t+\Delta t) = 1 - \exp\left[-\psi_s\,\beta^{(E)}\,\frac{\sum\limits_i I_i(t)}{N_E^{\eta_E}}\,\Delta t\right]$$

$\to$ time-dependent infectiousness $I_j(t)$, and
$\to$ susceptibility ψ_s, and
$\to$ normalisation exponent η_E to capture different contact frequencies in environments E

- outcome of infection from continuously updated public health data

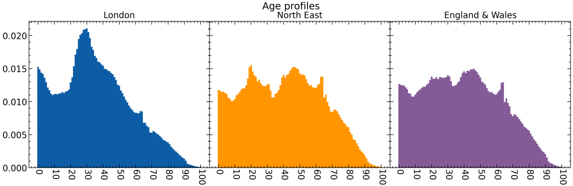
(they started as soon as the first data came out of Wuhan)

JUNE

motivation: why granularity matters

- impact of COVID=19 highly age-dependent
  $\longrightarrow$ **need geographical granularity for regional planning**

(coincidence: Durham hosts & maintains England & Wales census data of past decades)

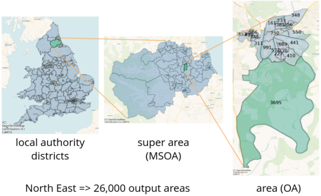


inputs: demographics

- last census (2011)

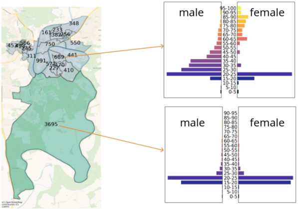
  (data freely available from Office for National Statistics)

- hierarchical data structure



- OA's with $\sim$ 250 residents, with similar characteristics

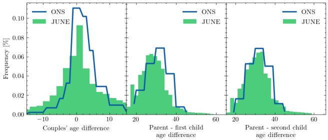
- build virtual population in OA: age, gender, ethnicity, deprivation index
- example: Durham



virtual households

- correct compositions important: primary place for infections
- household composition in 20 categories at OA level

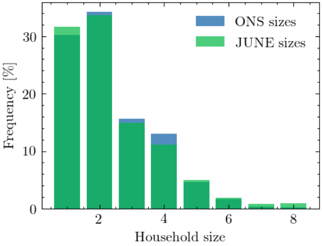
(children, young adults, adults, old adults)



- also: communal facilities

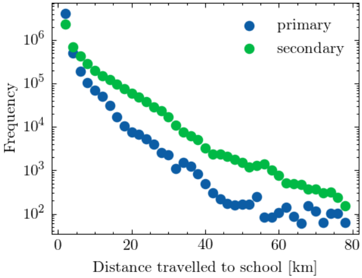
(carehomes)

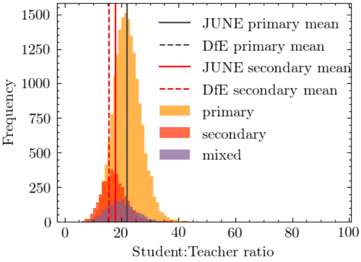
- further test: interplay with social mixing
- example: North-East England



virtual schools

- information about schools: age range of children and locations
- send kids to nearest age-appropriate structure
- could modulate this with school sizes, if necessary

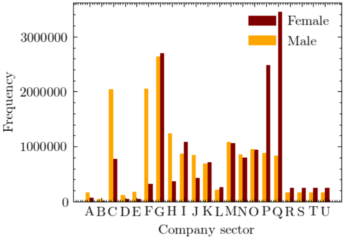


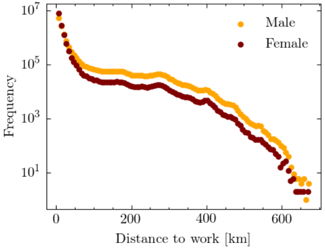


work & virtual companies

- workers and companies in ~ 20 macro-sectors at MSOA level
- know age/sex distribution in sectors nation wide
- distribute workers over companies (we know their sizes in bins)

  (construct a big origin-destination matrix & optimise)
- information about commute mode: public vs. private

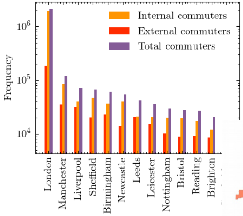


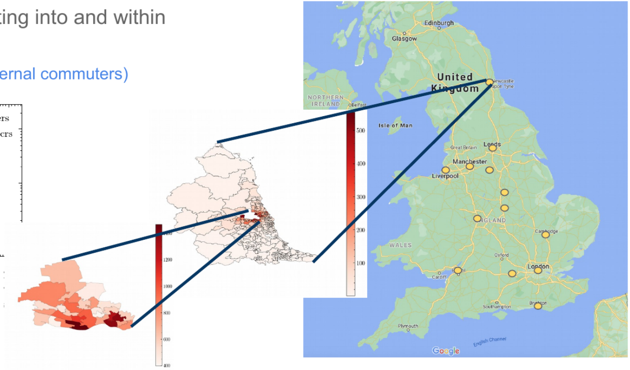


commuting patterns

- distinguish commuting into and within metropolitan areas

  (external vs. internal commuters)



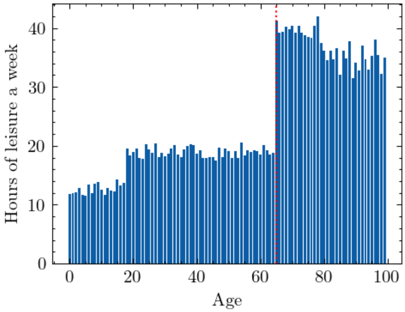


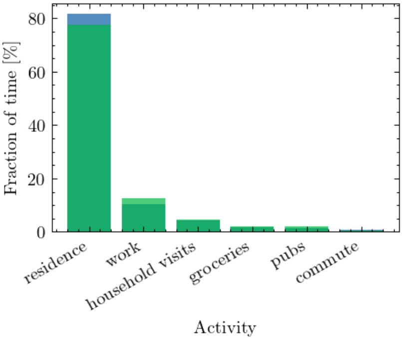
simulating daily structure

- JUNE allows flexible daily routines, separation of weekday/weekend
- time spent on activities known from ONS surveys

(this changes under lock-down)

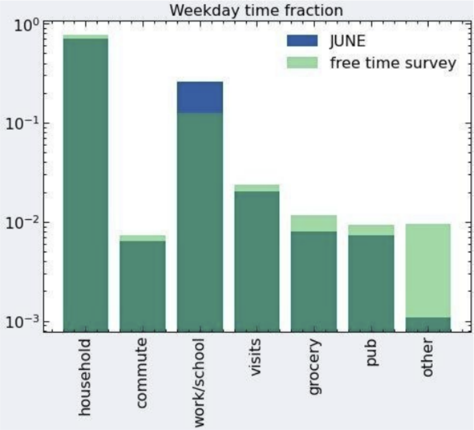
- can translate into age-dependent probabilities for activities





validation: average time budget

- compare our simulation with data from ONS
- average time spent per day in different activities



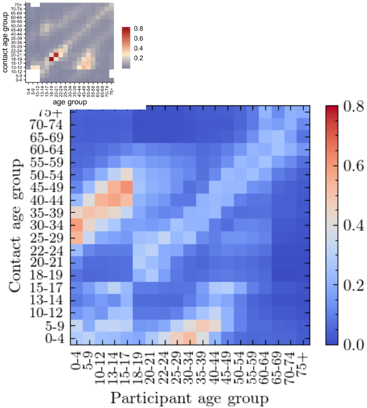
social mixing matrices

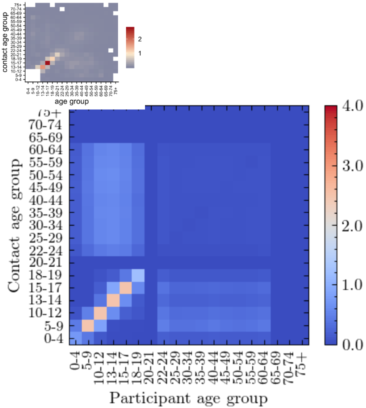
- move on to the environment for the interactions
- use social mixing matrices from POLYMOD and BBC Pandemics project
  J.Mossong et al., PLoS Med 5(3) e74, https://doi.org/10.1371/journal.pmed.0050074;
  P.Klepac et al., https://www.medrxiv.org/content/10.1101/2020.02.16.20023754v2
- denote number of contacts of person with age i with person of age j
- somewhat tricky format: averages over full population
  $\longrightarrow$ have to normalise it to our fractured social environment
  $\longrightarrow$ interplay with household sizes etc. (may have to be fitted?)

JUNE output

- example: household (left) & schools (right) (BBC results as inlays)
- broad agreement with input from surveys: interesting closure test

(in JUNE contacts also depend on composition of environment)





disease transmission

- probabilistic process:

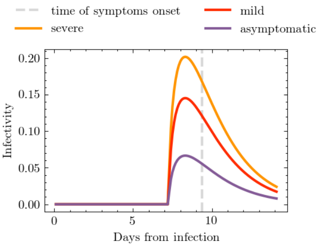
$$\mathcal{P}_{\int\rangle}(t,\, t+\Delta t) \,=\, 1 - \exp\left[-\psi_s\beta_{si}^{(E,g)} \int\limits_t^{t+\Delta t} \mathrm{d}t' \sum_{i\in g} I_i(t')\right]\,,$$

- ψ_s is susceptibility (reduced for children)
- $\beta_{si}^{(E,g)}$ is constant "closeness" in environment E, modulated by contact matrix:

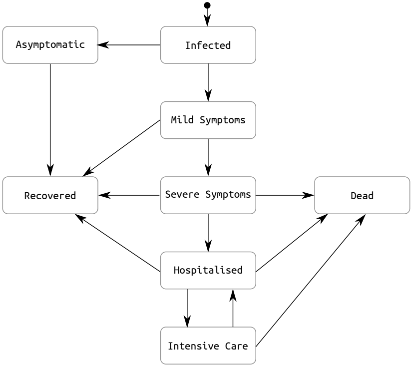
$$\beta_{si}^{(E,g)} = \beta^E \cdot \frac{\chi_{si}^{(g)}}{N_g}$$

with size of group N_g

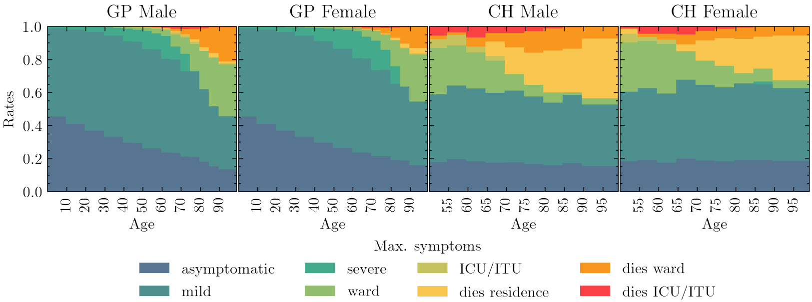
- $I_i(t')$ is infectiousness profile



outcomes of infection

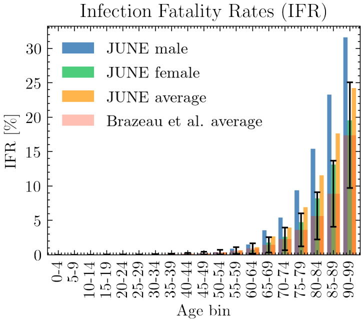


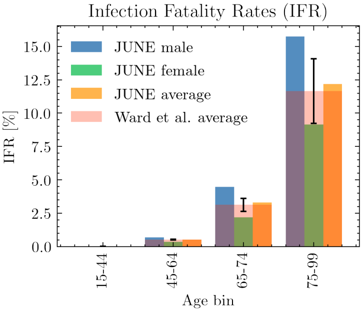
- this was a tiring data-mining exercise with inconsistent and often contradictory data
- extra difficulty: include care homes (CH) vs. general population (GP)



validation

- compare infection fatality rate with other codes



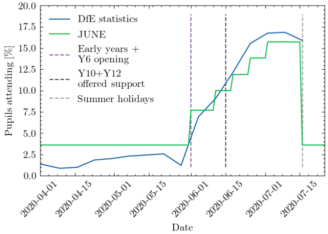


policies

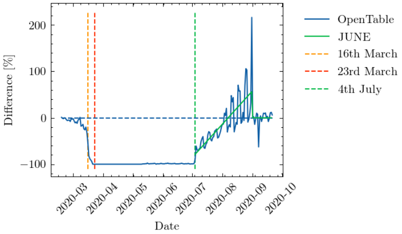
- below a list of relevant policies for results
- realised by modifications of β (e.g. social distancing) or by making activities unavailable

Date (dd/mm/yy)	Policy
12/03/20	case isolation at home
16/03/20	voluntary household quarantine, work from home, avoidance of leisure activities, encourage social distancing, stop all non-essential travel & contact, shielding of over 70's
20/03/20	closure of schools and universities
21/03/20	closure of leisure venues
11/05/20	multiple trips outside are allowed in England only
13/05/20	encourage to go back to work while distancing
01/06/20	meeting in groups of up to 6 outside allowed, shielding of over 70s relaxed, school reopening for Early Year and Year 6 students
13/06/20	'support bubbles' allowed
15/06/20	school reopening for Year 10 and 12 students for face-to-face support
04/07/20	leisure venues allowed to reopen, household-to-household visits permitted along with overnight stays
24/07/20	Mask wearing compulsory in grocery stores
01/08/20	shielding is paused, 'Eat Out to Help Out' scheme introduced
31/08/20	'Eat Out to Help Out' scheme ends
01/09/20	schools and Universities allowed to reopen, 'Rule of 6'
14/10/20	Tiered local lock-down system introduced

validation

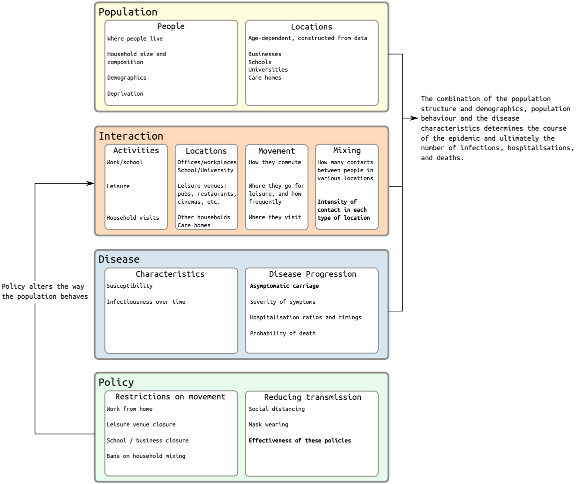


DfE statistics
JUNE
Early years + Y6 opening
Y10+Y12 offered support
Summer holidays
Pupils attending [%]
Date
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
2020-04-01
2020-04-15
2020-05-01
2020-05-15
2020-06-01
2020-06-15
2020-07-01
2020-07-15



OpenTable
JUNE
16th March
23rd March
4th July
Difference [%]
Date
200
100
0
−100
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09
2020-10

JUNE simulation content - summary

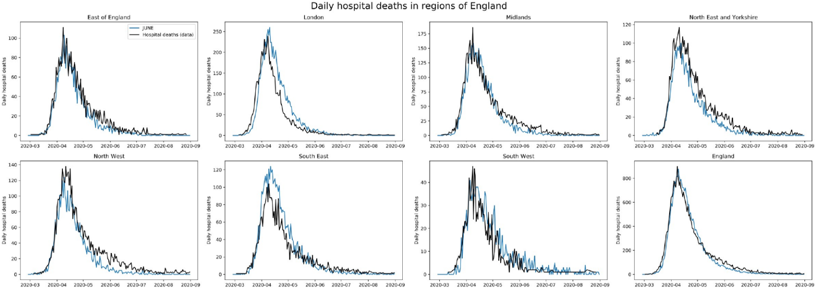


JUNE results for the first wave

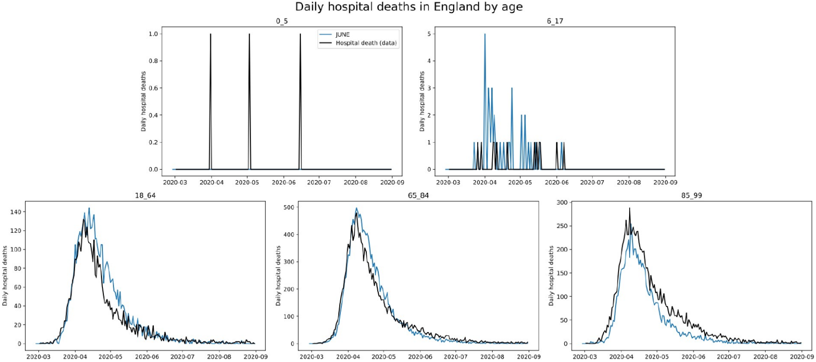
June results

- below a selection of June results for the 1$^{\mathrm{st}}$ wave in England
- some aspects of the code changed in the meantime:
  - seeding of infections - must reorganise due to multiple CoVid streams
  - added gyms as leisure venues
  - take into account ethnicity in household compositions
  - added vaccination protocols
- refitting underway

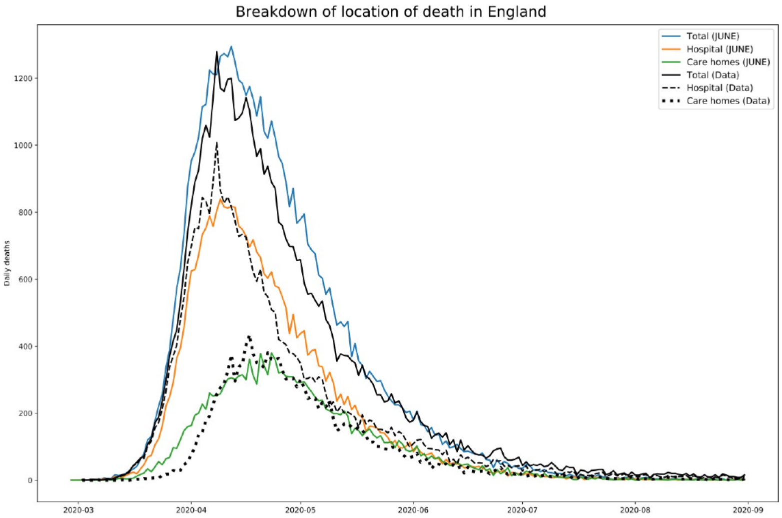
- 1^{st} wave: deaths in hospitals - regional distribution



- 1^{st} wave: deaths in hospitals - age distribution

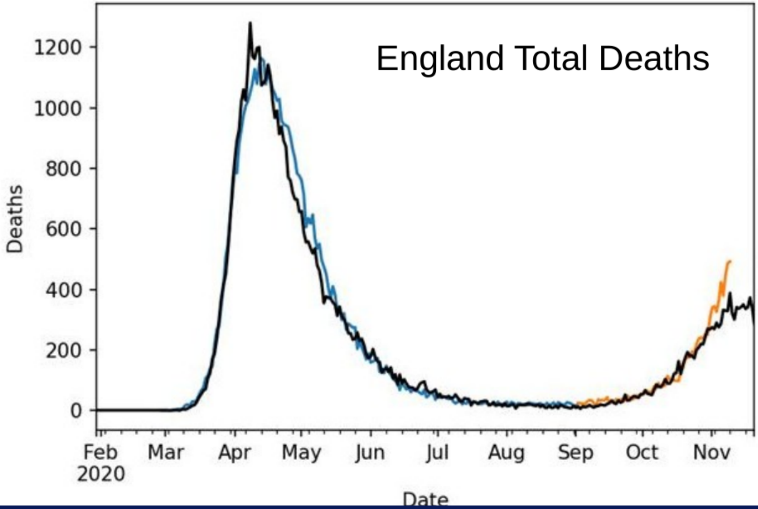


- 1^{st} wave: all deaths - distribution of location



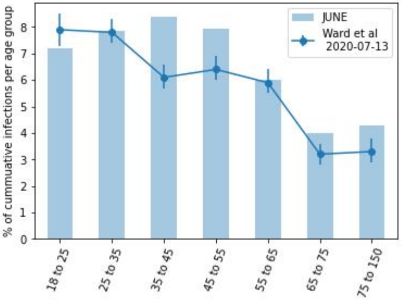
- extrapolation to second wave from September 1$^{\mathrm{st}}$

fitted parameters from first wave only!

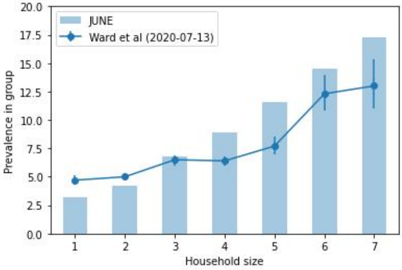


social imbalances

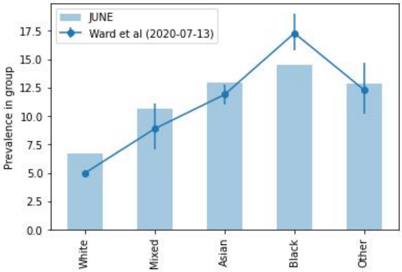
- look at cumulative infection rates until July 2020 in dependence on
  - age band
  - household size
  - ethnicity
- keep in mind: all imbalances only due to regional and sociological differences encoded in census data



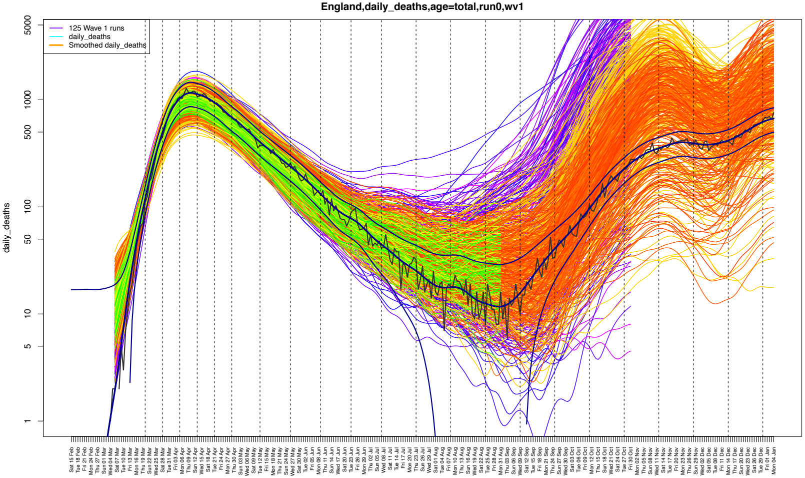
impact of household size



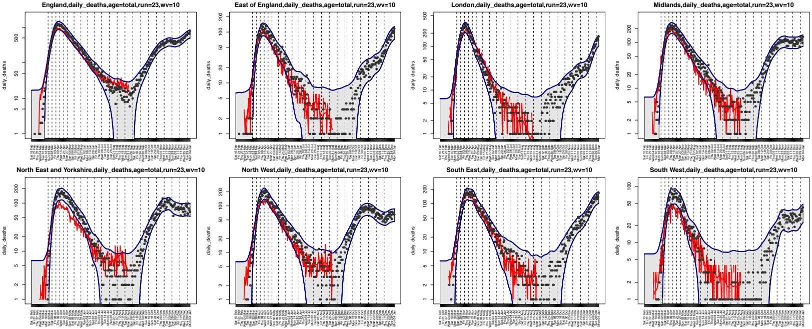
impact of ethnicity



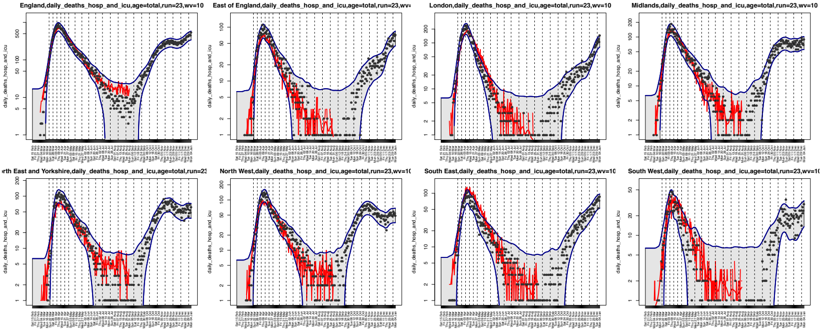
refitting snapshot



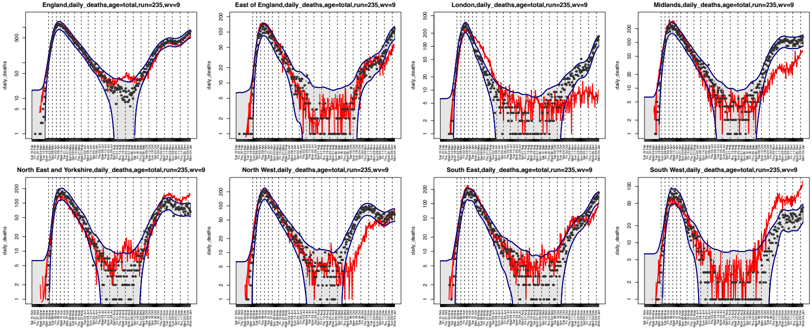
improved 1$^{\rm st}$ wave: total deaths by region (example run)



improved 1^{st} wave: deaths in hospital by region (example run)



first stab at 2$^{\text{nd}}$ wave (example run)



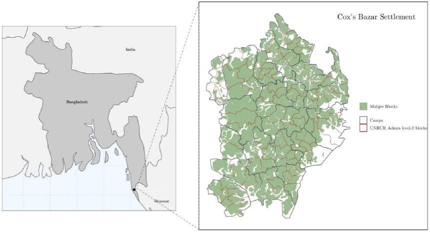
a spin-off:

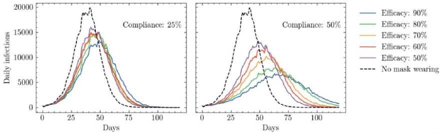
JUNE for Cox's Bazaar Refugee Operation

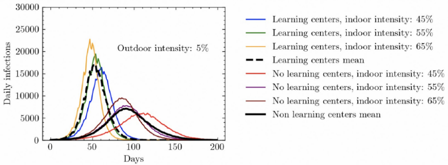
- one of our students (J.Bullock) has spent 6 months of placement with UN Global Pulse in New York
- ongoing contact, and he talked with them about our code
- as a result, three students collaborate with WHO, UN-Global Pulse, and IBM-MIT Watson AI lab to provide tools for scenario planning for Cox's Bazaar in Bangladesh

(huge refugee camp, Rohingya crisis in Myanmar)

- largest settlement in the world
- in some areas, the settlement is denser than New York City
- high risk of COVID transmission




Daily infections
20000
15000
10000
5000
0
0 25 50 75 100
Days
Compliance: 25%
Compliance: 50%
Efficacy: 90%
Efficacy: 80%
Efficacy: 70%
Efficacy: 60%
Efficacy: 50%
No mask wearing


Daily infections
30000
25000
20000
15000
10000
5000
0
0 50 100 150 200
Days
Outdoor intensity: 5%
Learning centers, indoor intensity: 45%
Learning centers, indoor intensity: 55%
Learning centers, indoor intensity: 65%
Learning centers mean
No learning centers, indoor intensity: 45%
No learning centers, indoor intensity: 55%
No learning centers, indoor intensity: 65%
Non learning centers mean

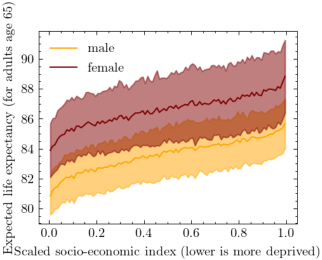
conclusions & outlook

summary

- epidemiology is an interesting field of science
- constructed an individual-based model with supreme granularity: demography, geography, sociology
- model informed operational planning of NHS:
  - $\rightarrow$ early warning of second wave
  - $\rightarrow$ understanding of transmission sociology
- code is highly flexible:
  - $\rightarrow$ addition of new effects & policies relatively painless
  - $\rightarrow$ adaptation to new environments: Cox's Bazaar
  - $\rightarrow$ adaptation to Germany underway (M.Schott, U Mainz)
- challenge to widespread perception in computational sociology: more and better detail often helps

outlook

- continue decision support for NHS:
  - impact of mutations:
    added Kent variant B1.1.7
  - impact of vaccination protocols:
    take-up, efficacy
  - new fits underway
    (difficulty: Xmas mixing)
- follow-up studies on social imbalances:
  ethnicity, IFR in dependence on
  socio-economic deprivation
- use for historical studies:
  Spanish Flu, Black Death



meet the JUNE team

- work started in March
- publication on medRxiv (https://www.medrxiv.org/content/10.1101/2020.12.15.20248246v2)
- very interdisciplinary research, programming and data crunching mainly driven by **PhD students** from Durham CDT

Joseph Bullock[1,3], **Carolina Cuesta-Lazaro**[1,2], **Arnau Quera-Bofarull**[1,2],
Miguel de-Icaza-Lizaola[1,2], **Aidan Sedgewick**[1,2], **Henry Truong**[1,3],
Aoife Curran[1,2], **Edward Elliott**[1,2],
Richard Bower[1,2], Tristan Caulfield[4], Kevin Fong[4],
Ian Vernon[5], Julian Willams[6] & FK[1,3]

1 Institute for Data Science Durham
2 Institute for Computational Cosmology Durham
3 Institute for Particle Physics Phenomenology Durham
4 University College London & NHS
5 Mathematical Sciences, Durham
6 Institute for Hazard, Risk & Resilience Durham