



Fine-scale spatial clustering of measles nonvaccination that increases outbreak potential is obscured by aggregated reporting data

Nina B. Masters^{a,1}, Marisa C. Eisenberg^a, Paul L. Delamater^b, Matthew Kay^c, Matthew L. Boulton^{a,d,2}, and Jon Zelner^{a,e,1,2}

^aDepartment of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48109; ^bDepartment of Geography, University of North Carolina, Chapel Hill, NC 27514; ^cSchool of Information, University of Michigan, Ann Arbor, MI 48104; ^dDepartment of Internal Medicine, Division of Infectious Disease, University of Michigan Medical School, Ann Arbor, MI 48109; and ^eCenter for Social Epidemiology and Population Health, University of Michigan School of Public Health, Ann Arbor, MI 48109

Edited by Alfred Sommer, Johns Hopkins University, Baltimore, MD, and approved September 2, 2020 (received for review June 11, 2020)

The United States experienced historically high numbers of measles cases in 2019, despite achieving national measles vaccination rates above the World Health Organization recommendation of 95% coverage with two doses. Since the COVID-19 pandemic began, resulting in suspension of many clinical preventive services, pediatric vaccination rates in the United States have fallen precipitously, dramatically increasing risk of measles resurgence. Previous research has shown that measles outbreaks in high-coverage contexts are driven by spatial clustering of nonvaccination, which decreases local immunity below the herd immunity threshold. However, little is known about how to best conduct surveillance and target interventions to detect and address these high-risk areas, and most vaccination data are reported at the state-level—a resolution too coarse to detect community-level clustering of nonvaccination characteristic of recent outbreaks. In this paper, we perform a series of computational experiments to assess the impact of clustered nonvaccination on outbreak potential and magnitude of bias in predicting disease risk posed by measuring vaccination rates at coarse spatial scales. We find that, when nonvaccination is locally clustered, reporting aggregate data at the state- or county-level can result in substantial underestimates of outbreak risk. The COVID-19 pandemic has shone a bright light on the weaknesses in US infectious disease surveillance and a broader gap in our understanding of how to best use detailed spatial data to interrupt and control infectious disease transmission. Our research clearly outlines that finer-scale vaccination data should be collected to prevent a return to endemic measles transmission in the United States.

measles | epidemiology | simulation model | disease dynamics | vaccination clustering

The Global Vaccine Action Plan set a goal of measles elimination in five World Health Organization (WHO) regions by 2020. However, re-emergence of measles in ostensibly post-elimination settings and slow progress in endemic settings have thwarted these international control efforts, with 187/194 (96%) of WHO member states reporting measles cases in 2019 (1). Globally, the first half of 2019 witnessed the most reported measles cases since 2006, with 791,143 suspected cases in 2019, compared to 484,077 in 2018, a 63% increase (2, 3). Recent drops in vaccination coverage have threatened the WHO American Region's measles elimination status, attained in 2000 (4).

In the United States, a 2014 measles outbreak originating at Disneyland was the largest, most-publicized outbreak event since the declaration of elimination (5). Majumder et al. estimated that the vaccination rate among those infected in this outbreak was between 50 and 86%, much lower than California's state average of 92.8% ($\pm 3.9\%$) (6, 7) and the national average of 91.9% (6). Local variability in measles vaccine coverage likely contributed to the size of the outbreak, with Pingali et al. finding

93 regions, or “coldspots,” encompassing 31% of California's primary schools, where many kindergarteners were not up-to-date for recommended vaccinations (8). This demonstrates how fine-scale clustering of nonvaccination can increase the likelihood of outbreaks while “flying below the radar” of statewide statistics. Such exemption clusters have also been responsible for outbreaks of pertussis in Michigan (9) and Florida (10) and of measles in Oregon (11). Vaccination heterogeneity is a key threat to measles elimination and control: in the United States alone, 2019 saw 1,282 cases of measles in 31 states, the most since 1992, making a return to endemic measles likely if these trends are not rapidly reversed (12).

Redefining Vaccination Coverage Targets

To meet global elimination goals, WHO has set vaccination coverage targets of 95% for the first and second doses of the pediatric measles-containing vaccine (MCV) (13, 14). High coverage of MCV is necessary because measles is highly contagious

Significance

The United States witnessed large, persistent measles outbreaks in 2019, nearly losing its elimination status, despite achieving national measles vaccination coverage above the World Health Organization recommendation of 95%. Previous research showed that measles outbreaks in high-coverage contexts are driven by spatially clustered nonvaccination, locally depressing immunity levels. We perform a series of computational experiments to assess the impact of clustering of nonvaccination on outbreak potential and how disease risk predictions might be biased by measuring vaccination rates at coarse spatial scales. When nonvaccination is locally clustered, reporting aggregated data can result in substantial underestimates of outbreak risk. This research illustrates that finer-scale vaccination data should be collected to prevent a return to endemic measles transmission in the United States.

Author contributions: N.B.M., M.C.E., M.L.B., and J.Z. designed research; N.B.M. performed research; M.K. assisted with figure development and data visualization; P.L.D., M.K., and J.Z. contributed analytic tools and support; N.B.M., M.C.E., and J.Z. analyzed data; and N.B.M. and J.Z. wrote the paper with assistance from all authors.

Competing interest statement: P.L.D. has received research funding from Merck for an unrelated project.

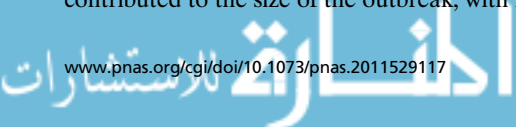
This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: mastersn@umich.edu or jzelner@umich.edu.

²M.L.B. and J.Z. contributed equally to this work.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011529117/-DCSupplemental>.



with a basic reproduction number (R_0) of 12 to 18, among the highest known values, although estimates of the R_0 are quite variable (15, 16). Although the measles–mumps–rubella vaccine is highly immunogenic, with two doses conferring 97% protection (17), the proportion of the population that needs to be vaccinated or have natural immunity from prior disease to prevent outbreaks, known as the critical vaccination fraction (V_c)*, is nonetheless very high, around 94 to 95% (7, 18). A key assumption underlying most estimates of V_c is that the population is evenly mixed and that all susceptible, infectious, and immune individuals contact each other with equal probability. However, when nonvaccinated individuals are geographically clustered, this formula can underestimate V_c by as much as 3%, so that outbreaks remain possible despite statewide vaccination coverage targets being met or exceeded (19).

What Is the Right Scale of Surveillance?

While the role of heterogeneous mixing and infectiousness in populations in increasing outbreak risk for vaccine-preventable diseases (VPDs) has been demonstrated in prior studies (8, 9, 20–26), public health surveillance systems typically report vaccination coverage at the county and state level, obscuring this risk. For example, in Michigan, 4.54% of kindergarteners statewide had vaccination waivers for the 2018 to 2019 school year, meeting the WHO threshold of 95% overall vaccination. That same year, a large measles outbreak occurred in Oakland County, where the waiver rate was 7.14%, but school-district waiver rates ranged from 0 to 23.4%, and two schools reported >50% waivers (Fig. 1A). Additionally, many clinical preventive services have been suspended in the wake of the COVID-19 pandemic, with many individuals fearful of doctors and nonemergent visits delayed, which has led to plummeting pediatric vaccination rates nationally; an estimated 400,000 fewer MCV doses were ordered from January 6 to April 18, 2020 than were ordered over the same period last year (27). In Michigan, vaccination rates have dropped to dangerously low levels for measles in particular, with only 70.9% of 16-mo-old children currently up to date for MCV, down from 76.1% last year (28). As such, understanding the role that clustering of nonvaccination for measles plays in outbreak risk is especially important as existing clusters are likely to be magnified by plummeting pediatric vaccination rates. Furthermore, elucidating at what scale aggregate surveillance data are too unreliable to capture such fine-scale heterogeneity will be necessary to successfully implement control strategies for both emergent measles outbreaks and ongoing COVID-19 infections. Because granular vaccination data are not readily available to researchers, this paper uses a simplified, schematic model to provide proof-of-concept and understand the mechanisms by which clustering of nonvaccination, and aggregation of such data, impact population health and outbreak risk.

Methods

Simulated Environment. To understand how aggregation of surveillance data may impact outbreak risk assessment, we constructed a spatial measles transmission model in a simulated city of 256,000 people laid out on a 16×16 grid. Our model includes four nested levels analogous to those found in real vaccination data: 1,000-person blocks (1 cell), 4,000-person tracts (4 cells), 16,000 person neighborhoods (16 cells), and 64,000 person quadrants (64 cells). This configuration allowed us to fix the population average vaccination coverage while varying the spatial distribution of coverage at multiple scales to isolate the specific impact of clustering at different levels. Our model encoded contact between individuals within each block and with contiguous blocks, as school-aged children have primarily local contacts.

*This can be calculated as $R = R_0 * ((1 - (V_c * VE)))$, where $1 - (V_c * VE)$ is the proportion of the population that remains susceptible after vaccination. The V_c can be expressed in terms of infectiousness and vaccine efficacy: $V_c = 1 - 1/ROVE$, where VE is the proportion of vaccinated individuals protected from disease, or the vaccine efficacy (19).

Contact between blocks used queen's contiguity, in which all surrounding cells are considered neighbors (cells which share an edge or a corner with the index cell, such that cells in the center of the grid would have eight neighbors). The spatially dependent force of infection was split such that 50% of transmission occurred within cells, and 50% of transmission was split between all neighboring cells equally. We fixed population-wide measles vaccine coverage at the WHO threshold for measles (95%) while varying the spatial distribution and intensity of local clustering of vaccination (Fig. 1B and C). In all simulations, R_0 was fixed at 16, and the average community vaccination coverage was 95%, which represents a scenario in which a completely homogeneous model would predict that an outbreak is not possible.

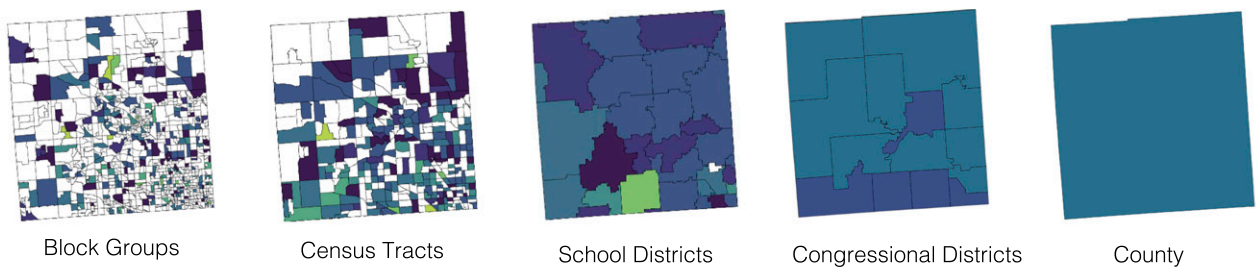
Clustering Motifs of Nonvaccination. Clustering motifs were generated using stratified random sampling at the quadrant, neighborhood, tract, and block levels to produce different landscapes and spatial distributions of nonvaccinators within this population. The motifs were created by sampling 12,800 (5% of the total population) total unvaccinated individuals into the environment's individual cells with probability proportional to the intensity of clustering at each of the four nested spatial levels, allowing us to explore the difference in outcomes between motifs with equivalent vaccination coverage but with large-scale vs. fine-scale clustering and vice versa. A depiction of this process is shown in *SI Appendix, Fig. S1*. In all simulations, we assigned the top-left quadrant to be the most highly clustered quadrant and explored scenarios in which 85% of the nonvaccinators were in that quadrant and the remaining 15% were evenly distributed among the remaining quadrants, to the least clustered case in which a quarter of nonvaccinators were deposited in each quadrant. Three additional sets of probabilities generated the full set of clustering motifs: 70, 58, and 40% of nonvaccinators in the top-left quadrant, distributing the remaining 30, 42, and 60% of nonvaccinators evenly among remaining quadrants, respectively. Of the 625 potential clustering motifs representing every combination of probabilities, 336 were consistent with a scenario of 95% vaccination coverage at the population level, i.e., where the proportion of nonvaccinators in each cell was ≤ 1 .

Model Structure. We modeled transmission using a deterministic, compartmental, Susceptible-Infected-Recovered (SIR) model where the clustering motifs representing different landscapes of nonvaccination were used as initial conditions for the compartmental transmission model (*SI Appendix, Fig. S2*) (19, 23). For simplicity, no vital dynamics were included due to a simulation time of 1 y.

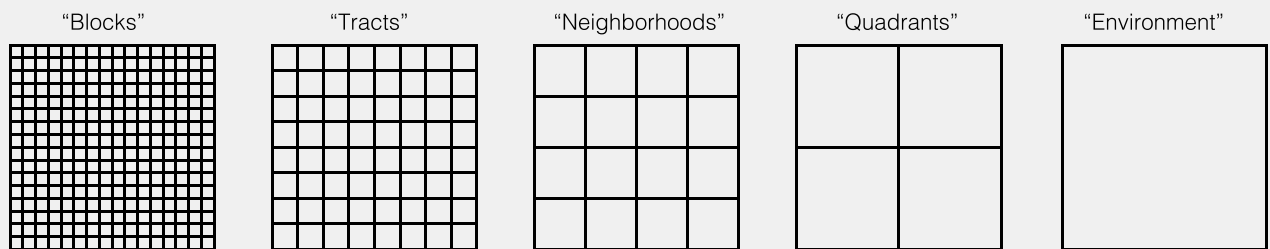
Measuring Clustering. Clustering of nonvaccination in each motif was measured using Moran's I , a measure of global spatial autocorrelation (29), and the isolation index, a measure of the proportion of within-group contacts in a population with two main subgroups (i.e., vaccinated and unvaccinated) (30). Moran's I (*SI Appendix, Supplementary Methods, Eq. S2*) ranges from -1 to 1 , where a value of -1 corresponds to perfect clustering of dissimilar values (e.g., high-low clustering), 0 indicates no autocorrelation, and 1 indicates perfect clustering of similar values (e.g., high-high) (29). By contrast, the isolation index (*SI Appendix, Supplementary Methods, Eq. S3*) measures exposure, specifically the extent to which nonvaccinated individuals contact each other: if there is little systematic separation of the groups, the value of isolation will approach the global percentage of nonvaccinators and will approach 1 when nonvaccinators are highly concentrated in one geographic location (30).

Measuring Aggregation Effects. To examine how the resolution of vaccination data impacts model-based risk predictions, we created counterfactual simulations to see how much error was incurred by coarsening the spatial vaccination data. This is analogous to quantifying the "type M" errors of magnitude described by Gelman et al. (31) The clustering motifs described above were regarded as the "true" vaccination data, with resolution at the block level. The grid was coarsened by moving up the levels of aggregation shown in Fig. 1: block-level data were aggregated up to the tract level, where the four cells that belong to each tract were averaged and nonvaccinators were redistributed to the contributing cells. This process was then repeated at the neighborhood and quadrant level. Once these aggregated motifs were generated, we ran the SIR model on the coarsened grids to see how the predicted case burden differed from that of the block-level, true data, using the difference in these predictions to characterize the bias from aggregating this data.

A Childhood Vaccination Waiver Rates from Oakland County, 2018, at Different Levels of Aggregation



B Aggregation levels of 'blocks' – each individual cell (each cell contains 1,000 individuals), tracts' – groups of four cells (4,000 people), 'neighborhoods' – groups of 16 cells (16,000 people), and 'quadrants' – groups of 64 cells (64,000 people), make up our simulated 'environment' (256,000 total people) – the level at which an overall vaccination level is fixed.



C Example data from one set of simulated vaccination conditions, with overall vaccination level across the environment at 95%

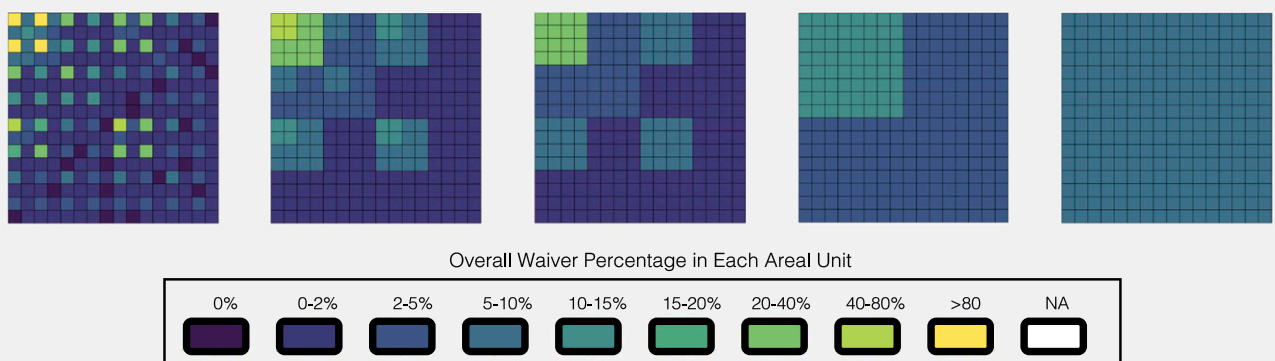


Fig. 1. Impact of spatial aggregation of vaccination data on coverage estimates. (A) Vaccination coverage data from Oakland County, Michigan, at five different levels of spatial scale: block groups, census tracts, school districts, congressional districts, and the county level. (B) Schematic illustrating the spatial model used in this study with a 256-grid cell environment, each of which contains 1,000 individual people, divided into spatial scales of "blocks" (all grid cells), "tracts" (groups of 4 cells), "neighborhoods" (groups of 16 cells), "quadrants" (groups of 64 cells), and, finally, the entire vaccination "environment" (all 256 cells aggregated to one unit), the level at which overall vaccination percentages are fixed for analysis (i.e., at 95 and 98%). (C) Example data from one simulated set of vaccination conditions, fixed at 95% overall vaccination, showing the impact of aggregation to these different scales on loss of granularity of block-level data.

Statistical Analysis and Simulation Protocol. Simulations were conducted in R version 3.6.0 using the *deSoIve* package. The SIR model was simulated across the clustering motifs, and the outbreak potential and cumulative incidence were calculated for four scenarios: an initial seed case dropped in the center of each quadrant to capture spatially heterogeneous outcomes based upon the location of the introduced case. The attack rate (AR) after 1 y of simulation time was calculated, with $AR = 1\text{ y cumulative incidence}/\text{initial susceptibles}$. For each motif, 10 simulations were run for a seed case dropped in each quadrant to capture stochastic variation due to the multinomial probability distribution used to generate the motifs themselves, generating 40 simulated runs for each motif. The Moran's I and isolation index of the starting motifs were calculated by generating the motifs 30 times each and taking the average value to account for sampling differences. The isolation

index was normalized using the following formula: $\text{normalized isolation} = (\text{isolation index} - \text{minimum isolation}) / (\text{maximum isolation} - \text{minimum isolation})$ for easier interpretation. For assessing outbreak potential, we defined an outbreak as a simulation with five or more secondary cases. Code used to generate all simulations, motifs, and datasets can be found at <https://github.com/epibayes/Measles-Spatial-Clustering-and-Aggregation-Effects/> (32).

Sensitivity Analysis. Numerous sensitivity analyses were conducted to evaluate the robustness of our findings against different assumptions. Our baseline model uses density-dependent transmission in which the force of infection for neighbor-driven transmission is dependent on the number of neighbors, and we assessed the model instead with a frequency-dependent force of infection. Additionally, the baseline model assumed that 50% of

transmission occurred within the cells, and 50% was divided between neighboring cells. We varied this percentage of between-cell transmission from 10 to 75% to examine the impact of changing neighbor-driven transmission. Finally, the overall percentage of vaccination was modified from

the baseline scenario of 95%, with sensitivity analyses using 94, 98, and 99% overall vaccination (yielding a total number of possible motifs that did not exceed cell-level populations over 1,000 of 296, 543, and 620, respectively). We also assessed combinations of different vaccination percentages and

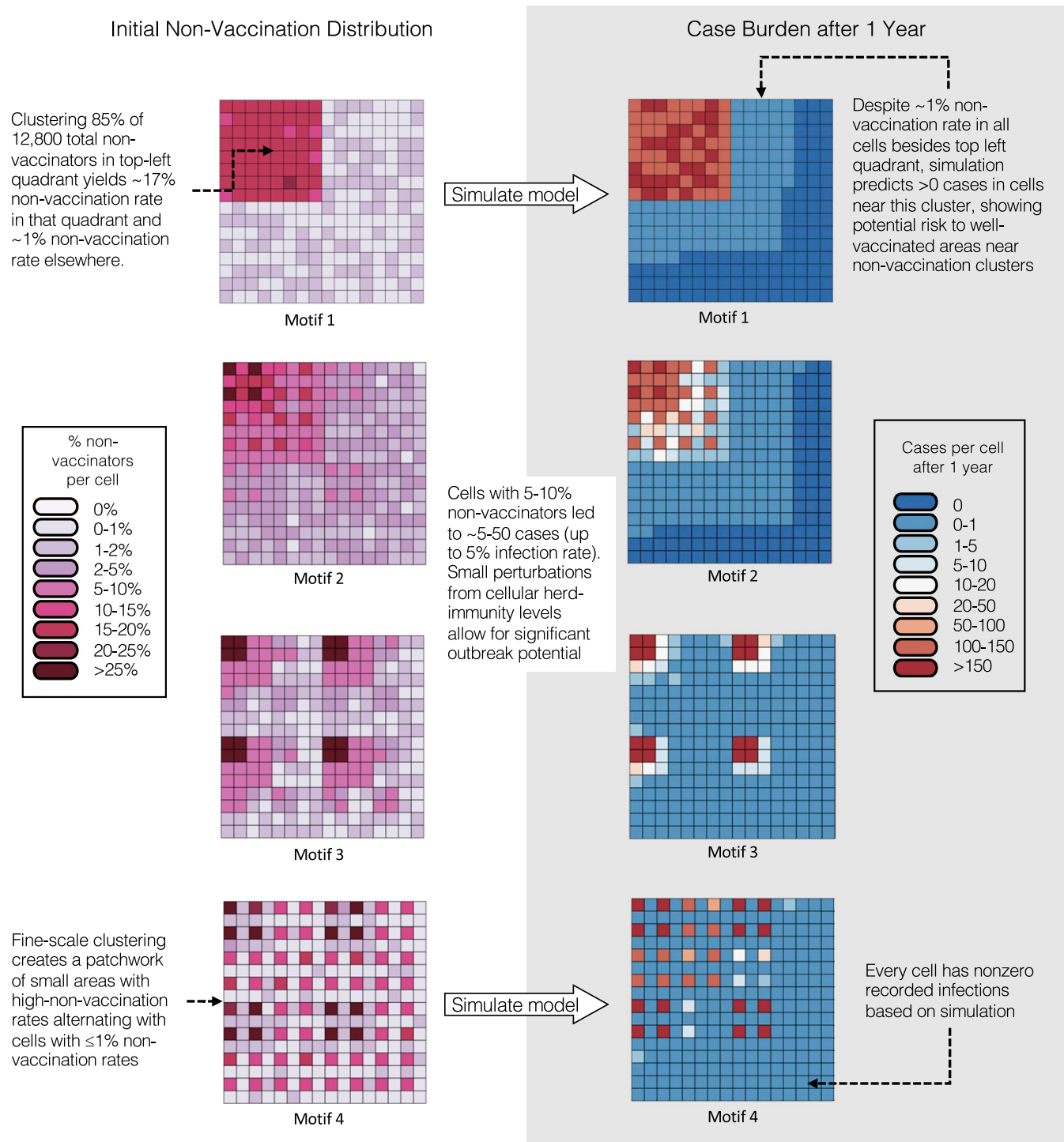


Fig. 2. Distribution of nonvaccination at baseline (*Left*) and case burden after 1 y (*Right*) for four selected clustering motifs with 95% overall vaccination coverage ($n = 12,800$ nonvaccinated individuals at baseline). In each case, a seed infection was introduced into the top left quadrant and cases spread throughout and beyond the demarcated boundaries of high-risk unvaccinated regions, as can be seen for motifs 1 and 2, where there is a band of nonzero case burden around the high-risk cluster of nonvaccinated individuals, even though the vaccination rate in this region is 0 to 1%. For motif 3, the four foci of nonvaccination with >25% unvaccinated proportions are the hardest hit in terms of attack rate after 1 y, with >150 cases per cell (>15% attack rate), but the surrounding cells, with 5 to 10% nonvaccination, see 10 to 50 cases after 1 y, representing a 1 to 5% attack rate. Finally, for motif 4, a fine-scale clustering pattern creates local cells with high attack rates, but all cells have a nonzero attack rate.

between-cell transmission rates to explore the impact of varying both parameters at once.

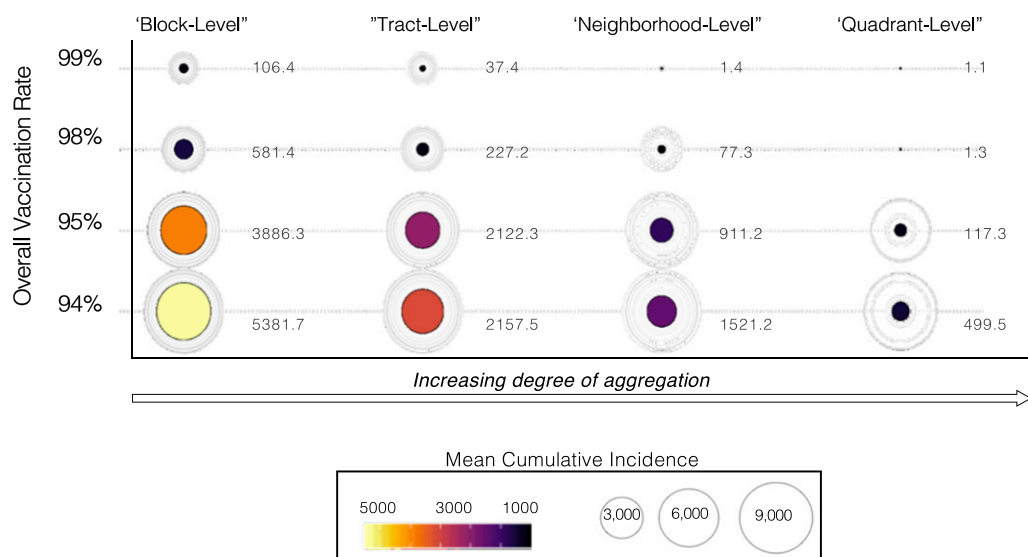
Results

Impact of Clustering on Outbreak Probability and Size. The intensity of clustering of vaccination and contact between nonvaccinators was assessed using Moran's I (29) and the isolation index (30). In both univariate and multivariate models, for 95% overall vaccination, a change from the minimum to maximum values of normalized isolation was associated with an 80% increase in AR (~7,325 cases), while no association was observed for Moran's I (SI Appendix, Table S1). This suggests that isolation better captures the central role of clustering of susceptible individuals than does Moran's I, which is agnostic about the nature of clustering measured (i.e., of nonvaccination or vaccination).

Impact of Clustering on Outbreak Risk and Magnitude. Simulations from this model at 95% coverage across all possible clustering motifs ($n = 296$) yielded an average cumulative AR of 35.6% (SI Appendix, Table S4). Sensitivity analyses evaluating the cumulative incidence and AR at 94, 98, and 99% coverage showed that large outbreaks were possible at all coverage rates when nonvaccination was spatially clustered. By contrast, a full environment-level simulation (with spatially randomly distributed nonvaccinators, i.e., no encoded clustering), revealed that, at 95% vaccination coverage and above, there was fewer than 1 secondary case, and only 1.24 secondary cases were observed for 94% overall vaccination, indicating that herd immunity is upheld when there is no spatial clustering of nonvaccination (SI Appendix, Table S5). In all simulations, when the initial case was seeded in the quadrant inhabited by the majority of nonvaccinators, a larger outbreak was predicted as compared to seeding cases in the other quadrants, with introductions to the quadrant farthest in cartesian distance from the low-vaccination area, resulting in the fewest overall cases and the longest time-to-peak of cases. Most cases occurred in cells with low vaccination rates, although there was spillover to adjacent cells due to high levels of infection pressure from their low-coverage neighbors (Fig. 2). Sensitivity analyses of frequency-dependent transmission yielded similar cumulative incidence counts to the density-dependent baseline model (SI Appendix, Table S6).

Our simulations consistently showed that increasing clustering at each level of aggregation (blocks, tracts, neighborhoods, and quadrants) corresponded to a higher cumulative incidence of cases (SI Appendix, Figs. S7–S10). In addition to exploring the outbreak size as an outcome, we evaluated outbreak probability, defining three thresholds for an outbreak: 5, 10, and 20 cases over the course of 1 y. For 94% overall vaccination, 93.5% of simulation runs yielded outbreaks (defined as five or more cases), and there was a 92.3% probability of an outbreak with a threshold of 20 cases (SI Appendix, Table S7, and Figs. S15 and S16). For 95% overall vaccination, 89.0% of simulation runs generated a 5+ case outbreak, and 87.4% of simulation runs generated a 20+ case outbreak. For 99% vaccination coverage, the outbreak probability was much lower: 19.3% of simulation runs generated 5 or more cases, and 18.1% of runs generated 20 or more cases. These results show that outbreak probability decreases as coverage increases, yet in this clustered landscape of nonvaccination, even for 99% overall vaccination rates, there was a sizable proportion of simulation runs that were able to generate outbreaks.

Impact of Measurement Scale on Outbreak Size Prediction Errors. Our design analysis consisted of taking the block-level "ground truth" results of each simulation and aggregating these data up to each of the levels in Fig. 1. This resulted in large downward biases in both the simulated probability of observing outbreaks and their predicted size. The expected outbreak size for simulations at 95% overall vaccination was predicted to be 3,886 (AR = 30.4%) cases using unaggregated data; 2,122 (AR = 16.6%) using tract-level aggregation (45.4% reduction); 911 (AR = 7.1%) using neighborhood-level aggregation (76.5% reduction); and no secondary cases when aggregated to the quadrant level (99.9% reduction) (Fig. 3). Fig. 4 illustrates how this aggregation process obscures fine-scale spatial heterogeneity for three selected motifs, where three very different underlying patterns of nonvaccination and resultant outbreak potential converge to an identical motif with an expected AR of 51% when aggregated to the quadrant level. Across all motifs, the downward bias in the estimated isolation index increased with the intensity of aggregation (SI Appendix, Fig. S11).



The mean cumulative incidence (outbreak potential) across all clustering motifs, varying levels of overall vaccination (94%, 95%, 98%, 99%), shows notable reduced predicted outbreak size as data is aggregated up from the block-level, or 'ground truth' in simulation results, indicating that evaluating data at high levels of aggregation might not only miss potential local clusters of heterogeneity that can prompt a large-scale outbreak, but poorly predict outbreak risk.

Fig. 3. Impact of aggregation on estimated outbreak risk. Cumulative incidence at four different levels of vaccination coverage: 94, 95, 98, and 99%, including nonaggregated vaccination data (block-level) resolution to tract level (4-cell) resolution, to neighborhood (16 cell), and, finally, quadrant-level (64 cells) shows the reduction in estimated case burden as aggregation increases, a pattern that holds true across all levels of vaccination.

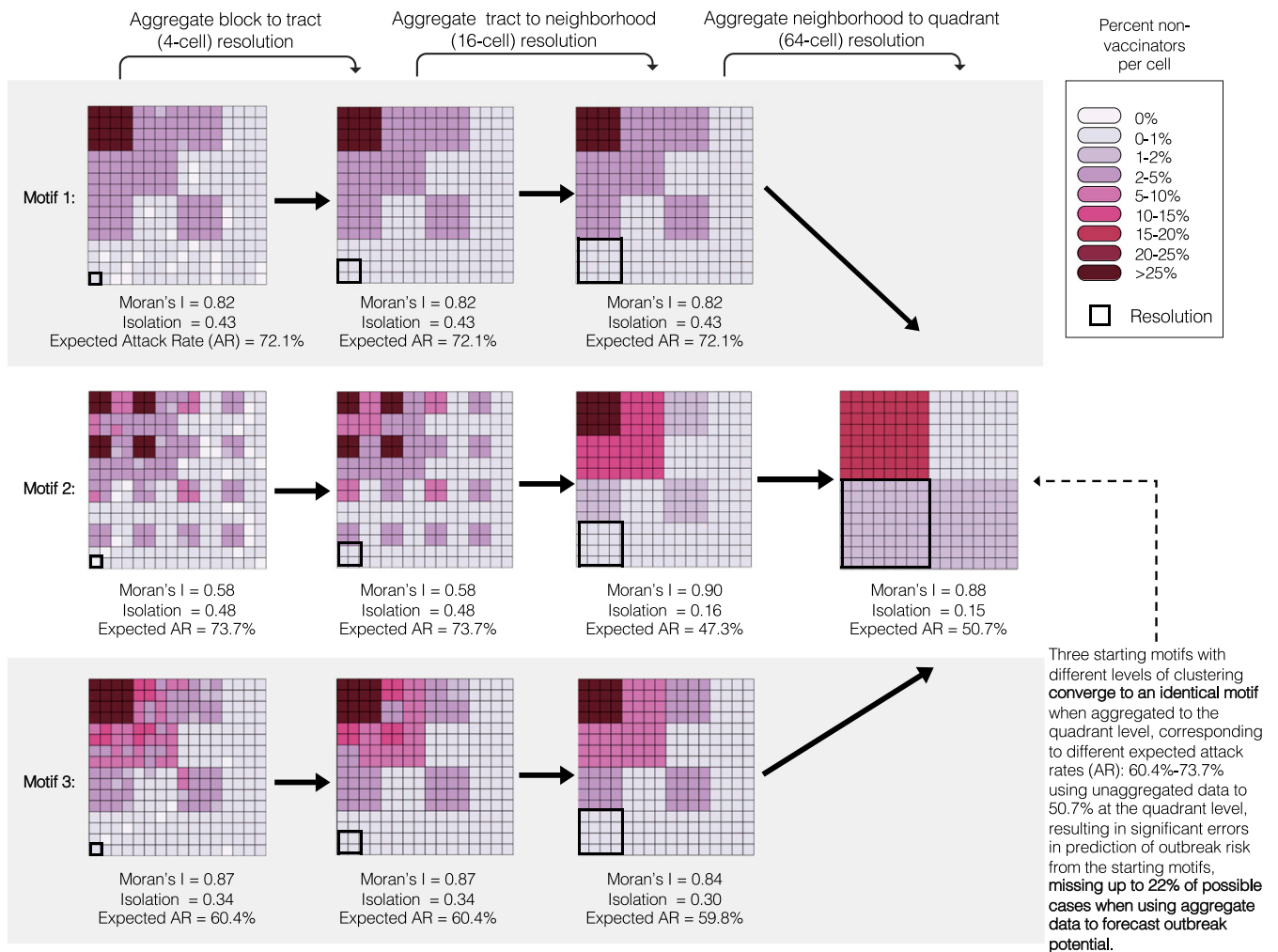


Fig. 4. Aggregated vaccine coverage systematically downplays outbreak risk. Aggregation from “true” 256-cell (block-level) resolution to tract level (4-cell) resolution to neighborhood (16-cell), and, finally, to quadrant-level (64-cell) resolution using a starting vaccination motif with overall vaccination at 95%. Three different motifs with different clustering patterns were subsequently aggregated up these three levels and yielded the same aggregate motif at the quadrant level, illustrating that large-scale vaccination data can mask significant heterogeneity at finer scales.

Aggregating vaccination data resulted in consistent underestimates of outbreak potential, with this bias growing as a function of the intensity of clustering in the input motif and the level of aggregation (Fig. 5). This trend was observed across all motifs, with models using data aggregated to the tract level predicting 41 to 65% fewer cases than simulations using nonaggregated data, and neighborhood-level aggregation resulting in 72 to 99% fewer cases detected (at 94 and 99% overall vaccination, respectively) (SI Appendix, Table S8). Quadrant-level aggregation resulted in greater than 90% reduction in detected cases at all tested vaccination levels. The proportion of expected cases plotted by isolation index of the initial motif can be seen in Fig. 5A; however, it is important to recall that an increasing isolation index corresponds to an increased simulated cumulative incidence, and thus higher levels of aggregation yield reduced accuracy in predicting outbreak potential, with greater numbers of cases missed, as vaccination landscapes become more clustered (Fig. 5B). This phenomenon was observed for all simulated vaccination levels (SI Appendix, Figs. S12–S14).

Discussion

Our results illustrate how failure to account for fine-scale heterogeneity in susceptibility can result in overly optimistic estimates

of outbreak potential. This mismatch between assumptions of homogeneous mixing which underlie the classical calculation of the V_c and the reality of local clustering of nonvaccination can lead to missed opportunities for preventing outbreaks. This is underscored by the finding that, even at 99% overall vaccination coverage, theoretically far exceeding the V_c for measles, deviations from homogeneity permitted outbreaks to occur. We found that increasing isolation of nonvaccination predicted an increased cumulative incidence at all vaccination levels, suggesting that the isolation index can be used to assess area-level outbreak vulnerability.

Additionally, our models show that aggregation-based estimates of outbreak risk relying on assumptions of homogeneity have the potential to mischaracterize the population at risk. As fine-scale vaccination data were aggregated, or “coarsened,” a large downward bias resulted in the projected number of cases, which grew with successive levels of aggregation. This has immediate implications for vaccine-coverage surveillance in the United States, highlighting that finer-scale data are needed to fully understand community susceptibility to outbreaks of measles and other VPDs. This accords with Truelove et al. (19) and Brownwright et al.’s suggestions (33) that setting the classical V_c as a national or state-wide vaccination target may ultimately

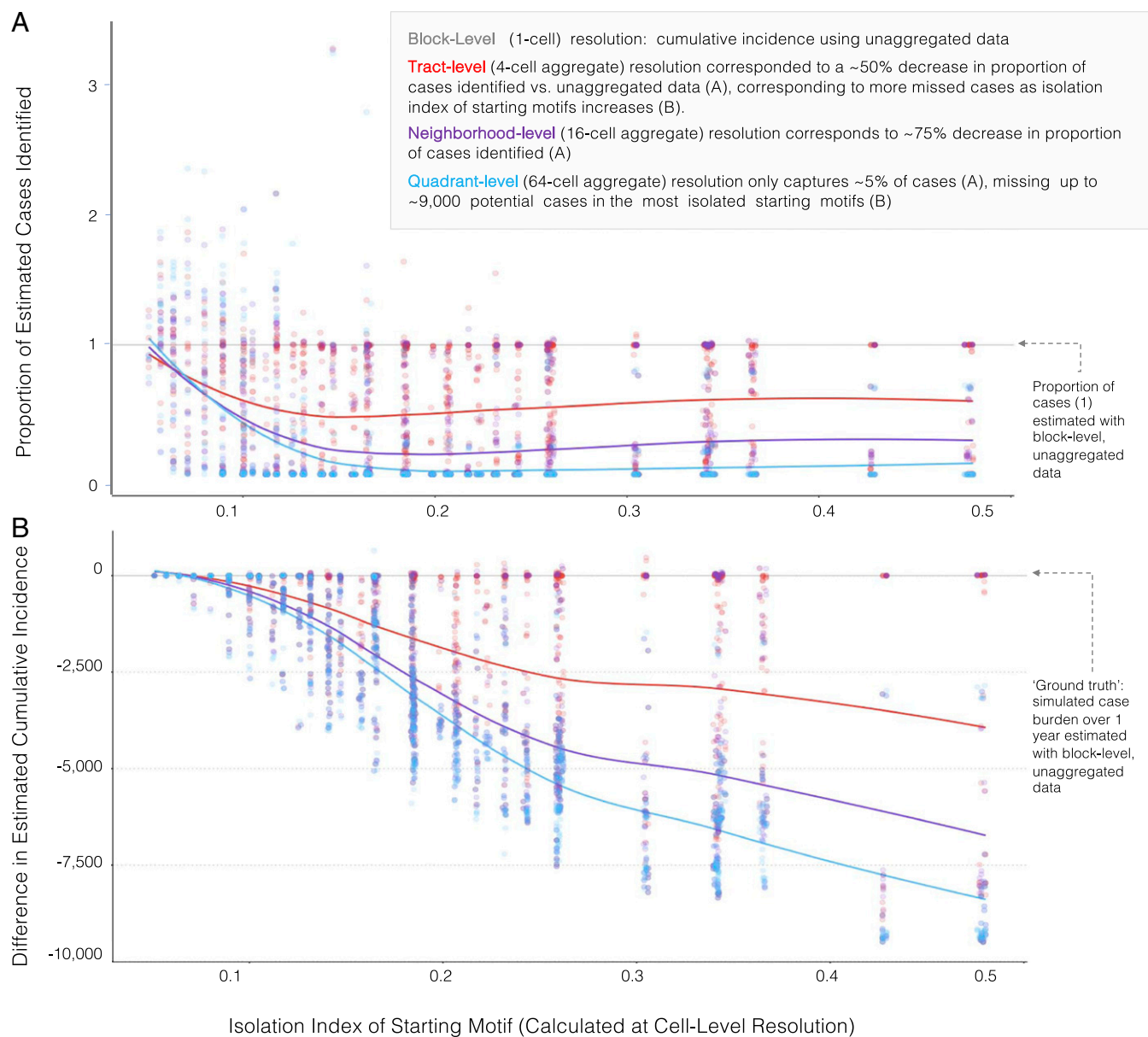


Fig. 5. Underestimate of outbreak risk grows with intensity of isolation of nonvaccinators. (A) Proportion of estimated cases identified, treating the block-level or individual-cell level simulation results as “truth,” in gray, when motifs are aggregated to the tract, neighborhood, and quadrant levels, sorted by the isolation index of the starting motif. (B) Difference in number of estimated cases, or cumulative incidence, by aggregation level and isolation index of initial motif, illustrating greater loss in predicted number of cases as both aggregation level and isolation index increase.

permit endemic transmission, necessitating a greater focus on assessments of finer-scale vaccination levels. Similarly, Tatem (34) argues that fine-scale analysis can better highlight communities at risk, although public health surveillance would need to be strengthened and enhanced, requiring a greater structural investment for this to be carried out effectively. Additionally, as shown in Fig. 14, regions without available vaccination data are often aggregated up into areal estimates of vaccination coverage, propagating errors associated with this missingness upward, which only further highlights the need for collection and dissemination of finer-scale vaccination data in order to make informed decisions about populations at risk.

An important caveat is that, while vaccination data are collected at the school-level for entry requirements, publicly released data instead are typically aggregated to the county- or

state-level despite the existence of finer-scaled data, representing a lost opportunity for improving surveillance. Leslie et al. found that only 20 US states report school-level data, 4 report school-district level data, 19 report county-level data, and 2 report health-department level data, but only a subset ($n = 26$) provide such data online, with 14 states providing data only after onerous Freedom of Information Act Requests (35). Additionally, the Centers for Disease Control and Prevention receives state-level vaccination data, which is far from the granular scale needed to set national policies that are sensitive to local vulnerability to measles (36).

Identifying the scale at which vaccination data are reported and available for analysis is not straightforward and comes with important trade-offs between privacy, feasibility, and cost. Many policy benchmarks are set at the national level, which may fail to

account for transmission dynamics playing out on a smaller scale, as coverage estimates of large regions cannot assume herd immunity is maintained at the scale of transmission. When defining such a spatial scale, relevant considerations comprise the potential intervention, the scale of surveillance, the reality of obtaining high-quality granular data, and the level at which vaccination coverage estimates are meaningful and actionable.

A number of different spatial scales have been explored in the literature, with notable heterogeneity in vaccination coverage identified at the subcontinental level, subnational (33), and regional levels (37). If the geographic level of data is mismatched to the scale of an intervention (38), reliance on aggregated data may result in diminished effectiveness of aid and interventions, leading to erroneous conclusions about what works for preventing VPD outbreaks (39). To address varied findings at different levels of analysis, some authors have also attempted to use multiple spatial scales, although such studies have yielded poor predictive ability (40, 41). At the finest spatial scales, such as human individual movement (42, 43) or mobility data using cell phone records (44), there is significant potential for the introduction of too much noise, yielding fewer informative results (45). As such, it is important to acknowledge that more research must be done to elucidate a feasible and actionable spatial scale to evaluate vaccination coverage, especially in countries nearing measles elimination where significant heterogeneity may undermine elimination efforts if unidentified.

Strengths and Limitations. This study has many strengths. Much of the literature surrounding spatial clustering of nonvaccination utilizes complex methods of identifying “hotspots” of infection in an environment with many complicating factors surrounding the reliability and accuracy of geographic and immunization coverage data, such as data that is spatially “jittered” to preserve anonymity (19, 33). Our work provides a much needed proof-of-concept, illustrating that fixing vaccination coverage and adjusting only the degree of clustering has large impacts on the risk and magnitude of outbreaks. Additionally, the literature on spatial heterogeneity in vaccination coverage is typically focused on patterns observed in vaccination coverage or serology data. Our use of simulation in an idealized environment allows for a better understanding of the implications of the types of clustering identified in these earlier analyses for outbreak risk.

This study has some limitations as well. We used a SIR model, which does not use an incubation period (which could be encoded using a susceptible-exposed-infected-recovered (SEIR) model with a compartment for latent infection) because the time dynamics of transmission were not a key focus of this paper, and both models will result in the same predictions of epidemic size. We also did not consider vaccine failure (i.e., assumed 100% vaccine effectiveness), and thus our results likely underestimate the number of cases that could occur in a worst-case scenario. Additionally, we used a deterministic transmission model to highlight the impact of clustering of nonvaccination and aggregation, yet the occurrence and size of outbreaks is in reality a function of both stochasticity in the population distribution of susceptibility—which we model explicitly—and demographic stochasticity in transmission dynamics, which our model omits. The

use of a deterministic model allowed us to focus specifically on the stochastic variation of the spatial distribution of nonvaccination, but our results should be interpreted in light of this choice. Finally, a square grid with fixed population size of 256,000 individuals is a stylized, simplified representation of a city and is not meant to directly represent the complexity of real-world contact networks, but instead seeks to capture a mix of local and nonlocal transmission. Making optimal use of these findings necessitates understanding how this heterogeneity impacts dynamics in the context of more heterogeneous and multilayered contact networks. Finally, the model’s dynamics are dependent upon our choice to analyze a population smaller than the critical population size of ~400,000 to 500,000, above which endemic circulation becomes possible. This allowed us to focus on the types of outbreak scenarios that are currently of the most pressing concern, but limits applications of this research to endemic transmission.

Conclusions

We show that the assumptions of spatially homogeneous vaccination coverage and contact result in an underestimation of the true number of individuals who need to be vaccinated to prevent outbreaks. Fine-scale clustering, as measured by high values of the isolation index, produced scenarios with the greatest outbreak potential. Since such fine-scale vaccination data are not broadly available in the United States, it is difficult to allocate resources, plan vaccination strategies, and respond to imported measles cases in a way that is responsive to this type of localized clustering. Especially given the ongoing pandemic, it is imperative to better understand and control the spread of preventable diseases such as measles—focusing on concrete ways to reduce case burden and health service utilization—as the coming school year is likely to see unprecedented challenges as COVID-19 cases grow and the fall influenza season approaches. The approach here is also likely to have important implications for managing COVID-19 therapeutic/vaccine distribution, as clustering of susceptibility and immunity are likely to occur in the communities both least and most hard hit in the first waves of transmission. As noted by Truelove et al., fine-scale clustering of the sort described here resulted in the largest increases in the critical vaccination fraction for diseases with lower values of R_0 . This suggests that issues around spatial clustering of susceptibility to COVID-19, which has an R_0 roughly four times lower than measles, may be as or more acute as in the scenarios described here (19). This research thus motivates the need not only for increased vaccination coverage, but also for the collection of finer-scale vaccination data to create “susceptibility maps” that can guide policy-makers and health practitioners to preferentially direct resources to those areas at highest risk of outbreaks.

Data Availability. All model code and data generated for this analysis are available for download on the GitHub Repository: <https://github.com/epibayes/Measles-Spatial-Clustering-and-Aggregation-Effects/> (32).

ACKNOWLEDGMENTS. J.Z. was supported by a Catalyst award from the Michigan Institute of Computational Discovery and Engineering and a Michigan Institute for Clinical and Health Research Pathway Award. N.B.M.’s PhD research is funded by M.L.B.

1. World Health Organization (WHO), Measles and rubella surveillance data (2020). https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/. Accessed 11 August 2020.
2. World Health Organization (WHO), Immunization, Vaccines, and Biologicals: New measles surveillance data for 2019 (2019). <https://www.who.int/immunization/newsroom/measles-data-2019/en/>. Accessed 11 August 2020.
3. World Health Organization (2020), Global measles and rubella monthly update: March, 2020. World Health Organization, Immunization, Vaccines and Biologicals, Geneva. https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/. Accessed 11 August 2020.

4. R. Silverberg, J. Caceres, S. Greene, M. Hart, C. H. Hennekens, Lack of measles vaccination of a few portends future epidemics and vaccination of many. *Am. J. Med.* **132**, 1005–1006 (2019).
5. A. D. Pananos et al., Critical dynamics in population vaccinating behavior. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13762–13767 (2017).
6. H. A. Hill, L. D. Elam-Evans, D. Yankey, J. A. Singleton, V. Dietz, Vaccination coverage among children aged 19–35 Months: United States, 2015. *MMWR Morb. Mortal. Wkly. Rep.* **65**, 1065–1071 (2016).
7. M. S. Majumder, E. L. Cohn, S. R. Mekaru, J. E. Huston, J. S. Brownstein, Substandard vaccination compliance and the 2015 measles outbreak. *JAMA Pediatr.* **169**, 494–495 (2015).
8. S. C. Pingali et al., Associations of statewide legislative and administrative interventions with vaccination status among kindergartners in California. *JAMA* **322**, 49–56 (2019).

9. S. B. Omer *et al.*, Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. *Am. J. Epidemiol.* **168**, 1389–1396 (2008).
10. J. Matthias *et al.*; Centers for Disease Control and Prevention, Notes from the field: Outbreak of pertussis in a school and religious community averse to health care and vaccinations—Columbia County, Florida, 2013. *MMWR Morb. Mortal. Wkly. Rep.* **63**, 655 (2014).
11. S. G. Robison, J. Liko, The timing of pertussis cases in unvaccinated children in an outbreak year: Oregon 2012. *J. Pediatr.* **183**, 159–163 (2017).
12. Centers for Disease Control and Prevention, “US measles cases in first five months of 2019 surpass total cases per year for past 25 years” (Press Release, May 30, 2019).
13. W. J. Moss, D. E. Griffin, Global measles elimination. *Nat. Rev. Microbiol.* **4**, 900–908 (2006).
14. World Health Organization (WHO), “Global measles and rubella strategic plan: 2012–2020” in *Tech. Report No. ISBN 978 92 4 150339 6*, (World Health Organization, Geneva, Switzerland, 2012).
15. P. E. M. Fine, J. A. Clarkson, Measles in England and Wales—I: An analysis of factors underlying seasonal patterns. *Int. J. Epidemiol.* **11**, 5–14 (1982).
16. F. M. Guerra *et al.*, The basic reproduction number (R_0) of measles: A systematic review. *Lancet Infect. Dis.* **17**, e420–e428 (2017).
17. Centers for Disease Control and Prevention, Transmission of measles (2018). <https://www.cdc.gov/measles/transmission.html>. Accessed 11 August 2020.
18. Centers for Disease Control and Prevention (CDC), “Measles” in *Epidemiology and Prevention of Vaccine-Preventable Diseases*, J. Hamborsky, A. Kroger, C. Wolfe, Eds. (Public Health Foundation, 13th edition, 2015), pp. 209–229.
19. S. A. Truelove *et al.*, Characterizing the impact of spatial clustering of susceptibility for measles elimination. *Vaccine* **37**, 732–741 (2019).
20. P. A. Gastañaduy *et al.*, A measles outbreak in an underimmunized Amish community in Ohio. *N. Engl. J. Med.* **375**, 1343–1354 (2016).
21. P. L. Delamater, T. F. Leslie, Y. T. Yang, A spatiotemporal analysis of non-medical exemptions from vaccination: California schools before and after SB277. *Soc. Sci. Med.* **168**, 230–238 (2016).
22. P. L. Delamater, T. F. Leslie, Y. T. Yang, Examining the spatiotemporal evolution of vaccine refusal: Nonmedical exemptions from vaccination in California, 2000–2013. *BMC Public Health* **18**, 458 (2018).
23. J. K. Olive, P. J. Hotez, A. Damania, M. S. Nolan, The state of the antivaccine movement in the United States: A focused examination of nonmedical exemptions in states and counties. *PLoS Med.* **15**, e1002578 (2018).
24. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
25. D. E. Sugerman *et al.*, Measles outbreak in a highly vaccinated population, San Diego, 2008: Role of the intentionally undervaccinated. *Pediatrics* **125**, 747–755 (2010).
26. M. Salathé, S. Bonhoeffer, The effect of opinion clustering on disease outbreaks. *J. R. Soc. Interface* **5**, 1505–1508 (2008).
27. J. M. Santoli *et al.*, Effects of the COVID-19 pandemic on routine pediatric vaccine ordering and administration: United States, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 591–593 (2020).
28. C. A. Bramer *et al.*, Decline in child vaccination coverage during the COVID-19 pandemic: Michigan care improvement registry, May 2016–May 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 630–631 (2020).
29. P. A. P. Moran, Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
30. J. Iceland, D. H. Weinberg, E. Steinmetz, U.S. Census Bureau, Series CENSR-3, “Racial and Ethnic Residential Segregation in the United States: 1980–2000. Appendix B. Measures of Residential Segregation” (2002) U.S. Government Printing Office, Washington, DC. <https://www.census.gov/library/publications/2002/dec/censr-3.html>. Accessed 11 August 2020.
31. A. Gelman, J. Carlin, Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspect. Psychol. Sci.* **9**, 641–651 (2014).
32. N. Masters, Measles-Spatial-Clustering-and-Aggregation-Effects (Version v1.0.0). GitHub. <http://doi.org/10.5281/zenodo.4050445>.
33. T. K. Brownwright, Z. M. Dodson, W. G. van Panhuis, Spatial clustering of measles vaccination coverage among children in sub-Saharan Africa. *BMC Public Health* **17**, 957 (2017).
34. A. J. Tatem, Innovation to impact in spatial epidemiology. *BMC Med.* **16**, 209 (2018).
35. T. F. Leslie, E. J. Street, P. L. Delamater, Y. T. Yang, K. H. Jacobsen, Variation in vaccination data available at school entry across the United States. *Am. J. Public Health* **106**, 2180–2182 (2016).
36. H. A. Hill, L. D. Elam-Evans, D. Yankey, J. A. Singleton, M. Kolasa, National, state, and selected local area vaccination coverage among children aged 19–35 months: United States, 2014. *MMWR Morb. Mortal. Wkly. Rep.* **64**, 889–896 (2015).
37. B. Sartorius *et al.*, Identifying high-risk areas for sporadic measles outbreaks: Lessons from South Africa. *Bull. World Health Organ.* **91**, 174–183 (2013).
38. D. Ntirampeba, I. Neema, L. Kazembe, Modelling spatio-temporal patterns of disease for spatially misaligned data: An application on measles incidence data in Namibia from 2005–2014. *PLoS One* **13**, e0201700 (2018).
39. A. Kundrick *et al.*, Sub-national variation in measles vaccine coverage and outbreak risk: A case study from a 2010 outbreak in Malawi. *BMC Public Health* **18**, 741–751 (2018).
40. A. Wesolowski *et al.*, Measles outbreak risk in Pakistan: Exploring the potential of combining vaccination coverage and incidence data with novel data-streams to strengthen control. *Epidemiol. Infect.* **146**, 1575–1583 (2018).
41. N. C. Lo, P. J. Hotez, Public health and economic consequences of vaccine hesitancy for measles in the United States. *JAMA Pediatr.* **171**, 887–892 (2017).
42. B. T. Grenfell, O. N. Bjørnstad, J. Kappey, Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
43. J. Lessler, H. Salje, M. K. Grabowski, D. A. T. Cummings, Measuring spatial dependence for infectious disease epidemiology. *PLoS One* **11**, e0155249 (2016).
44. N. Bharti *et al.*, Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334**, 1424–1427 (2011).
45. C. E. Utazi *et al.*, High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine* **36**, 1583–1591 (2018).