Modelling the Spread of Disease Considering Clustering in Vaccination Status Within Social Networks

Lauren Hughes

Supervised by: Gavin Woolman

July – August 2026

Institute for Particle Physics Phenomenology, Durham University

Funded by Laidlaw Scholars Leadership & Research Programme

Contents

Abstract

In this report the effects of the clustering of individuals with the same vaccination status, vaccinated or unvaccinated, within their respective social networks, people they interact with, on the severity of an outbreak of measles is explored. MAY and JUNE software that simulate the world and a disease spreading within it are utilised and developed upon to examine this via a simulation. This report is a pilot, the simulation is run on a smaller area, the county of Wiltshire, to test the software developed and gather preliminary conclusions on the hypothesis. It shows so far that greater clustering can cause there to be more cases of infection, i.e. a worse outbreak. The unique insight this process provides is that the software encodes social networks via a literal list of associates allowing true clustering as opposed to similar works that use proxies for social networks i.e. locational distance.

1. Introduction

Motivations

Public health in the UK, America and other western countries is increasingly threatened by many vaccine curable diseases [1]. For example, the WHO stated the number of measles cases in the European region had doubled to 127,350 in 2024 [2]: the highest level seen in 25 years. This has deepfelt impacts; 5 people died from measles in the UK in 2019 and 2024 the highest number of deaths since 1988. [3]. But this issue is tractable, the UK has previously had measles elimination status and lost it.

This worrying trend is largely due to a stalling and decline in vaccine uptake in some countries [1] often in the west. For example, there were declines in 21 of 36 high-income countries for at least one type of vaccine doses [4]. In the UK in 2025 the uptake of the MMR vaccine (Measles, Mumps and Rubella) decreased to 91.8% and 83.7% for the first and second dose respectively from highs of 95% in 2017 for the first dose and 88.6% in 2015 for the second [5].

Looking to the future this means National Immunisation Strategy must be as effective as ever and understand these new dynamics in order to create lifesaving policies. An underexamined precedent not accounted for in many herd immunity models is the clustering of unvaccinated individuals. This is proven to be dangerous reality,[6] occurring often due to social selection where non-vaccinted individuals share socio-demographic traits which are themselves clustered and social influence in which vaccine hesistancy is an ideology spread social networks. [7] This precedent is expected to change outbreak risks for the worse. [8] [9] [10] [11] For example the WHO sets a target for countries of 95% vaccination coverage; however, some studies show this level of overall population coverage isn't enough to prevent serious outbreaks when accounting for low uptake clusters [12] [10]. Making this an important focus for countries achieving higher rates of coverage [10] where clustering effects play a larger role but may be overlooked.

Modelling clustering requires an understanding of the social networks within society. Previous research done on the implications of clustered un-vaccinated individuals on the spread of disease use proxies for social networks such as, physical proximity [11], same schools, or households' [8]. The software called MAY, used in this report, allows us to explicitly extract individual's social networks i.e. which other people in a population they have contact with, giving a unique insight into the effects of literal social clusters.

Introduction to the Software

MAY builds a 'digital model' of the world within which the spread of a disease can be simulated. It creates a record of every person (agent) in the population and their attributes like age and sex, including features that would affect an individual's response to disease i.e. comorbidities. Agents are allocated to a location in England and a household with a family; venues they attend i.e. for work and leisure are also assigned. Social connections are created from certain pools of relevant people. These factors are required to determine who interacts and where to track the transmission of disease. Real world data on all these factors is used to construct an accurate picture of the England. A complete explanation can be found here https://idas-durham.github.io/MAY/ .
JUNE simulates the spread of a specific disease within the world handed to it built by MAY. An initial seed of disease is given to some agents in the population. Then who it spreads to over time is determined considering the encounters occurring and the transmission likelihood of these encounters based on the venues and the individual's status, i.e. vaccinated or immune, of the encounter. JUNE keeps a record of the status, called a symptom tag, of all agents each day, the possibilities are healthy, exposed, asymptomatic, mild, sever, hospitalised, intensive care, death at home, death at hospital and recovered. A complete explanation can be found here https://github.com/IDAS-Durham/JUNE2/blob/main/USER_GUIDE.md .

Vaccines and Measles

Vaccines have been in use for over 200 years and are considered 'after clean water … the most successful public health measure in history'. [4] In the UK a measles vaccine was first introduced 1968 and then the combined MMR was introduced in 1988.
Measles is a highly contagious infection with an R0, basic reproduction number, of 12-18, meaning the average number of secondary cases caught from one infected individual in an entirely susceptible population, compared to covid which is 2.6. This was motivation to use measles in this report as they will exacerbate any additional spread because of clustering. Measles symptoms include a fever, cough and the infamous rash. Severe complications include pneumonia, blindness and seizures.

In this report we will explore how social clustering of un-vaccinated individuals affects the spread of measles in England and how we can model this clustering and simulate the spread of disease using MAY and JUNE software.

2. Methodology

Firstly, MAY was used to build a digital model of England, in which the vaccination status of each individual was assigned probabilistically using real UK COVER (Cover of Vaccination Evaluated Rapidly) data. This formed the non-clustered world. Two more worlds were created with varying degrees of clustering. The previous world was a basis and individuals with the same vaccination status were made to be more likely to be 'friends' i.e. in each other's social network. Secondly, JUNE was adapted to model measles. Finally, simulations were run using both the non-clustered and clustered worlds, and the resulting outcomes were compared.

This methodology provides an abstraction of the process; the implementation primarily involved creating and modifying YAML configuration files in MAY and JUNE. YAML is a human-readable data-serialisation format used to store and exchange structured data. These files will hopefully be published.

Assigning data accurate vaccination status to agents in England

The aim was to represent the English population as accurately as possible so that the resulting model would provide a realistic, applicable basis for investigating measles transmission. The adapted simulation was based on the data and YAML configurations already collated for 2021, which can be found in the github JUNE files references in the introduction. A new vaccination-status attribute was probabilistically assigned to individuals using MAY's attribute-assignment function.
MAY allows the probability of an agent being assigned a particular attribute to be determined from a CSV file based on demographic characteristics. In this case, the aim was to determine the probability that an individual had been vaccinated based on their age and local authority. UK COVER data [13] records the percentage of relevant birth cohorts vaccinated in each local authority at a specified age. The coverage data were therefore converted to represent the estimated vaccination status of individuals in 2021. For each year of available data, the corresponding cohort's age in 2021 was calculated using:
Age in 2021 = 2021 − year of data collection + age at which vaccination coverage was measured.
The local-authority labels in the COVER dataset also had to be matched to the corresponding local government unit (LGU) labels used by MAY.
Vaccination probabilities were assigned differently according to age. For individuals under seven years of age, the probability of vaccination was determined using both age and local authority. For individuals over seven, only age was used. This distinction was made to account for population movement, as the data on specific local-authority coverage for individuals older than seven was deemed to be too inaccurate to use.
To simplify the model within the available timescale, the model did not distinguish between individuals who had received only an MMR1 dose, those who had subsequently received MMR2, or those who had received a booster. Instead, vaccination status was represented using the proportion of individuals who had received the first dose MMR1.
Individuals who were already older than one year when measles vaccination was introduced in 1968 were not assigned a vaccination status. This was intended to distinguish individuals who had been born before the introduction of vaccination from individuals who were subsequently eligible for vaccination but remained unvaccinated which becomes relevant when creating clustering. The possible states were therefore vaccinated, unvaccinated, and no vaccination status.

Creating the clustered worlds

Individuals in MAY have social connections representing relationships with other people with whom they may interact as discussed in the introduction. The aim was to increase the proportion of social connections between individuals who had the same vaccination status. This is referred to here as the edge-match rate, where an edge represents a connection between two individuals.
Initially, an approach was investigated that did not require the entire social network to be rebuilt, with the aim of minimising computational requirements. This was considered important because the eventual aim is to run the model for the whole of England on a supercomputer, where it's important to reduce computational time and costs.
The first approach attempted to increase the edge-match rate by reassigning vaccination statuses after the social network and world had been generated. Python code compared the vaccination status of connected individuals within the same local authority and swapped vaccination statuses when this could produce a better match. This preserved the total number of individuals assigned each vaccination status within a local authority. However, when the overall edge-match rate was calculated, it had decreased rather than increased. This was likely because too few swaps were made and because the algorithm did not consider the edge-match rate across all of an individual's connections when selecting swaps. This approach was therefore abandoned in favour of time,

although a more sophisticated reassignment algorithm could be investigated in future work.
The second approach altered the population from which friendships could be selected during social-network generation. Individuals were filtered according to vaccination status so that connections were more likely to occur between individuals with the same status. According to the social-network configuration, most individuals have an average of approximately six connections: three associated with shared venues, two associated with similar age and geographical location, and one associated with a wider age range and geographical location.
For Cluster 1, only the two connections based on similar age and the same MGU were constrained to be between individuals with the same vaccination status. For Cluster 2, an additional connection was constrained according to vaccination status, resulting in a greater degree of vaccination-status clustering.
The edge match rates for all worlds in seen in Table 1. Because a large proportion of the population was vaccinated, approximately 95% in Wiltshire in 2021, the overall edge-match rate from vaccinated to vaccinated connections is expected to be high even in the absence of deliberate clustering. Therefore, the edge-match rate calculated for only unvaccinated individuals was also considered, as this provides a more informative measure.

World	Edge match rate	Unvaccinated edge match rate average	Unvaccinated edge match rate median
Non-Clustered	0.75	0.29	0.33
Cluster 1	0.87	0.63	0.67
Cluster 2	0.94	0.82	0.83

Table 1: Edge match rate for all worlds

Configuring JUNE for Measles

Configuring JUNE for a specific pathogen means supplying a set of YAML configuration files that together define the disease's natural history, how it spreads through the population, and any interventions, vaccination, and potentially any other policies, acting on it.
The following mechanics, not covered in the introduction, are relevant to understanding the measles configuration described below:
Transmission of a disease can happen through two modes, respiratory contact (shared air/aerosol exposure) and physical contact (direct touch). Each venue i.e. care home, office and classroom, has a contact matrix a structured way to represent frequency of interaction between different population groups. These specify a separate "beta", per-contact transmission probability, for respiratory contact and physical contact. A pathogen's disease.yaml file declares a mode_transmissibility_multiplier for each mode, and the two betas are scaled independently to reflect how a given pathogen actually spreads. This is the mechanism used below to encode measles' overwhelmingly airborne transmission route.
Each infected individual is assigned one clinical trajectory. The disease's natural history is defined as a set of mutually exclusive trajectories (e.g. "exposed → mild → recovered" or "exposed → mild → hospitalised → dead_hospital"), each built from an ordered sequence of symptom stages with a probability distribution governing how long the individual spends in each stage. Which trajectory a given person is assigned is drawn from an age- and sex-stratified probability table (the outcome_rates_csv), so two people of different ages infected on the same day can end up on very different clinical courses. Individual infectiousness over time is separately modelled as a gamma-shaped curve (disease.yaml's transmission block), whose peak timing relative to symptom onset can be shifted — this is how a pathogen's pre-symptomatic infectious window is encoded.
Vaccines act on two distinct probabilities. A dose's infection_efficacy reduces the probability of becoming infected at all (applied to susceptibility), while its symptom_efficacy shifts probability mass away from the more severe trajectories conditional on infection occurring. Doses ramp up over days_to_effective, hold at full effect until days_to_waning, then decay to days_to_finished. Who receives a vaccine is controlled by campaigns, which run for a defined date window inside the simulation and check each person against a set of eligibility criteria (age, sex, or any custom per-person attribute written into the world by MAY, such as a pre-assigned vaccination status).
Behaviour can change in response to symptoms or policy. policies.yaml can define symptom-triggered rules, e.g. "isolate at home once a person's status reaches a given symptom tag", as well as time-bounded, population-wide policies, e.g. school closures over a date range, each with a compliance rate rather than being deterministic.
Outbreaks are seeded, not assumed. infection_seeds.yaml specifies how and when initial cases enter the population.
Every event is logged for later analysis. Infections, symptom-stage transitions, hospital/ICU admissions and discharges, deaths, vaccinations, and social/network events are all timestamped into a simulation_events.h5 output file, which supporting Python tooling (analysis_tools/june_events) can load to analyse.

A full scenario configuration (configs/config_measles/) was built by adapting JUNE's COVID-19 template (configs/config_2021/), which shares the same file structure but requires every disease-specific parameter to be re-derived from measles epidemiology.

Clinical course of disease (disease.yaml)
Measles' symptom_tags and trajectories were built to reflect its well-documented clinical course: an incubation period modelled as a single exposed stage of approximately 11–12 days on average (range 7–21, drawn from a beta distribution) before rash onset, branching into six mutually exclusive outcomes, inapparent/modified infection, uncomplicated rash illness, complications managed at home, hospitalisation, encephalitis/ICU-level care, and death (split between home and hospital, reflecting the two death pathways). The infectiousness curve's peak was shifted to begin ~4 days before rash onset, matching the documented infectious window of 4 days before to 4 days after rash appears.

Age and sex-stratified outcome probabilities (data/infection_outcome_rates_measles.csv)
Because JUNE draws each infected person's trajectory from an age/sex-indexed probability table, this table had to be constructed rather than reused from the COVID-19 template. Probabilities for hospitalisation, ICU-level care, and death were built as age-dependent curves reflecting measles' characteristic U-shaped severity profile, highest risk under age 5 and rising again from around age 50, lowest through school age, with the residual "severe" (complicated, non-hospitalised) category calculated so that all outcome probabilities for a given age/sex bin sum to exactly 1.

Transmission intensity (contact_matrices.yaml)
Rather than altering the underlying contact structure, who meets whom, and how often, at each venue type, which is disease-agnostic and left untouched, the per-mode transmission probabilities were rescaled to reflect measles' epidemiology: the respiratory beta was scaled up ×7 to reflect its much higher basic reproduction number R0 ≈ 12–18, versus ≈ 2.5–3.5 for the COVID-19 configuration this file was adapted from, while the physical contact beta was scaled slightly less ×2, since measles transmission is more so airborne than touch-mediated. Note that an upper threshold approximation of the beta was used to ensure the outbreak was large enough to analyse. These multipliers are starting assumptions rather than a calibrated fit to a specific observed outbreak.

Outbreak seeding (infection_seeds.yaml)
Rather than seeding infection broadly across the population, appropriate for modelling a pandemic's arrival, measles outbreaks in a largely vaccinated population typically begin from a small number of imported or index cases concentrated in one household or community. This was modelled using JUNE's clustered seeding mode, placed 40 initial cases within a single MGU geographic unit.

Vaccination (vaccines.yaml)
An MMR vaccine was defined with dose-based efficacy against infection, informed by published effectiveness estimates, of 96% and symptomatic disease of 100%. The population's vaccinations were meant to have occurred at several occasions prior to the similation, however JUNE didn't have a function for this so a vaccination campaign was run on day 1 to imitate this. Eligibility for this was evaluated against a person's vaccination status attribute. This yaml was also used to encode immunity assigning 85% of individuals older than 53 (when a measles vaccine was first introduced) as immune by giving this age category 85% infection efficacy with 0% symptom efficacy.

Case isolation (policies.yaml)
A symptom-triggered isolation policy was configured to reflect standard public-health guidance for measles: cases are moved to home isolation once they reach the rash-illness stage, with hospitalisation applied for those progressing to severe or ICU-level disease. No population-wide restrictions were modelled, consistent with real-world measles control relying on case isolation and targeted vaccination rather than broad measures such as school closures.

Verification
Once configured, the presence and content of /events/infections, /events/vaccinations, /events/deaths, and related datasets in each run's simulation_events.h5 output were inspected using JUNE's june_events analysis package to confirm that seeding, transmission, and vaccination campaigns were behaving as configured, for example, checking that vaccination events matched the intended eligible sub-population and that recorded case counts and outcome proportions were consistent with the configured age-stratified probabilities.

3. Results

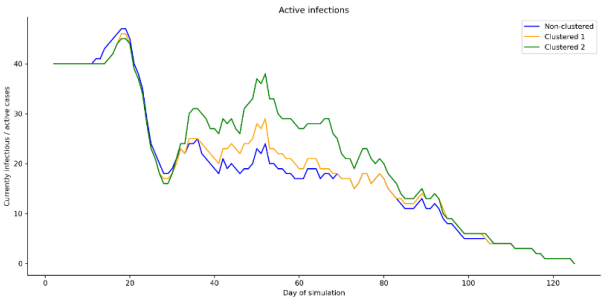

Fig 1: Active infections per day of the simulation run in County Wiltshire for all three worlds

In Fig 1 you can see the active infections per day of the simulation in the county Wiltshire for all 3 different worlds with increasing level of edge match rate, using the original world realistic JUNE2 for measles as described in the method.

World	Total Infections	Hospitalisations	Deaths
Non-Clustered	106	20	0
Cluster 1	110	18	0
Cluster 2	124	22	0

Table 2: Statistics from the simulation run in County Wiltshire for all three worlds

4. Discussion

Evaluation of Results

Overall, Fig 1 shows the more clustered the world the greater the outbreak, with a greater number of cases. The maximum number of cases for the cluster 2 world is 36 whereas for the non-clustered it is 25 which is a 44% increase. Table 1 also shows a consecutive increase in the total infections as clustering increases, total infections being 17% larger for cluster 2 than the non-clustered world. For this difference to be a result of an uncertainty on total infections both the non-clustered and cluster 2 world would have to have an absolute uncertainty of 8 people which seems unrealistic meaning it is reasonable to assume this correlation is somewhat accurate. Table 1 and

Table 2 also suggests there is not a linear relationship between unvaccinated edge match rate and total cumulative infections. However, there are only three data points so this cannot be confidently asserted, future work could explore this correlation.
It must be noted as these results involve so few cases as these simulations were run on a small area, Wiltshire has a population of 510310, there is lots of noise and so uncertainty in these conclusions. Next time we would run the simulation for all the of England to reduce this.
To evaluate the severity of impact to public health our metric should relate to the number of life's impacted and gravity of this impact, death could be more heavily weighted than hospitalisation which is worse than infection alone. For example, an outbreak could have many more cases but fewer deaths and hence under some scales be better. The specific weightings of this in itself is an interesting question on ethics that needs to be considered when creating policy. As these outbreaks were not run on a large enough scale simulation to have lots of data, specifically there were no deaths, it is harder to draw an overall conclusion to which reality damages public health the most. Also, the hospitalisation rates do not follow a linear trend like total infections as seen in Table 1. In future works when there is more relevant data, we would like to come up with or research a more rigid metric for severity.
Originally the non-clustered world cases peak the most when introduced to the infection seed. Only after around the 30th day do cases begin to rise again and a characteristic outbreak occurs presumably due to the incubation period of secondary cases from the original seed being themselves infective.

Sources of error and limitations

Two broad sources of uncertainty should be considered when interpreting these results: uncertainty in the estimated effect of vaccination-status clustering, and uncertainty arising from differences between the models and the real world.

Uncertainty in the estimated effect of clustering
The first source of uncertainty relates to how accurately the simulations isolate the effect of vaccination-status clustering. Although the clustered worlds were designed to increase the edge-match rate, other characteristics of the social network may also influence transmission. The methods used to create the clustered worlds were relatively simple and were primarily designed to increase the proportion of connections between individuals with the same vaccination status. Consequently, the resulting networks may not accurately represent the way vaccination status is clustered in real populations. This means that any observed difference between the clustered and non-clustered worlds should be interpreted as the effect of the particular clustering structures implemented in this study, rather than as a definitive estimate of the effect of clustering in the real world.

Limitations of the representation of the real world
The second source of uncertainty arises from the representation of the population in MAY and the measles disease in JUNE.
The attribution of vaccination status in MAY could be made more realistic by distinguishing between MMR1, MMR2 and subsequent booster doses, including their individual uptake rates and efficacy in JUNE. For example, in 2021, reported uptake differed substantially between MMR1 and MMR2, at approximately 94.3% and 86.6%, respectively. Vaccine effectiveness also varies depending on the dose, with estimates reported in the literature ranging from approximately 93% to 97% for prevention of measles of MMR1 versus MMR2 respectively [14]. This is a potentially important source of uncertainty because accurate vax cluster!!!!!!!!!
It would also have been preferable to incorporate vaccination coverage data at a smaller geographical scale. Government vaccination data was not available at smaller geographical levels apart from by GP practice but this would have required us reconfiguring JUNE for GP constituency geographical units which was not done due to time constraints. This could produce more realistic local variation in vaccination coverage and therefore improve the representation of vaccination-status clustering per geographical unit.
Catch-up vaccination campaigns could also have been considered in greater detail. Some catch-up activity may be indirectly represented in the available coverage data, particularly where vaccination coverage was recorded at age five. However, the original catch-up activity associated with the introduction of measles vaccination in 1968 was not investigated in sufficient detail. Instead, the model assumed that individuals who were older than one year in 1968 had not received vaccination. This is a simplifying assumption and may result in an inaccurate representation of immunity among older age groups. A study was identified that incorporated catch-up vaccination campaigns [14], but it was not used because there was insufficient time to validate its methodology and its data differed substantially from the UK data used in this study but this may be useful for use in future works. Nevertheless, the discrepancy suggests that further investigation of historical catch-up vaccination campaigns could be important for improving the model.
The assignment of natural immunity was another substantial simplification. A value of 85% was used to represent the proportion of individuals considered naturally immune. This was partly based on UK government policy, which treats individuals of this age as effectively immune, and on the high transmissibility of measles, which would suggest that a large proportion of individuals would have been exposed before widespread vaccination. An alternative approach was investigated in which the probability of natural immunity was estimated for each age group from examining probability of being infected and recovering in each year of life from recorded measles cases. However, this produced substantially lower levels of immunity than expected based on the original assumptions. One possible explanation is that historical measles cases were underreported, particularly in periods when surveillance and case ascertainment were less developed. Future work could therefore investigate methods for estimating the probability of natural

immunity by age using adjusted historical case rates or other epidemiological evidence.
The epidemiological parameters used in the JUNE measles configuration were derived from published sources, including the CDC, WHO, NEJM and other epidemiological literature, rather than being fitted to a specific observed measles outbreak. They should therefore be considered a reasonable starting configuration rather than a validated calibration of the model. In particular, the contact-matrix β scaling factors, uptake rate for outbreak-response vaccination campaigns and compliance rates associated with isolation policies are assumptions that could be calibrated against observed measles case, hospitalisation and mortality data from real outbreaks. Calibration against observed outbreak data would provide greater confidence that the model reproduces the epidemiological behaviour of measles under realistic conditions.

Future work

Firstly, the simulations could be run using these configurations for the whole of England as mentioned. This would provide a larger population and a greater number of cases reducing noises effect.
A further area of investigation would be the nature and extent of vaccination-status clustering. Under the time constraints of this study, the primary objective was to increase the edge-match rate. Future work could instead systematically vary the parameters controlling clustering and investigate how different degrees and structures of clustering affect measles transmission. This could allow thresholds of clustering associated with substantially increased outbreak risk to be identified. The realism of the clustering could also be improved by incorporating vaccination coverage data from smaller geographical areas, as discussed above.
Another approach would be to ensure that unvaccinated clusters have demographic characteristics that more closely reflect those observed in the real population. Groups with lower vaccination uptake may be associated with factors including healthcare accessibility, deprivation, rurality and cultural differences [15] [16]. Some of these factors are indirectly represented by the local-authority structure of the current model, as areas with different socioeconomic and demographic characteristics may have different vaccination coverage. However, demographic data are unlikely to fully capture individuals who refuse vaccination for specific ideological or belief-based reasons. These attitudes may also spread through social media or other online networks, meaning that physical-contact networks alone may not fully represent the mechanisms through which vaccine hesitancy can cluster.
An alternative direction for future work would be to investigate the vaccination-coverage threshold at which outbreaks can occur under different levels of clustering. Rather than comparing outbreaks at a single overall vaccination coverage, simulations could be run across a range of vaccination coverage levels for each clustering configuration. This would allow the relationship between overall vaccination coverage, the structure of vaccination-status clustering and outbreak risk to be investigated.
Finally, future work could investigate the implications of these results for immunisation policy, either through a review of the existing literature or by explicitly modelling different interventions. For example, some papers suggested specific geographically targeted interventions are beneficial in light of clustering, [11] this could be investigated to determine whether vaccination strategies tailored to local patterns of coverage and clustering produce different outcomes from uniform interventions. JUNE could potentially provide a framework for testing interventions under different local conditions, allowing strategies to be evaluated according to the specific characteristics of different populations and areas rather than assuming that the same intervention will have equal effectiveness everywhere.

5. Conclusions

Explicitly modelling social-network structure, rather than assuming uniform mixing, shows that clustering of unvaccinated individuals meaningfully increases measles outbreak size, even over the modest range of clustering achieved here: an unvaccinated edge-match rate rising from 0.29 to 0.82 produced a 17% increase in total infections and a 44% increase in peak active cases. This supports the view that population-level coverage targets, such as the WHO's 95% threshold, can understate real outbreak risk when unvaccinated individuals are not evenly distributed through a community, an interaction largely absent from standard herd-immunity models. As a single-county pilot with only three clustering configurations and no recorded deaths, these results are best read as a proof of concept for the MAY/JUNE clustering methodology, not a calibrated estimate of real-world risk.

References

[1] E. Haeuser, et al., “Global, regional, and national trends in routine childhood vaccination coverage from 1980 to 2023 with forecasts to 2030: a systematic analysis for the Global Burden of Disease Study 2023.” The Lancet 406, 10500 (2025). doi: 10.1016/S0140-6736(25)01037-2.

[2] UNICEF and WHO Europe, “European Region reports highest number of measles cases in more than 25 years – UNICEF, WHO/Europe.” UNICEF. https://www.unicef.org/press-releases/european-region-reports-highest-number-measles-cases-more-25-years-unicef-whoeurope, (2025). Accessed 3 September 2026.

[3] UK Health Security Agency, “Measles: historic confirmed cases, notifications and deaths.” GOV.UK. https://www.gov.uk/government/publications/measles-historic-confirmed-cases-notifications-and-deaths, (2026). Accessed 3 September 2026.

[4] Royal College of Paediatrics and Child Health, “Vaccination in the UK: Access, uptake and equity.” RCPCH Commission on Immunisation policy report. https://www.rcpch.ac.uk/sites/default/files/2025-

06/Vaccination-in-UK-access-uptake-equity_policy-report-2025.pdf, (2025). Accessed 18 September 2026.

[5] UK Health Security Agency, "Childhood vaccination coverage." UKHSA data dashboard. https://ukhsa-dashboard.data.gov.uk/cover/childhood-vaccinations, (2026). Accessed 3 September 2026.

[6] R.M. Anderson and R.M. May, Infectious Diseases of Humans: Dynamics and Control (Oxford University Press, Oxford, 1991). 1st edition. ISBN: 978-0198540403.

[7] L.G. Alvarez Zuzek, et al., "Spatial clustering in vaccination hesitancy: The role of social influence and social selection." PLOS Computational Biology 18, e1010437 (2022). doi: 10.1371/journal.pcbi.1010437.

[8] E. Kuylen, et al., "Clustering of susceptible individuals within households can drive measles outbreaks: an individual-based model exploration." Scientific Reports 10, 20151 (2020). doi: 10.1038/s41598-020-76746-3.

[9] D. Klinkenberg, et al., "Social clustering of unvaccinated children in schools in the Netherlands." Epidemiology and Infection 150, e200 (2022). doi: 10.1017/s0950268822001455.

[10] A. de Figueiredo, "Forecasting sub-national trends in COVID-19 vaccine uptake in the UK." medRxiv (2020). doi: 10.1101/2020.12.17.20248382.

[11] S.A. Truelove, et al., "Characterizing the impact of spatial clustering of susceptibility for measles elimination." Vaccine 37, 732-741 (2019). doi: 10.1016/j.vaccine.2018.12.012.

[12] D. Klinkenberg, et al., "Social clustering of unvaccinated children in schools in the Netherlands." Epidemiology and Infection 150, e200 (2022). doi: 10.1017/s0950268822001455.

[13] UK Health Security Agency, "Cover of vaccination evaluated rapidly (COVER) programme: April 2024 to March 2025." GOV.UK. https://www.gov.uk/government/statistics/cover-of-vaccination-evaluated-rapidly-cover-programme-april-2024-to-march-2025, (2026). Accessed 18 September 2026.

[14] M. Edelstein, et al., "Triangulation of measles vaccination data in the United Kingdom of Great Britain and Northern Ireland." Bulletin of the World Health Organization 97, 754–763 (2019). doi: 10.2471/BLT.18.229138.

[15] A. Noroozi, et al., "Predictors of Childhood Vaccination Uptake in England: An Explainable Machine Learning Analysis of Longitudinal Regional Data (2021-2024)." arXiv 2504.13755 (2025). doi: 10.48550/arXiv.2504.13755.

[16] B. MacKenna, et al., "Trends, regional variation, and clinical characteristics of COVID-19 vaccine recipients: a retrospective cohort study in 23.4 million patients using OpenSAFELY." medRxiv 2021.01.25.21250356 (2021). doi: 10.1101/2021.01.25.21250356.