



# Characterizing the impact of spatial clustering of susceptibility for measles elimination



Shaun A. Truelove<sup>a,\*</sup>, Matthew Graham<sup>a,b</sup>, William J. Moss<sup>a</sup>, C. Jessica E. Metcalf<sup>c,d</sup>,  
Matthew J. Ferrari<sup>e,f</sup>, Justin Lessler<sup>a</sup>

<sup>a</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

<sup>b</sup> The Hospital for Tropical Diseases, Wellcome Trust Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam

<sup>c</sup> Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

<sup>d</sup> Office of Population Research, Woodrow Wilson School, Princeton University, Princeton, NJ, USA

<sup>e</sup> Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA

<sup>f</sup> Department of Biology, The Pennsylvania State University, University Park, PA, USA

## ARTICLE INFO

### Article history:

Received 12 June 2018

Received in revised form 7 December 2018

Accepted 11 December 2018

Available online 19 December 2018

### Keywords:

Heterogeneity

Clustering

Measles

Effective reproductive number

Critical vaccination threshold

Critical immunity threshold

Disease transmission

Outbreak risk

Vaccination

Stochastic individual-based model

## ABSTRACT

Measles elimination efforts are primarily focused on achieving and maintaining national vaccination coverage goals, based on estimates of the critical vaccination threshold ( $V_c$ ): the proportion of the population that must be immune to prevent sustained epidemics. Traditionally,  $V_c$  estimates assume evenly mixing populations, an invalid assumption. If susceptible individuals preferentially contact one another, communities may remain vulnerable to epidemics even when vaccination coverage targets are met at the national level. Here we present a simple method to estimate  $V_c$  and the effective reproductive number,  $R$ , while accounting for spatial clustering of susceptibility. For measles, assuming  $R_0 = 15$  and 95% population immunity, adjustment for high clustering of susceptibility increases  $R$  from 0.75 to 1.29,  $V_c$  from 93% to 96%, and outbreak probability after a single introduction from <1% to 23%. The impact of clustering remains minimal until vaccination coverage nears elimination levels. We illustrate our approach using Demographic and Health Survey data from Tanzania and show how non-vaccination clustering potentially contributed to continued endemic transmission of measles virus during the last two decades. Our approach demonstrates why high national vaccination coverage sometimes fails to achieve measles elimination, and that a shift from national to subnational focus is needed as countries approach elimination.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Despite significant improvements in global measles vaccination coverage (from 73% to 85% for a single dose between 2000 and 2015), large outbreaks continue to occur, making declared goals of measles elimination by 2020 across World Health Organization (WHO) regions realistically unattainable [1,2]. These outbreaks have occurred both in countries with low measles vaccination coverage and those with perceived high levels of measles control throughout Europe (e.g., France, Romania), Africa (e.g., Malawi), and Asia (e.g., the Philippines) [3]. The continued occurrence of outbreaks reflects the considerable logistical, political, and social challenges to measles elimination, highlighting the need for novel approaches [4,5].

WHO's vaccination coverage goals of 95% for the first and second doses of measles-containing vaccine (MCV) in the Global Measles and Rubella Strategic Plan derive, in part, from theoretical estimates of the immunity level needed to prevent ongoing measles virus transmission [1,6]. These estimates are based on the basic reproductive number ( $R_0$ ), the average number of people a single infected person is expected to infect in a fully susceptible population (usually assumed to be 12 to 18 for measles but varies substantially) [7–9]. Theory states that if a country can successfully immunize, via vaccination, a proportion of their population equal to  $1 - (1/R_0)$ , the population will have sufficient herd immunity to avoid significant epidemics if introduction of the pathogen occurs [7,10]. Hence, for measles, 92–94% of the population would need to be immune to achieve this critical vaccination threshold ( $V_c$ ).

Though useful, this rule of thumb is based on simple epidemic models that ignore population structure, assuming each infected

\* Corresponding author.

E-mail address: [shauntruelove@jhu.edu](mailto:shauntruelove@jhu.edu) (S.A. Truelove).

individual is equally likely to have a potentially infectious contact with every other individual in the population. If this assumption is violated, then predictions from these models may be invalid. Specifically, in a setting where susceptible individuals are more likely to be in contact with other susceptible individuals than would be expected by random chance, an introduced pathogen has a higher probability of producing uninterrupted chains of transmission than in a setting where contact is homogeneous. If vaccination coverage is geographically or socially heterogeneous, the vaccination process itself can lead to violation of this critical assumption [11–13].

The existence of heterogeneities in vaccination coverage and their association with disease incidence has been repeatedly demonstrated. Within African countries, heterogeneity in measles vaccination coverage is prevalent, with clusters of low vaccination common even in countries with high overall coverage [14]. In the United States, spatial clustering of nonmedical exemptions to school immunization requirements in Michigan and California were associated with pertussis incidence [15,16]. In 2014, an outbreak in Ohio of 380 confirmed measles cases occurred in a large Amish community [17]. The spatial clustering of this population, where vaccination coverage may be as low as 8%, resulted in the largest measles outbreak in recent U.S. history, despite high measles vaccination coverage for Ohio overall (>92% among kindergarteners) [18].

The impact of these heterogeneities on disease transmission is well supported by theoretical work. Recent work by Funk et al. demonstrated the increased importance of immunity in specific age groups, particularly 5–9 year-olds, due to age-specific heterogeneity in contact patterns [19]. Other studies demonstrated that assuming homogeneous contact in transmission networks can underestimate  $R_0$  [10,20,21]. Using network theory, Salathe and Bonhoeffer found that when heterogeneity in contact between susceptibles exists, a higher level of vaccination coverage is needed to eliminate disease transmission than estimated from homogeneous networks [21].

Simple, broadly applicable, evidence-based methods that provide guidance where heterogeneity in vaccine use exists may be useful in planning and evaluating measles control activities. Here we present a method for estimating the effective reproductive number and critical vaccination threshold in the presence of spatial heterogeneity in susceptibility. Using Demographic and Health Survey (DHS) data from Tanzania, we demonstrate its application and examine how spatial heterogeneity in vaccination coverage may have contributed to continued measles virus transmission while vaccination coverage, measured at the national level, was presumed to be high enough to achieve elimination.

**2. Methods**

*2.1. Approach*

The basic and effective reproductive numbers,  $R_0$  and  $R$ , are key to estimating disease outbreak potential and the vaccination coverage needed to interrupt transmission.  $R$  is how many people an average infectious individual is expected to infect in a population with some immunity. If everyone is equally likely to interact with everyone else, the probability that a contact is susceptible equals the proportion susceptible in the population. When immunity is exclusively vaccination-derived, this proportion susceptible is  $1 - v$ , where  $v$  is the proportion who successfully develop protective immunity following vaccination. Hence:

$$R = R_0 \times \text{Pr}(\text{contact is susceptible}) = R_0(1 - v) \tag{1}$$

Here we focus on maintaining measles elimination through successful vaccination. For simplicity, we ignore the impact of immunity acquired through infection, assuming application to populations where endemic measles has been eliminated (i.e., where we assume susceptibility is synonymous with not being successfully vaccinated).

In many populations spatial heterogeneity in vaccination arises from access or refusal, and individuals are, in general, more likely to contact people living near their home than those living farther away [22–25]. The more spatially clustered non-vaccination is, the higher the probability that contacts of a susceptible individual are also susceptible, and the higher the probability of onward transmission if a pathogen is introduced.

We can account for this heterogeneity in Eq. (1) by adjusting the probability that a contact is susceptible. We quantify the tendency of non-vaccination to cluster using the statistic  $\tau(r)$ , defined as the relative probability that an individual living at spatial distance  $r$  from an unvaccinated individual is also unvaccinated versus the probability that anyone in the population is unvaccinated [26,27]. Hence:

$$\begin{aligned} &\text{Pr}(\text{contact is susceptible} | \text{contact lives at distance } r) \\ &= (1 - v) \frac{1 - v(r)}{1 - v(\infty)} = (1 - v)\tau(r) \end{aligned} \tag{2}$$

where  $1 - v(\infty)$  is equivalent to  $1 - v$  for the full population. If we know  $g(r)$ , the probability that a contact occurs at distance  $r$  given that a contact occurred, we can calculate the number of infectious contacts that a susceptible individual is expected to make (i.e., the adjusted effective reproductive number,  $R^*$ ):

$$\begin{aligned} R^* &= \beta c(1 - v) \int_0^\infty g(r)\tau(r)dr = R_0(1 - v) \int_0^\infty g(r)\tau(r)dr \\ &= R_0(1 - v)\phi \end{aligned} \tag{3}$$

where  $c$  is an individual's expected number of potentially infectious contacts over the course of their infectious period,  $\beta$  is the per contact transmission probability (note that  $R_0 = \beta c$ ), and  $\phi$  is a clustering adjustment factor for  $R$  (i.e.,  $R^* = R\phi$ ).

The critical vaccination threshold,  $V_c$ , is calculated as the vaccination level where  $R < 1$ , or  $V_c = 1 - \frac{1}{R_0}$ . Assuming the susceptibility clustering and the distribution of contacts remain stable, we can account for clustering in an adjusted critical vaccination threshold,  $V_c^*$ , by including  $\phi$ :

$$V_c^* = 1 - \frac{1}{R_0\phi} \tag{4}$$

**3. Statistical methods**

For these analyses, we assumed that the relative probability that an individual at distance  $r$  from an unvaccinated (i.e., susceptible) individual is also unvaccinated follows an exponential decay function, of the form  $\tau(r) = \theta e^{-\lambda r} + 1$ , where  $\theta$  is the maximum relative probability of a contact being unvaccinated and  $\lambda$  is the rate of decay towards 1. In simulations, we assume successful vaccination to be the only source of immunity in the population. The distance from the home at which contacts occur is assumed to follow a gamma distribution,  $g(r)$  gamma( $r, \alpha, \beta$ ), where  $r$  is the distance of the contact. We note that the method could be applied using other parametric and empirical forms of  $\tau(r)$  and  $g(r)$ .

#### 4. Analytic examples and simulation studies

To illustrate our approach, we considered populations with four levels of spatial clustering of susceptibility: *none* ( $\tau(r)$  parameters:  $\theta = 0$ ,  $\lambda = 0.5$ ), *low* ( $\theta = 0.25$ ,  $\lambda = 0.5$ ), *moderate* ( $\theta = 0.5$ ,  $\lambda = 0.5$ ), and *high* ( $\theta = 1.0$ ,  $\lambda = 0.5$ ). We examined contact distributions fit to empirical data from three distinct populations representing different contact patterns:  $g_A(r)$ , fit to data from urban and rural locations in southern China ( $\alpha = 0.238$ ,  $\beta = 0.162$ );  $g_B(r)$ , fit to data from rural Zambia ( $\alpha = 0.086$ ,  $\beta = 0.042$ ); and  $g_C(r)$ , fit to cell phone data from the U.S ( $\alpha = 0.701$ ,  $\beta = 0.076$ ) (see [Supplemental Figure S1](#) and [Text S4](#) for full details) [28–30]. Contacts are most local in  $g_A(r)$ , while they are most dispersed in  $g_C(r)$ , reflecting US commuting culture. We considered three levels of successful vaccination coverage, 85, 90, and 95%, and assumed  $R_0 = 15$  (similar to measles).

To validate analytic calculations, we performed spatially explicit simulation studies of these scenarios using stochastically generated populations of 100,000 individuals for each level of vaccination coverage and clustering. We performed sensitivity analyses of the impact of population size and density ([Fig. 1](#), [Appendix S7](#)). Assuming a single introduction of a measles-like virus ( $R_0 = 15$ , serial interval of 2 weeks), we performed 10,000 simulations for each scenario.  $R^*$  was calculated analytically for each scenario using the pre-defined target  $\tau(r)$  and the exact  $g(r)$  from each simulation. For each scenario we compared the behavior (outbreak probability and final size distribution) of the spatially explicit model with that of a homogenous mixing model where  $R_0$  was equal to the corresponding  $R^*$ . An outbreak was defined as  $\geq 5\%$  of susceptible individuals becoming infected.

#### 5. Example application

We illustrate our approach using data from Demographic and Health Surveys (DHS) conducted in Tanzania. DHS are cluster randomized surveys that include reported and documented

individual-based vaccination histories, and, in the most recent surveys, GPS coordinates for sampling cluster locations [31]. These data provide a snapshot of vaccination coverage achieved through routine and supplemental immunization activities among children under 5 years. We use Tanzania to demonstrate the application of our approach because the country had multiple years of DHS surveys with spatial data and a recent history of increasing measles vaccination coverage. See [Supplemental Text S9](#) for full methods.

All analyses were performed using R 3.4.1 and RStudio 1.0.153.

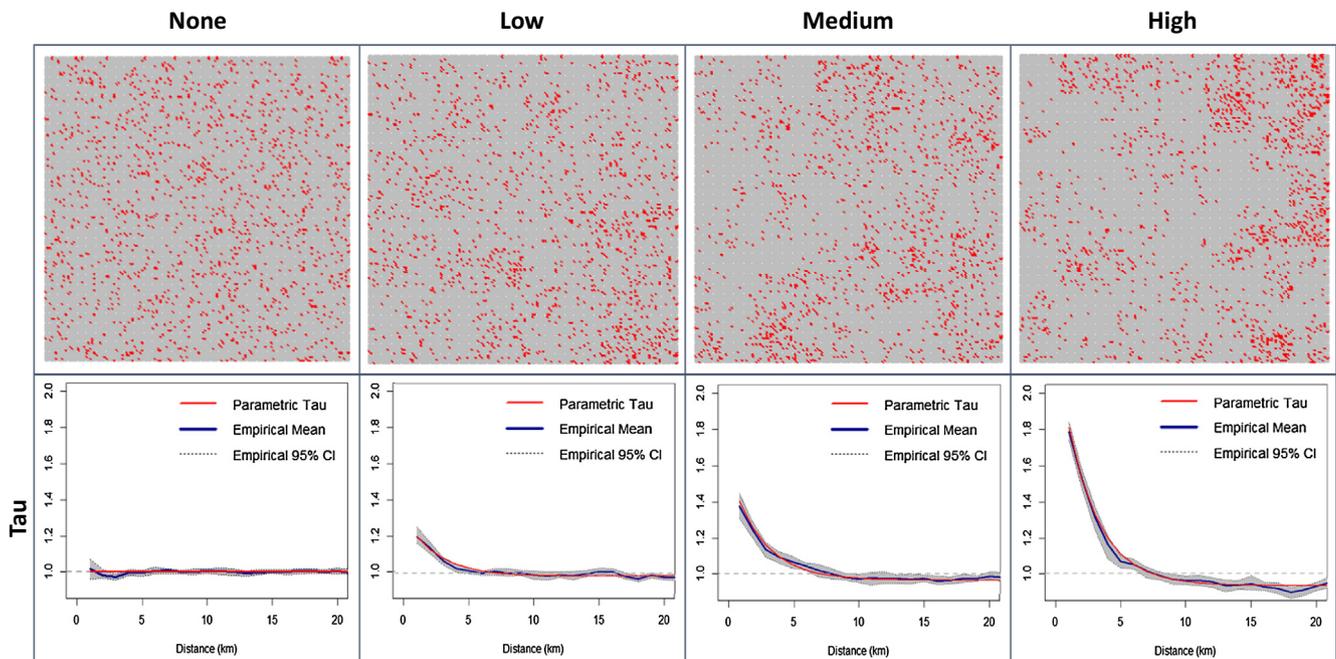
#### 6. Results

##### 6.1. Analytic Analysis

Analytic results show that clustering of non-vaccination can have a significant impact on both  $R$  and  $V_c$  ([Table 1](#) and [S1](#)). Under contact distribution  $g_A(r)$  and *high* clustering,  $R^*$  is nearly double that assumed by standard estimates that assume homogenous mixing ( $\phi = 1.71$ ), and  $R^* > 1$  at 95% successful vaccination coverage (i.e., herd immunity is not achieved, [Table 1](#)). These effects translate into significant increases in  $V_c^*$ , from 93% to 96% ([Table 1](#)).

The absolute increase in  $V_c^*$  due to clustering is inversely related to  $R_0$  ([Fig. 2](#)). For diseases with high  $R_0$  like measles ( $R_0 = 15$ ), the required  $V_c^*$  with *high* clustering is 3% higher than without ([Table 1](#)). Diseases with potentially lower  $R_0$  like mumps ( $R_0 = 7.7$ ), polio ( $R_0 = 6$ ), rubella ( $R_0 = 6$ ) or cholera ( $R_0 = 2.6$ ) have larger, yet possibly more manageable, increases in  $V_c^*$ , of 6% (87% versus 93%), 8% (83% versus 91%), and 19% (62% versus 81%) [32,10,33–35].

The shapes of both the contact distance and clustering distributions affect the resulting estimates of  $R^*$ ,  $V_c^*$ , and  $\phi$ . Of the three contact distributions applied,  $g_B(r)$  produced the highest  $R^*$  values in the presence of clustering due to having the highest proportion of contacts within short distances; the more dispersed  $g_C(r)$  produced the lowest  $R^*$  ([Table 1](#)). Curves of  $1 - G(r)$  (where  $G(r)$  is the cumulative distribution function of  $g(r)$ ) versus  $\tau(r)$  demonstrate the effect of overlap between  $\tau(r)$  and  $g(r)$  ([Fig. 3](#)). The



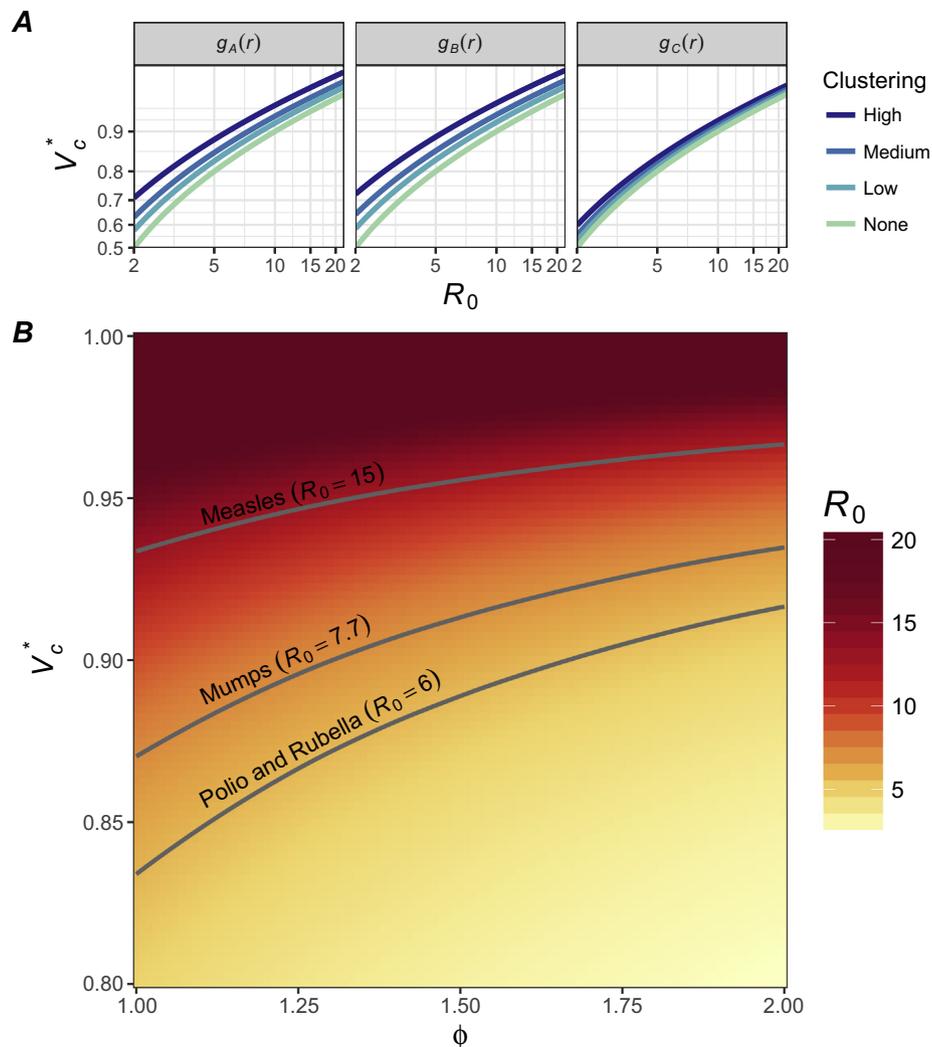
**Fig. 1.** Synthetic populations and corresponding  $\tau(r)$  functions with *no*, *low*, *medium*, and *high* spatial clustering of susceptibility. Individuals are spatially distributed evenly across the space, and susceptibility increasingly clustered (grey = immune, red = susceptible).

**Table 1**

Analytic  $R^*$  and  $V_c^*$  estimates from Eq. (1) assuming  $R_0 = 15$  and using contact distributions derived from Read et al. 2014 ( $g_A(r)$ ), Searle et al. 2017 ( $g_B(r)$ ), and Noulas et al. 2012 ( $g_C(r)$ ) [28–30].

Vaccination coverage	Clustering	$g_A(r)$		$g_B(r)$		$g_C(r)$	
		$R^*$	$V_c^*$	$R^*$	$V_c^*$	$R^*$	$V_c^*$
95%	None	0.75	93.3%	0.75	93.3%	0.75	93.3%
	Low	0.88	94.3%	0.90	94.4%	0.80	93.7%
	Medium	1.02	95.1%	1.05	95.2%	0.84	94.0%
	High	1.29	96.1%	1.35	96.3%	0.93	94.6%

† Assumes  $R_0 = 15$  and  $\lambda = 0.5$ .



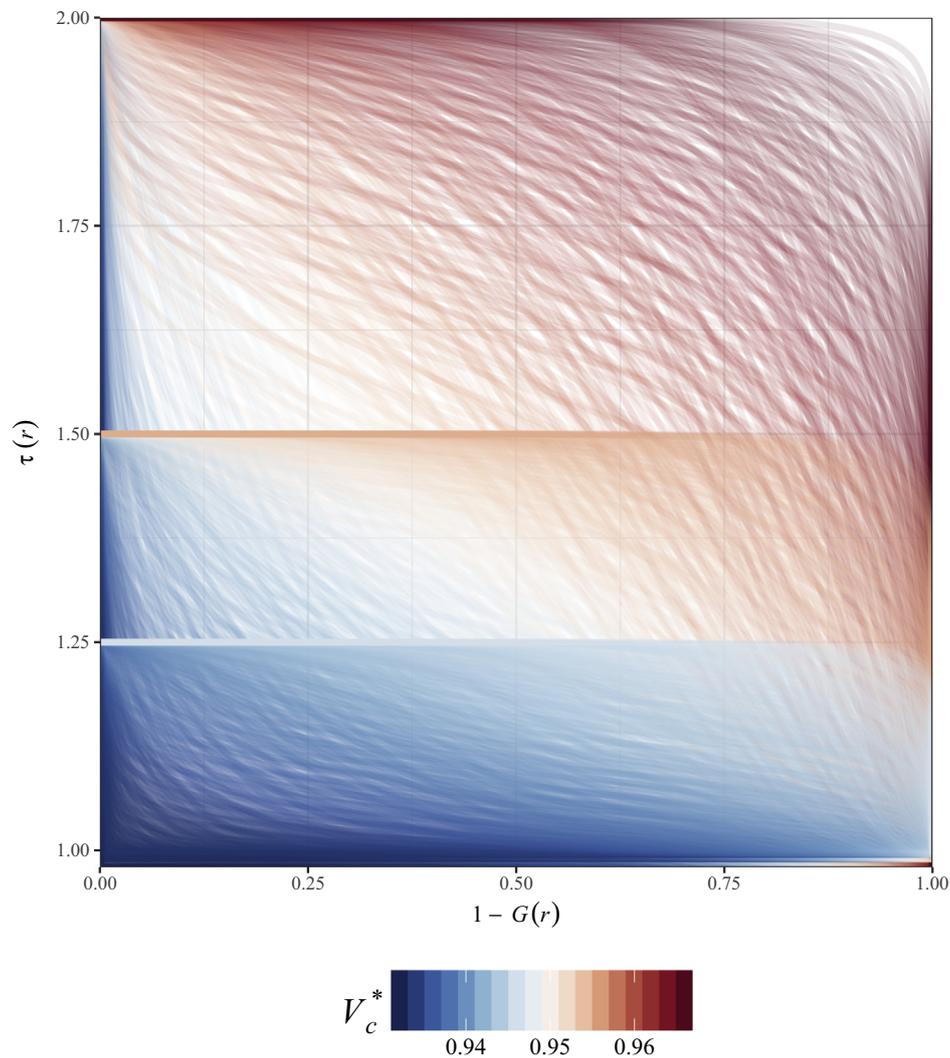
**Fig. 2.** The association between  $R_0$ ,  $V_c^*$ , and spatial clustering of non-vaccination ( $\phi$ ). As clustering increases, the required vaccination coverage to maintain  $R = 1$  increases. At low  $R_0$ , this increase due to clustering is much greater than at high  $R_0$ . (A) The association between  $R_0$  and  $V_c^*$  at the four defined levels of clustering, using the three  $g(r)$  distributions. The relative increase in  $V_c^*$  with increased clustering is relatively equivalent for  $g_A(r)$  and  $g_B(r)$ , but lower for  $g_C(r)$ . (B) The relationship between  $V_c^*$ ,  $\phi$ , and  $R_0$ , where  $\phi$  represents the relative impact of spatial clustering on  $R^*$ . This relationship is highlighted for measles ( $R_0 = 15$ ), mumps ( $R_0 = 7.7$ ), and polio and rubella ( $R_0 = 6$ ) [10,32,34,33]. For diseases like polio, with lower  $R_0$ , there is a substantially greater increase in absolute  $V_c^*$  with increasing  $\phi$ , as compared to diseases with higher  $R_0$ .

impact of clustering is maximized with perfect overlap between the  $g(r)$  probability and maximum  $\tau(r)$  value, occurring when 100% of contact between individuals occurs within the distance where  $\tau(r)$  is highest, in this case at  $r = 0$ . The impact of clustering is reduced when clustering and contact occur at different distances. Theoretically, if the shapes of the  $\tau(r)$  and  $g(r)$  functions were mismatched, we could see  $V_c^* < V_c$ . However, for this study we constrained these functions to monotonically decreasing distributions, thus  $V_c^*$  approaches  $V_c$  as  $g(r)$  approaches a uniform

distribution (i.e. equal contact distance probability, or homogeneous contact).

## 7. Simulation studies

We found near perfect correlation between the expected outbreak probabilities from homogeneous mixing SIR simulations and spatial mixing simulations (Pearson's  $\rho = 0.98$ ) and high



**Fig. 3.** Critical vaccination threshold, as defined by the overlapping densities of the  $\tau(r)$  and  $g(r)$  distributions. Plotted curves are  $\tau(r)$  where  $r$  is that which corresponds to the value of  $1 - G(r)$ , and color corresponds to the resulting  $V_c^*$  when  $R_0 = 15$ . Three levels of clustering ( $\theta = 0.25, 0.50, 1.0$ ) and 529  $g(r)$  distributions, with both  $\alpha$  and  $\beta$  range = 0.1–10.

**Table 2**

Outbreak probability from simulations with SIR and individual-based spatially explicit model, assuming  $\lambda = 0.5$  and  $g(r)$ .

Vaccination coverage	Clustering	$R^*$ ‡	Probability of outbreak <sup>†</sup>	
			Spatial simulations, mean (95% CI)	SIR simulations, mean (95% CI)
95%	None	0.75	0 (0–0.01)	0 (0–0)
	Low	0.88	0.01 (0.00–0.02)	0.01 (0.01–0.01)
	Medium	1.02	0.07 (0.06–0.08)	0.07 (0.07–0.08)
	High	1.24	0.23 (0.22–0.24)	0.23 (0.22–0.24)

<sup>†</sup> Outbreak is defined by  $\geq 5\%$  of the susceptible individuals becoming infected. Assumes introduction happens at equal rates randomly among susceptibles.

<sup>‡</sup>  $R^*$  calculated analytically using the empirical  $g(r)$  from spatial simulations and defined  $\tau(r)$  distributions.

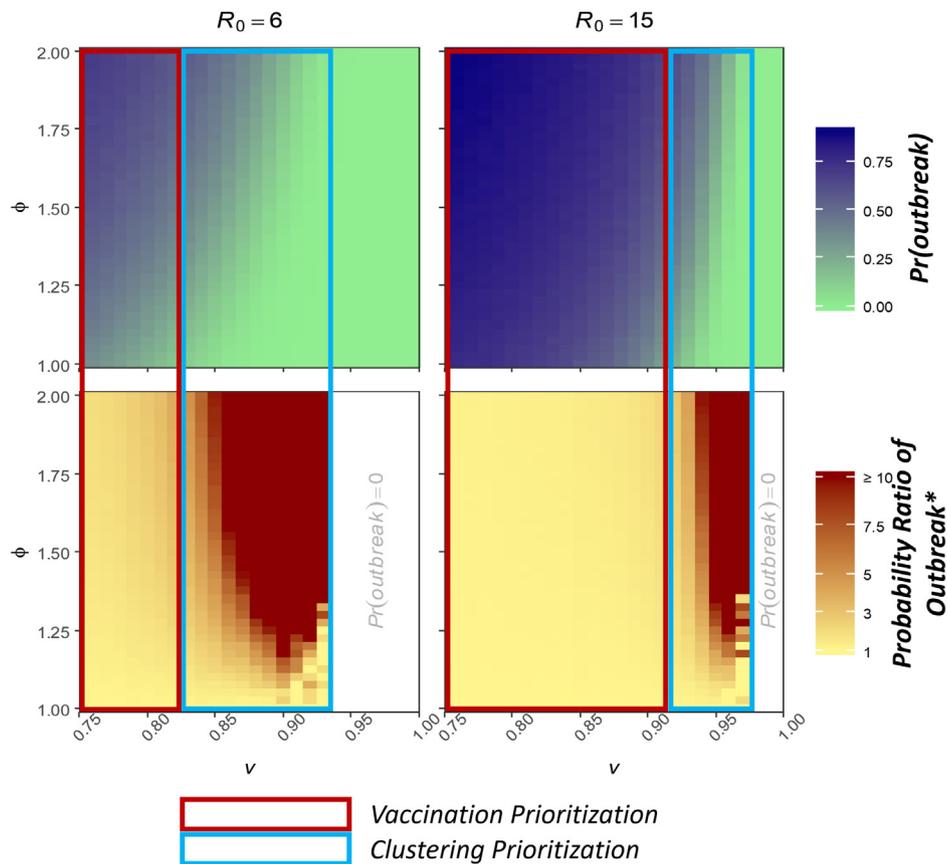
comparability between these estimates (Pearson's  $\chi^2$   $p = 0.24$ ) (Table 2). However, the final sizes of these epidemics differed significantly (51% [95% CI: 42–60%] vs 82% [95% CI: 80–84%] at  $R^* = 2.0$ ), due to previously described differences between spatial and homogenous mixing populations [36].

The relative impact of non-vaccination clustering on outbreak probability increases exponentially as vaccination coverage approaches the critical vaccination threshold (i.e., near elimination; Fig. 4). For measles, the ratio of the outbreak probability assuming high clustering to the outbreak probability assuming no clustering (hereafter PrR), with 85% immunity was 1.4 (95%

CI: 1.3–1.5) and with 90% immunity PrR = 1.9 (95% CI: 1.7–2.1), increasing rapidly to PrR > 20 with 95% immunity (Pr(outbreak) = 23% [95% CI: 22–24%] vs. 0% [95% CI: 0–0.1%]) (Table 3 and S3).

## 8. Application

We illustrate our approach using DHS data from Tanzania, which has a recent history of largely uninterrupted measles virus transmission concurrent with increasing vaccination coverage. Visual examination of cluster and district-level vaccination coverage shows apparent clustering of non-vaccination during both the



**Fig. 4.** Probability of an outbreak and the outbreak probability ratio given adjustment for clustering ( $\phi$ ) compared with homogeneous ( $\phi = 1$ ), for  $R_0 = 6$  and  $R_0 = 15$ . The probability ratio compares the clustering-adjusted estimates of outbreak probability for each level of clustering to the homogeneous estimates at each level of successful vaccination coverage. The probability of an outbreak decreases with increasing successful vaccination proportion ( $v$ ) and increases with increasing clustering of susceptibility ( $\phi$ ). Until vaccination becomes high enough, the contribution of clustering is negligible and public health focus should be on increasing vaccination coverage, as highlighted by the areas in the red boxes. As successful vaccination coverage approaches elimination levels (estimated at 83% for  $R_0 = 6$  and 93% for  $R_0 = 15$  with the traditional critical immunization threshold equation) the relative effect of clustering on the outbreak probability increases exponentially. When vaccination is sufficiently high, prioritization should shift from increasing national vaccination coverage to the clustering of susceptibility (blue box).

**Table 3**  
Incidence and estimated effective reproductive numbers for Tanzania from 1998 to 2014.

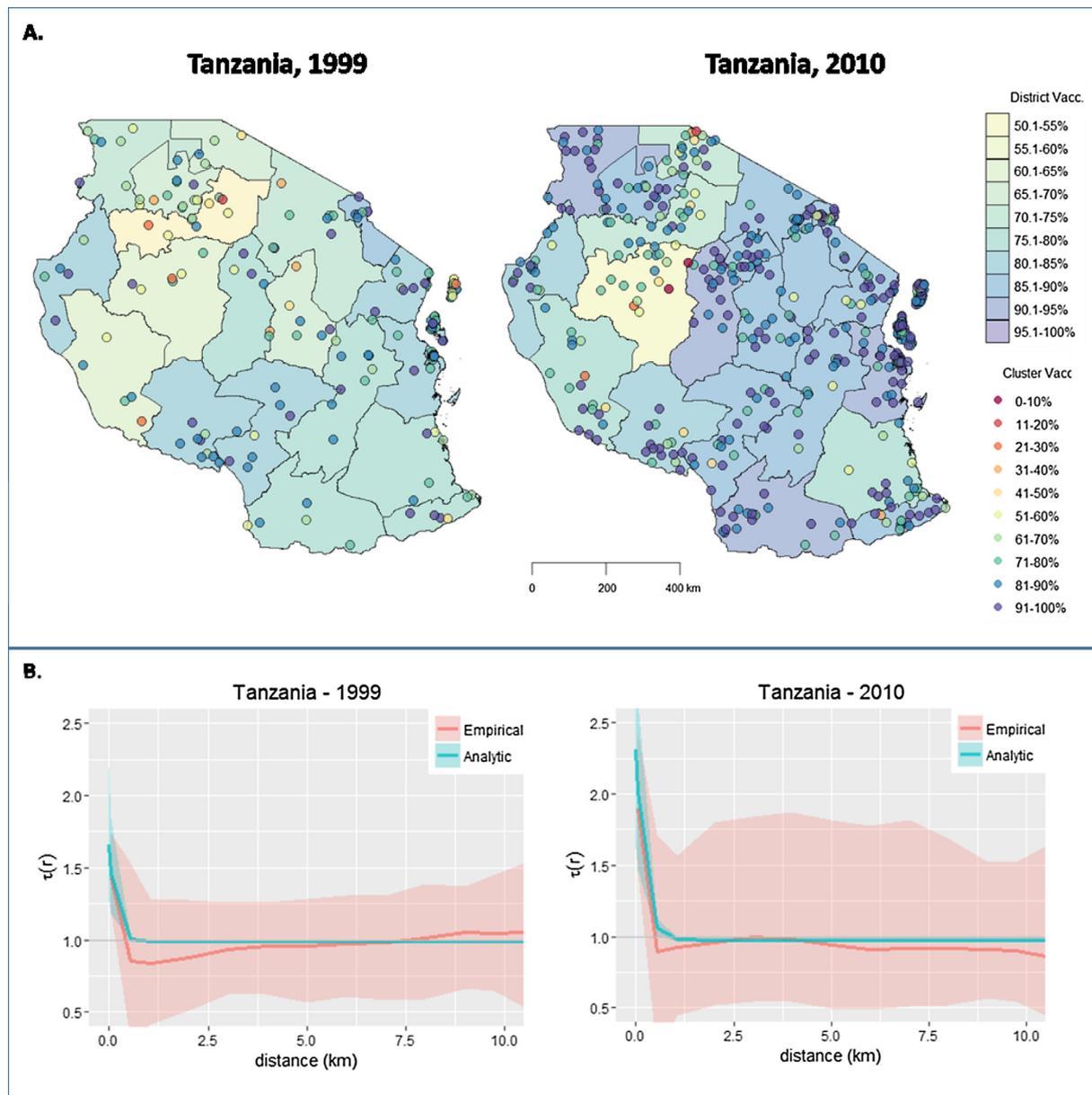
Year	5-yr routine vaccination	5-yr incidence (per 100,000)	R, 5-year mean		$V_c$	
			Unclustered	Clustered <sup>†</sup>	Unclustered	Clustered <sup>†</sup>
1999	76.2%	29.78	3.57	4.5 (3.9–5.3)	93.3%	94.7% (93.9–95.5%)
2010	92.2%	3.75	1.17	1.8 (1.5–2.3)	93.3%	95.7% (94.7–96.5%)

<sup>†</sup> Estimate mean and 95% CI.

1999 and 2010 Tanzania DHS surveys, with a possible increase in clustering in 2010 compared to 1999 (Fig. 5a). Fitting  $\tau(r)$  to DHS data confirms this observation, with significant clustering of non-vaccination in Tanzania during the 1999 and 2010 surveys, with  $\phi = 1.26$  (95% CI: 1.1–1.5) and 1.56 (95% CI: 1.2–1.9) (Fig. 5b). Assuming stable clustering and applying these  $\phi$ s to 5-year means of UNICEF/WHO vaccination coverage,  $R$  was likely substantially higher than implied by observed coverage in Tanzania. In 1999, with  $v = 76\%$  and assuming  $R_0 = 15$ ,  $R^* = 4.50$  (95% CI: 3.90–5.30) versus  $R = 3.57$ ; in 2010, with  $v = 92\%$ ,  $R^* = 1.83$  (95% CI: 1.46–2.25) versus  $R = 1.17$  (Table 3) [37]. Starting in 2012,  $R^* < 1$  and vaccination  $> V_c^*$ , assuming accurate vaccination coverages, and this was followed by measles incidence dropping and remaining below 1 per 100,000 (Table S2). Without reductions in clustering of non-vaccination, Tanzania will likely need to maintain successful vaccination coverage of at least 95–96% to maintain interrupted measles virus transmission as naturally-acquired immunity wanes.

### 9. Discussion

We demonstrate that the assumption of homogeneity under which standard methods function likely results in the underestimation of the effective reproductive number and critical immunity threshold, especially where high clustering of susceptibility exists. This can lead to situations where infectious diseases continue to spread despite perceived control and elimination. The effective reproductive number increases with increasing susceptibility clustering and the impact of this clustering on the relative probability of an outbreak becomes exponentially higher as countries approach elimination levels of vaccination (e.g., 95% for measles). Global MCV1 coverage currently stands at 85%, with 40% of 194 countries meeting the 95% MCV1 coverage target, putting these countries and others striving to meet it in the critical range where effective epidemic risk assessment and vaccination strategy prioritization now require accounting for clustering [38].



**Fig. 5.** (A) Vaccination coverage among children aged 1–5 years by district and survey sample in Tanzania during the 1999 and 2010 Demographics and Health Surveys. (B) Empirical  $\tau(r)$  functions and fitted exponential for Tanzania and Zambia, using DHS cluster data.

Our results can help guide prioritization between overall vaccination coverage and reducing clustering of non-vaccination (Fig. 4). Where national measles vaccination coverage remains well below targets (i.e.  $\leq 90\%$ ), efforts to increase overall vaccination coverage likely provide substantially more benefit than efforts to reduce or address heterogeneities in vaccination coverage. However, where national vaccination coverage is near classical herd immunity levels, non-vaccination clustering has a strong impact on transmission potential; thus, strategies to account for and address this clustering may be imperative to elimination success.

As demonstrated with our derivation of an adjusted critical immunity threshold equation (Eq. (4)), one theoretical strategy to counter susceptibility clustering may be increased national vaccination targets. This may be useful where clustering is relatively low, as  $V_c^* \leq 95\%$  for measles with *low* or *medium* clustering. Thus, assuming successful immunization, current 95% targets may be sufficient to counter the effects of clustering [1]. However, where

95% coverage is unlikely or clustering is high, alternative strategies to address clustering directly are likely necessary.

Country-specific strategies, such as targeting specific vaccine deployment or individual characteristics might be effective. For example, clustering of non-vaccination has been associated with distance from health care and lack of health education, as well as geographic isolation and belonging to nomadic groups [39,40]. The challenge of these strategies is that they rely on knowledge of the specific causes of heterogeneity across a population. A more effective universal approach may be to shift from national targets once vaccination coverage reaches a threshold, such as 90% for measles, to local (i.e., regional, provincial) vaccination targets, and following from these local targets, local vaccination activities.

Though developed using measles, application of these methods readily extends to other infectious diseases, including pertussis, mumps, rubella, and polio, all of which currently present challenges worldwide. For example, recent mumps outbreaks among

college students and hockey teams may be the result of clustering of susceptibility due not to non-vaccination but to immunity waning, stemming from clustering by age [41]. With  $R_0 = 7.7$ , booster vaccination coverage of 93% among the student populations should be highly successful at producing herd immunity despite high susceptibility clustering. For diseases with lower  $R_0$  values, like mumps, our methods may more readily translate into feasible adjusted overall vaccination targets.

To provide a useful general rule of thumb for vaccination, we made several simplifying assumptions, in particular focusing on immunity acquired through successful vaccination. As such, our model did not explicitly account for naturally-acquired immunity, vaccine efficacy (VE), or multiple doses of vaccine, though it is possible to incorporate these effects, given available data (Supplemental Text 10.2).

Naturally-acquired immunity presents challenges to accurately estimating  $R^*$ , and if it occurs in a spatially localized manner, such as through an outbreak in a community, the nature of susceptibility clustering may change. However, as noted above, the greatest impact of clustering, and where this framework is most valuable, is when populations are close to disease elimination. While naturally-acquired immunity likely provides protection that our model does not account for, without continued transmission in these populations susceptibility will eventually mirror effective vaccination. For our example country Tanzania, because of existing natural immunity, our  $R^*$  estimates may be inflated. However, to sustain the recent successful interruption of transmission in Tanzania, immunization programs will likely need to maintain vaccination coverage higher than 95%, especially as the proportion of the population with natural immunity declines. Use of serological surveys to detail susceptibility (rather than vaccination coverage surveys) could circumvent the challenge posed by natural immunity, as serological surveys have the additional benefit of capturing clustering of naturally-acquired immunity [42,43].

Describing human contact introduces further challenges. We simplified contact into a spatial proximity construct defined by a gamma probability model, a simplistic approximation that we acknowledge does not fully capture the complexity of human social networks and mixing. Future explorations of the effects of susceptibility clustering will likely require more complex models of social interaction and behavior. Among these, age-assortativity, which has been well-described by the POLYMOD study, has previously been shown by Funk et al. to produce a similar effect to spatial clustering if susceptibility is age-clustered [44,19]. Other assortativity research, such as that defined by the Social Identity Theory and Self-Categorization Theory, suggests contact is also clustered in perceived social categories or identities, in which behaviors and perceptions, including vaccination skepticism, also cluster [45]. Building on work in behavioral modeling and social psychology, we may be able to quantify this social clustering, providing a better understanding of both transmission risk, and how to specifically target these socially-defined non-vaccination clusters and prevent outbreaks. This may be especially important as populations near elimination. Research on behavioral modeling and game theory suggests that, as people perceive less risk from the disease due to lower prevalence, their incentive to vaccinate may also decrease [46]. This may already be true in the U.S. and Europe [47].

Acquiring high-quality and sufficient data is critical to approximating susceptibility clustering and human contact patterns. For spatially-defined non-vaccination clustering and contact, we were able to leverage widely available DHS data, with which accurate estimation of spatial susceptibility clustering is likely possible without collecting new data. In settings where clustering of non-vaccination and contact patterns are less spatially-structured,

available data, such as mobile phone, social networks, and school-based data, may be useful. Ideally, as recent reviews have concluded, human behavior presented within infectious disease models should be based on detailed behavioral data, particularly that of actual behavior and not just intention, and should be specific to the target population and disease [48,49].

Even minimal clustering of non-vaccination, and resultant susceptibility, can produce substantial increases in outbreak risk, particularly for populations that are close to elimination targets. Through simple adjustments to the current rules of thumb guiding vaccination policy, we can estimate the impact of this heterogeneity and adjust strategies accordingly. For countries that are well-below current vaccination targets, the strategy should remain the same: continued efforts to achieve national vaccination coverage targets. For those near elimination, strategies that account for, or directly address, clustering of non-vaccination may yield more dividends. The approach presented here can serve to guide appropriate decision making and target setting.

## 10. Policy implications

### Policy Implications

1. The potential for disease transmission and outbreaks is greater than expected when vaccination coverage and susceptibility are clustered.
2. Vaccination targets are likely too low to achieve herd immunity in many settings because of susceptibility clustering.
3. Susceptibility clustering has the strongest impact as vaccination coverage approaches herd immunity levels.
4. Strategies should depend on the state of the country's disease control and vaccination coverage:
  - a. Low coverage: Continue national vaccination coverage targets.
  - b. High coverage: Shift focus to local vaccination coverage likely needed.

### Ethics

This research did not require ethical approval as it used publicly available de-identified data (DHS).

### Competing interests

We have no competing interests.

### Authors' contributions

Conception and design of study: SAT and JL.  
Acquisition of data: SAT.

Development and/or verification of analytic methods: SAT, JL, MG, CJEM.

Analysis and/or interpretation of data: SAT, JL, WJM.  
 Drafting the manuscript: SAT and JL.  
 Revising the manuscript: SAT, JL, MG, WJM, CJEM, MJF.  
 Approval of the final manuscript: SAT, JL, MG, WJM, CJEM, MJF.

## Acknowledgements

We thank Kelly Searle for providing data for estimation of the contact distance probability function for Zambia; Bryan Grenfell for providing his expertise and insight on the manuscript; and the Bill & Melinda Gates Foundation for funding this work.

## Funding

This work was supported by a Bill and Melinda Gates Foundation grant No. 90057624.

## Data, code and materials

Data for the Demographic and Health Surveys are publicly available from the DHS Program through <https://dhsprogram.com/>. Vaccination coverage data are available from the WHO at [http://apps.who.int/immunization\\_monitoring/global\\_summary/timeseries/tswucoveragemcv1.html](http://apps.who.int/immunization_monitoring/global_summary/timeseries/tswucoveragemcv1.html). Measles incidence data are available from the WHO at [http://www.who.int/immunization\\_monitoring\\_surveillance/burden/vpd/surveillance\\_type/active/measles\\_monthlydata/en/](http://www.who.int/immunization_monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/).

Code developed for the analyses, tables, and figures are available at: <https://github.com/shauntruelove/SpatialClustering>. The main equations and additional functions for the above analyses have been implemented as a package available at: <https://github.com/shauntruelove/SpatialClusteringPkg>

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.vaccine.2018.12.012>.

## References

- [1] World Health Organization. Global Measles and Rubella Strategic Plan, 2012–2020. Geneva, Switzerland: World Health Organization; 2012.
- [2] Measles elimination by 2020. WHO Reg Off South-East Asia SEARO 2013. <http://www.searo.who.int/mediacentre/releases/2013/pr1565/en/index.html> [accessed December 10, 2013].
- [3] World Health Organization. Measles and Rubella Surveillance Data. WHO n.d. [http://www.who.int/immunization/monitoring\\_surveillance/burden/vpd/surveillance\\_type/active/measles\\_monthlydata/en/](http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/) [accessed August 22, 2017].
- [4] Cutts FT, Lessler J, Metcalf CJ. Measles elimination: progress, challenges and implications for rubella control. *Exp Rev Vaccines* 2013;12:917–32. <https://doi.org/10.1586/14760584.2013.814847>.
- [5] Takashima Y, Schluter WW, Mariano KML, Diorditsa S, de Quiroz Castro M, Ou AC, et al. Progress toward measles elimination—Philippines, 1998–2014. *MMWR Morb Mortal Wkly Rep* 2015;64:357–62.
- [6] Plans-Rubió P. Evaluation of the establishment of herd immunity in the population by means of serological surveys and vaccination coverage. *Hum Vaccines Immunother* 2012;8:184–8. <https://doi.org/10.4161/hv.18444>.
- [7] Anderson RM. The concept of herd immunity and the design of community-based immunization programmes. *Vaccine* 1992;10:928–35.
- [8] Fine PE. Herd immunity: history, theory, practice. *Epidemiol Rev* 1993;15:265–302.
- [9] Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles: a systematic review. *Lancet Infect Dis* 2017. [https://doi.org/10.1016/S1473-3099\(17\)30307-9](https://doi.org/10.1016/S1473-3099(17)30307-9).
- [10] Anderson RM, May RM, Anderson B. *Infectious Diseases in Humans. Dynamics and Control*. Oxford: Oxford University Press; 1992.
- [11] Sakai R, Fink G, Wang W, Kawachi I. Correlation between pediatrician supply and public health in Japan as evidenced by vaccination coverage in 2010: secondary data analysis. *J Epidemiol* 2015;25:359–69. <https://doi.org/10.2188/jea.IE20140121>.
- [12] Seither R, Calhoun K, Knighton CL, Mellerson J, Meador S, Tippins A, et al. Vaccination coverage among children in kindergarten - united states, 2014–15 school year. *MMWR Morb Mortal Wkly Rep* 2015;64:897–904.
- [13] Shrivastwa N, Gillespie BW, Kolenic GE, Lepkowski JM, Boulton ML. Predictors of vaccination in India for children aged 12–36 months. *Vaccine* 2015;33 (Supplement 4):D99–D105. <https://doi.org/10.1016/j.vaccine.2015.09.034>.
- [14] Takahashi S, Metcalf CJ, Ferrari MJ, Tatem AJ, Lessler J. The geography of measles vaccination in the African Great Lakes region. *Nat Commun* 2017;8:15585. <https://doi.org/10.1038/ncomms15585>.
- [15] Omer SB, Enger KS, Moulton LH, Halsey NA, Stokley S, Salmon DA. Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. *Am J Epidemiol* 2008;168:1389–96. <https://doi.org/10.1093/aje/kwn263>.
- [16] Atwell JE, Otterloo JV, Zipprich J, Winter K, Harriman K, Salmon DA, et al. Nonmedical vaccine exemptions and pertussis in California, 2010. *Pediatrics* 2013;132:624–30. <https://doi.org/10.1542/peds.2013-0878>.
- [17] Gastañaduy PA, Budd J, Fisher N, Redd SB, Fletcher J, Miller J, et al. A Measles outbreak in an underimmunized Amish community in Ohio. *N Engl J Med* 2016;375:1343–54. <https://doi.org/10.1056/NEJMoa1602295>.
- [18] Williamson G, Ahmed B, Kumar PS, Ostrov BE, Ericson JE. Vaccine-preventable diseases requiring hospitalization. *Pediatrics* 2017;140:e20170298. <https://doi.org/10.1542/peds.2017-0298>.
- [19] Funk S, Knapp JK, Lebo E, Reef SE, Dabbagh AJ, Kretsinger K, et al. Target immunity levels for achieving and maintaining measles elimination. *BioRxiv* 2017:201574. <https://doi.org/10.1101/201574>.
- [20] May RM. Network structure and the biology of populations. *Trends Ecol Evol* 2006;21:394–9. <https://doi.org/10.1016/j.tree.2006.03.013>.
- [21] Salathé M, Bonhoeffer S. The effect of opinion clustering on disease outbreaks. *J R Soc Interface* 2008;5:1505–8. <https://doi.org/10.1098/rsif.2008.0271>.
- [22] Ferguson NM, Donnelly CA, Anderson RM. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* 2001;413:542–8. <https://doi.org/10.1038/35097116>.
- [23] Bhoomboonchoo P, Gibbons RV, Huang A, Yoon I-K, Buddhari D, Nisalak A, et al. The spatial dynamics of dengue virus in Kamphaeng Phet Thailand. *PLoS Negl Trop Dis* 2014;8:e3138. <https://doi.org/10.1371/journal.pntd.0003138>.
- [24] Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 2001;294:813–7. <https://doi.org/10.1126/science.1065973>.
- [25] Salje H, Cummings DAT, Lessler J. Estimating infectious disease transmission distances using the overall distribution of cases. *Epidemics* 2016;17:10–8. <https://doi.org/10.1016/j.epidem.2016.10.001>.
- [26] Salje H, Lessler J, Endy TP, Curriero FC, Gibbons RV, Nisalak A, et al. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc Natl Acad Sci* 2012;109:9535–8. <https://doi.org/10.1073/pnas.1120621109>.
- [27] Lessler J, Salje H, Grabowski MK, Cummings DAT. Measuring Spatial Dependence for Infectious Disease Epidemiology. *PLoS One* 2016;11: e0155249. <https://doi.org/10.1371/journal.pone.0155249>.
- [28] Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social mixing patterns in rural and urban areas of southern China. *Proc Biol Sci* 2014;281:20140268. <https://doi.org/10.1098/rspb.2014.0268>.
- [29] Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C. A tale of many cities: universal patterns in human urban mobility. *PLoS One* 2012;7:e37027. <https://doi.org/10.1371/journal.pone.0037027>.
- [30] Searle KM, Lubinda J, Hamapumbu H, Shields TM, Curriero FC, Smith DL, et al. Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. *R Soc Open Sci* 2017;4:170046. <https://doi.org/10.1098/rsos.170046>.
- [31] The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs n.d. <http://www.dhsprogram.com/> [accessed June 27, 2016].
- [32] Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol* 2006;164:936–44. <https://doi.org/10.1093/aje/kwj317>.
- [33] Kanaan MN, Farrington CP. Matrix models for childhood infections: a Bayesian approach with applications to rubella and mumps. *Epidemiol Infect* 2005;133:1009–21. <https://doi.org/10.1017/S0950268805004528>.
- [34] Colson KE, Zúñiga-Brenes P, Ríos-Zertuche D, Conde-Glez CJ, Gagnier MC, Palmisano E, et al. Comparative estimates of crude and effective coverage of measles immunization in low-resource settings: findings from Salud Mesoamérica 2015. *PLoS One* 2015;10:e0130697. <https://doi.org/10.1371/journal.pone.0130697>.
- [35] Chao DL, Halloran ME, Longini IM. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci* 2011;108:7081–5. <https://doi.org/10.1073/pnas.1102149108>.
- [36] Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University Press; 2011.
- [37] World Health Organization. Data, statistics and graphics. WHO n.d. <[http://www.who.int/immunization/monitoring\\_surveillance/data/en/](http://www.who.int/immunization/monitoring_surveillance/data/en/)> [accessed July 1, 2016].
- [38] Dabbagh A. Progress toward regional measles elimination – worldwide, 2000–2017. *MMWR Morb Mortal Wkly Rep* 2018;67. <https://doi.org/10.15585/mmwr.mm6747a6>.
- [39] Metcalf C, Tatem A, Bjornstad O, Lessler J, O'Reilly K, Takahashi S, et al. Transport networks and inequities in vaccination: remoteness shapes measles

- vaccine coverage and prospects for elimination across Africa. *Epidemiol Infect* 2015;143:1457–66. <https://doi.org/10.1017/S0950268814001988>.
- [40] Brownwright TK, Dodson ZM, van Panhuis WG. Spatial clustering of measles vaccination coverage among children in sub-Saharan Africa. *BMC Public Health* 2017;17:957. <https://doi.org/10.1186/s12889-017-4961-9>.
- [41] Centers for Disease Control and Prevention. Mumps | Cases and Outbreaks n.d. <https://www.cdc.gov/mumps/outbreaks.html> [accessed June 12, 2017].
- [42] Metcalf CJE, Farrar J, Cutts FT, Basta NE, Graham AL, Lessler J, et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet* 2016;388:728–30. [https://doi.org/10.1016/S0140-6736\(16\)30164-7](https://doi.org/10.1016/S0140-6736(16)30164-7).
- [43] Lessler J, Metcalf CJE, Cutts FT, Grenfell BT. Impact on epidemic measles of vaccination campaigns triggered by disease outbreaks or serosurveys: a modeling study. *PLoS Med* 2016;13:e1002144. <https://doi.org/10.1371/journal.pmed.1002144>.
- [44] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008;5:e74. <https://doi.org/10.1371/journal.pmed.0050074>.
- [45] Attwell K, Smith DT. Parenting as politics: social identity theory and vaccine hesitant communities. *Int J Health Gov* 2017;22:183–98. <https://doi.org/10.1108/IJHG-03-2017-0008>.
- [46] Shim E, Grefenstette JJ, Albert SM, Cakouros BE, Burke DS. A game dynamic model for vaccine skeptics and vaccine believers: measles as an example. *J Theor Biol* 2012;295:194–203. <https://doi.org/10.1016/j.jtbi.2011.11.005>.
- [47] Why (some) parents don't vaccinate. IdeasTedCom 2016. <https://ideas.ted.com/why-some-parents-dont-vaccinate/> [accessed December 7, 2018].
- [48] Verelst F, Willem L, Beutels P. Behavioural change models for infectious disease transmission: a systematic review (2010–2015). *J R Soc Interface* 2016;13:20160820. <https://doi.org/10.1098/rsif.2016.0820>.
- [49] Weston D, Hauck K, Amlôt R. Infection prevention behaviour and infectious disease modelling: a review of the literature and recommendations for the future. *BMC Public Health* 2018;18:336. <https://doi.org/10.1186/s12889-018-5223-1>.