

## Supplementary Text 1: Data simulation

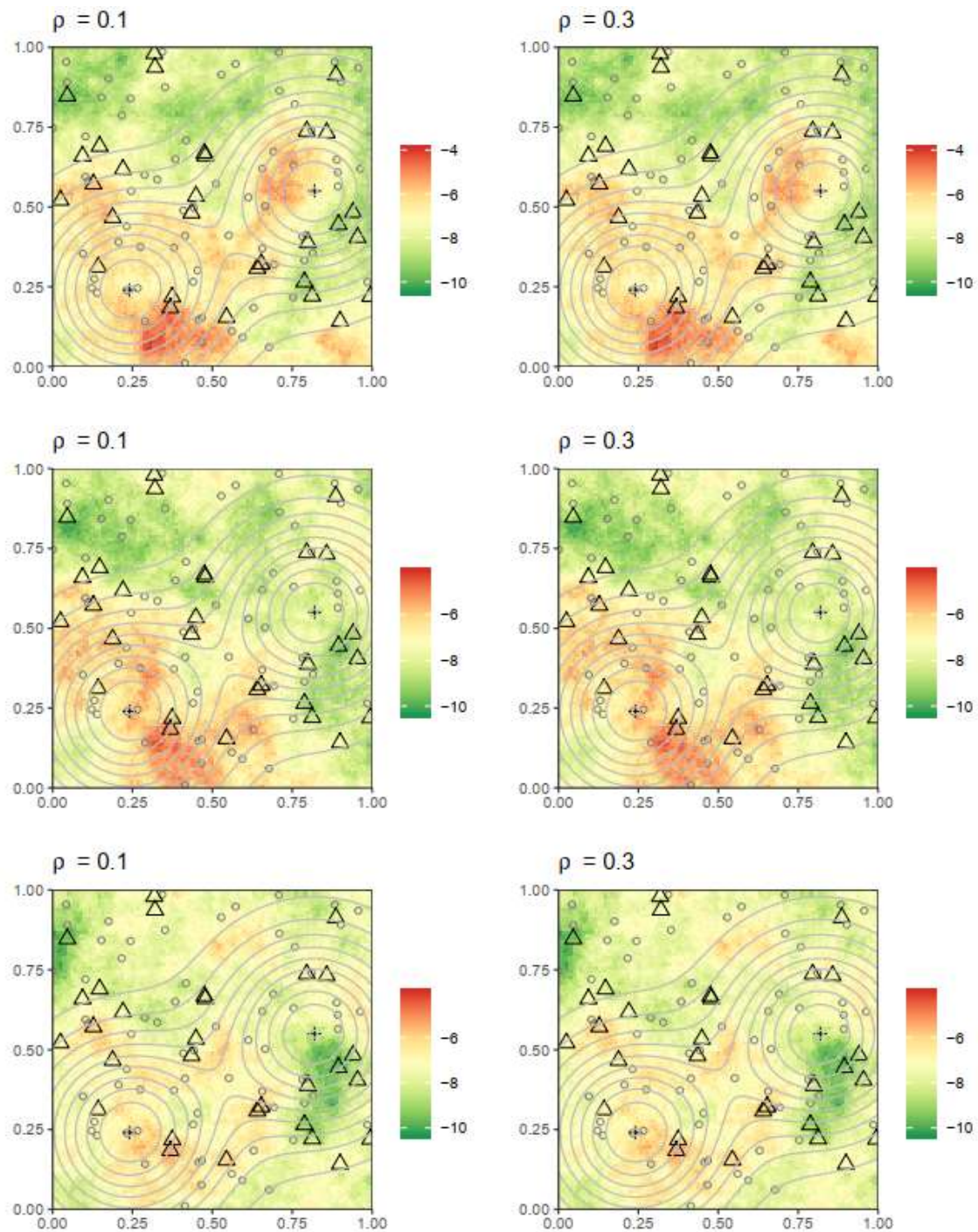
**Other input data.** A risk factor  $X$  that emanated from two point sources, such as the locations of mass gatherings or factories that emit air pollution, was generated. The level of the risk factor at  $s$  is determined by  $X(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d_{s,sp1}^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d_{s,sp2}^2}{2\sigma^2}}$ , with  $\sigma$  equals 0.2 and  $d_{s,spi}$  represents the distance between  $s$  and the  $i$ th point source.

**Epidemiologic data.** We simulate disease prevalence  $Y$  for 100 random locations within a unit square based on the known spatial distribution of a causal risk factor,  $X$ , and under two distinct scenarios of spatial autocorrelation in disease outcomes. The log prevalence for these sites is generated as:

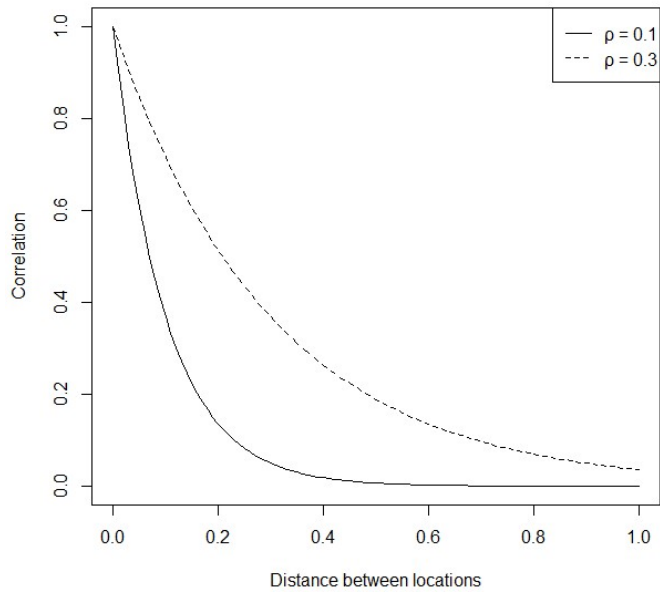
$$\log(Y) = \beta_0 + \beta_1 X + \eta + \varepsilon \quad (1)$$

where  $\beta_0$  represents the overall log mean prevalence,  $\beta_1$  represents the effect of a unit increase in risk factor  $X$ ,  $\eta$  represents a mean-zero Gaussian process accounting for the spatial correlation contributed by unmeasured risk factors, and  $\varepsilon$  represents i.i.d. mean-zero normally distributed noise with a variance of  $\sigma_\varepsilon^2$ . The spatially correlated error term  $\eta$  can be represented by a multivariate normal distribution with a variance-covariance matrix  $\mathbf{C}$ , in which each entry  $c_{ij}$  represents the covariance between the residuals at the  $i$ th and the  $j$ th location when  $i \neq j$ , and the spatial variance  $\sigma_s^2$  when  $i = j$ . We simulate our disease system to be consistent with the widely used second-order stationary assumption (the covariance between the residuals at two locations only depends on the distance between them) and exponential covariance function in specifying  $\mathbf{C}$ , so the covariance between sites  $i$  and  $j$  can be written as  $c_{ij} = \sigma_s^2 e^{-d_{ij}/\rho}$ . The spatial range parameter,  $\rho$ , defines the range of spatial autocorrelation, and it is this parameter that we vary to obtain scenarios of relatively smooth ( $\rho = 0.3$ ) and patchy ( $\rho = 0.1$ ) spatial variation in prevalence unexplained by  $X$ . The parameters  $\beta_0$ ,  $\beta_1$ ,  $\sigma_s$ , and  $\sigma_\varepsilon$  are set to be -8.5, 1, 1, and 0.1, respectively. Only the disease data at a randomly sampled 30 in-network sites  $\{s_1 \dots s_{30}\}$  are available to the designer. Data at the other 70 unmonitored locations  $\{s_{31} \dots s_{100}\}$  are treated as unknown.

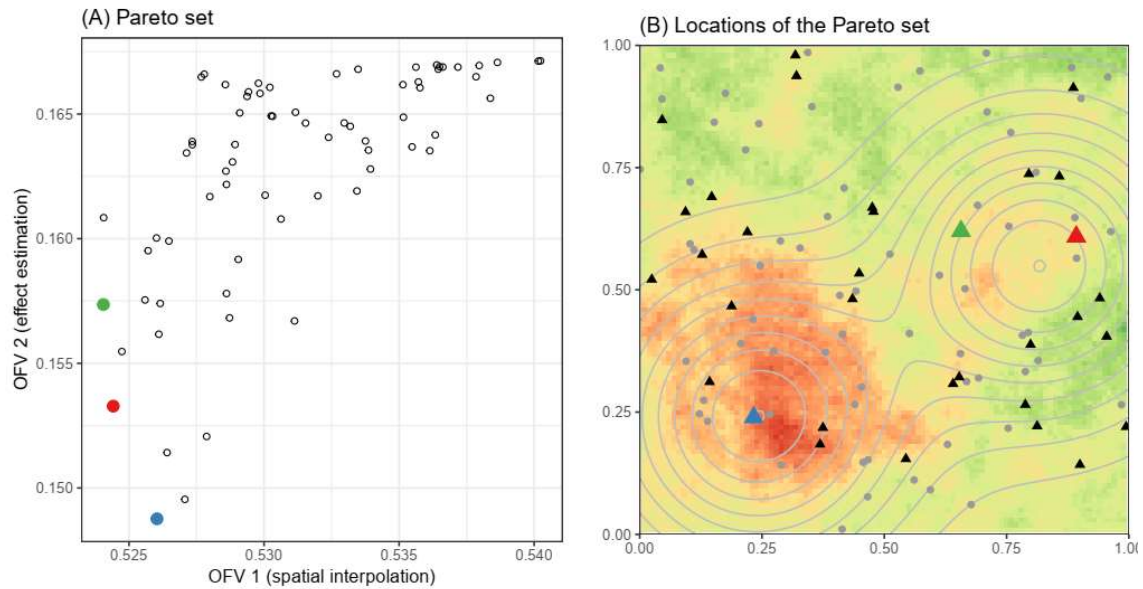
**Figure S1. Five realizations of the log prevalence surface when  $\rho = 0.1$  (left panels) or 0.3 (right panels).**



**Figure S2. Correlation as a function of distance between locations when  $\rho = 0.1$  (solid curve) or 0.3 (dashed curve).**

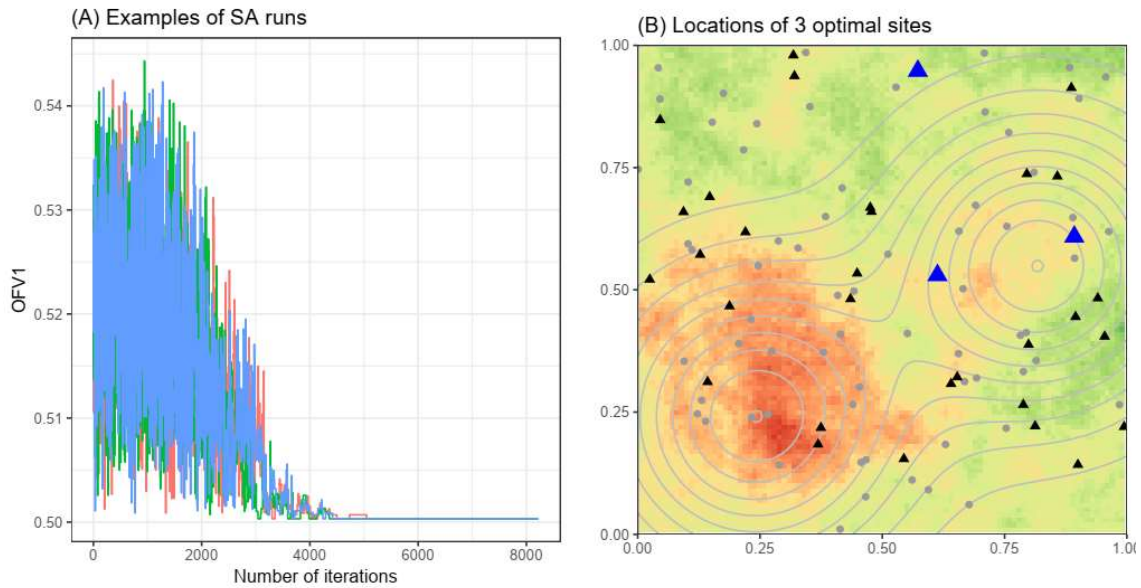


**Figure S3. Results from Pareto optimization ( $\rho=0.1$ ).** (A) OFV1 and OFV2 of the Pareto set (colored dots) and all other each candidate site (hollow dots). (B) Spatial locations of the Pareto set (colored triangles) colored by the same color scheme as in Panel A. Black triangles represent existing sites, and gray dots represent unchosen alternative sites. Background color represents log prevalence value when  $\rho = 0.3$  using the same color scheme as in Figure 2C, while contour lines represent levels of risk factor X.

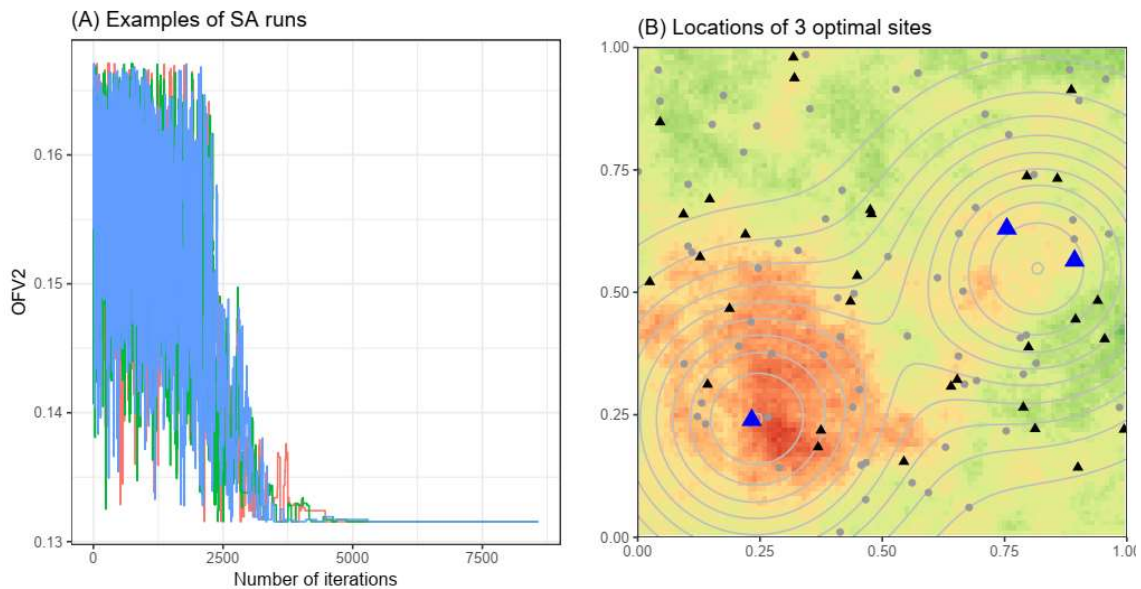


**Figure S4. Iterative optimization with simulated annealing (spatial interpolation,  $\rho=0.1$ ).** (A) OFV1 against the number of iterations in 3 SA runs. (B) The locations of the optimal 3 sites. Black

triangles represent existing sites, blue triangles represent the optimal additional sites, and gray dots represent unchosen alternative sites. Background color represents log prevalence value when  $\rho = 0.3$  using the same color scheme as in Figure 2C, while contour lines represent levels of risk factor X.



**Figure S5. Iterative optimization with simulated annealing (effect estimation,  $\rho=0.1$ ).** (A) OFV1 against the number of iterations in 3 SA runs. (B) The locations of the optimal 3 sites. Black triangles represent existing sites, blue triangles represent the optimal additional sites, and gray dots represent unchosen alternative sites. Background color represents log prevalence value when  $\rho = 0.3$  using the same color scheme as in Figure 2C, while contour lines represent levels of risk factor X.



**Figure S6. Iterative optimization with simulated annealing (effect estimation,  $\rho=0.3$ ).** (A) OFV1 against the number of iterations in 3 SA runs. (B) The locations of the optimal 3 sites. Black triangles represent existing sites, blue triangles represent the optimal additional sites, and gray dots represent unchosen alternative sites. Background color represents log prevalence value when  $\rho = 0.3$  using the same color scheme as in Figure 2C, while contour lines represent levels of risk factor X.

