Mathematical analysis to prioritise strategies for malaria elimination

Nakul Chitnis [a,b,*], Allan Schapira [a,b], Christian Schindler [a,b], Melissa A. Penny [a,b], Thomas A. Smith [a,b]

[a] *Swiss Tropical and Public Health Institute, Socinstrasse 57, Basel 4002, Switzerland*
[b] *University of Basel, Basel 4003, Switzerland*

A B S T R A C T

Malaria and some other tropical diseases are currently targeted for elimination and eventually eradication. Since resources are limited, prioritisation of countries or areas for elimination is often necessary. However, this prioritisation is frequently conducted in an *ad hoc* manner. Lower transmission areas are usually targeted for elimination first, but for some areas this necessitates long and potentially expensive surveillance programs while transmission is eliminated from neighbouring higher transmission areas. We use a mathematical model to compare the implications of prioritisation choices in reducing overall burden and costs. We show that when the duration of the elimination program is independent of the transmission potential, burden is always reduced most by targeting high transmission areas first, but to reduce costs the optimal ordering depends on the actual transmission levels. In general, when overall transmission potential is low and the surveillance cost per secondary case compared to the cost per imported case is low, targeting the higher transmission area for elimination first is favoured.

1. Introduction

The *Global Technical Strategy for Malaria 2016–2030*, released by the World Health Organization (WHO) (World Health Organization, 2015), sets the current global goals for malaria control and elimination by 2030 as: (i) reducing the number of malaria cases and deaths globally by 90% as compared to 2015; and (ii) eliminating and preventing re-establishment of transmission in at least 35 countries where malaria transmission was ongoing in 2015. The main strategies towards achieving these goals are (i) "control through universal access to malaria prevention, diagnosis and treatment"; (ii) intensifying efforts towards elimination and prevention of re-introduction; and (iii) "transforming malaria surveillance into a core intervention" of both control and elimination strategies (World Health Organization, 2015; 2017).

Malaria control activities are recommended in all locations where transmission persists (although it is sometimes not deployed in locations where financial and/or operational resources are insufficient). However, efforts to eliminate malaria are mainly focused on the fringes of its geographical range, for example in the Asia-Pacific region (Gosling et al., 2012) and in Southern Africa (Southern African Development Community, 2013).

There is a global health priority in eliminating foci of drug resistance in the Greater Mekong subregion (Gueye et al., 2014), and spatially progressive elimination may be rational where the risk of re-establishment of transmission is low (Lines et al., 2008; Smith et al., 2013), or where a small focus of transmission has a disproportionate economic importance. Targeting isolated islands and other areas with low transmission potential for malaria elimination may also have value as tests of new technologies or systems; but in general it is unclear whether targeting low transmission areas is a better strategy than focusing those resources on eliminating malaria from higher transmission areas, especially when these lower transmission areas face risks of malaria importation from neighbouring higher transmission areas.

The strategy of progressive elimination from the fringes has been criticised because of the implicit inequity of prioritising low burden areas (Shah, 2010). This strategy also ignores the important lesson from those programs that have been successful in eradicating a disease (smallpox (Henderson, 2009), or that have approached eradication (polio (Aylward et al., 2003) and dracunculiasis (Ruiz-Tiben and Hopkins, 2006), that eradication programs need to focus early in challenging areas which are likely to remain a threat after the disease is gone elsewhere. More generally, it seems likely that any eradication or elimination program

will be more efficient if core areas that export the infection are targeted at the start. This concurs with the experience of countries that have eliminated or approached elimination. Iran is now on its second attempt at national malaria elimination. Each time, interventions were rolled out nationally, burden reduction in the high transmission south east has been key to near-elimination in the north and west (Hemami et al., 2013).

An effective surveillance system is a key constituent intervention of intensified control efforts to eliminate malaria; and remains essential after malaria has been eliminated to prevent re-establishment of transmission (Kelly et al., 2012; World Health Organization, 2015). It must include a reactive component that is effective in detecting imported cases and preventing onward transmission from them. The surveillance system will need to be maintained as long as there is a risk of reintroduction, that is, until malaria has been eradicated. This can be operationally and financially challenging in tropical areas with high vectorial capacity (malaria transmission potential) (Roberts, 2010; Schapira and Zamani, 2012; Smith et al., 2011; World Health Organization, Global Malaria Programme and University of California, San Francisco, 2012).

There is therefore a need for decisions concerning elimination to be based on criteria that consider costs, overall disease burden and the risks associated with different options. This paper aims to provide a mathematical formulation to guide strategic thinking about how zones with different levels of disease burden should be targeted for elimination, on the assumptions that (i) malaria control is maintained in all areas; (ii) elimination is technically possible in all the areas being considered; (iii) but resources to intensify control programs to achieve elimination are limited to targeting one area at a time. The results are expressed in terms of general principles that may be applicable at different spatial scales across the whole range of malaria transmission intensities. To derive these principles, we only consider a simple economic model of two areas of equal population here, but some of this analysis may be extended to multiple areas. Although many studies, modelling and otherwise, have investigated the feasibility of malaria elimination (Moonen et al., 2010), and the persistence of elimination (Chiyaka et al., 2013; Smith et al., 2013), none have considered such mathematical economic models for ordering areas for intensified control efforts to achieve elimination.

We consider a simple system consisting of two connected geographical areas with similar populations but different initial levels of transmission, and equal (symmetric) movement of in both directions through the short term movement of people and possibly mosquito vectors. Symmetrical movement is a reasonable assumption here because imported malaria cases are usually not due to immigration but due to the short term movement of visitors from areas with higher transmission or returning residents. Therefore we assume that the importation of infection in either direction depends only on the prevalence in the source population.

Both areas are initially under control, with current tools, such as long lasting insecticidal nets and indoor residual spraying, maintaining the annual disease burden in each area at a relatively constant level. Although there are likely to be seasonal variations within each area, we do not explicitly consider them here because we are more interested in the general principles of the relationship between importation and elimination (which do not depend on seasonality), and not in the details of planning such elimination strategies (which depend on seasonality).

We assume that the technologies for time-limited elimination throughout the system are available (technical feasibility) with additional tools such as reactive case detection and reactive vector control (Moonen et al., 2010; World Health Organization, 2017). However, the human and/or financial resources to apply these additional intervention measures in both areas simultaneously are
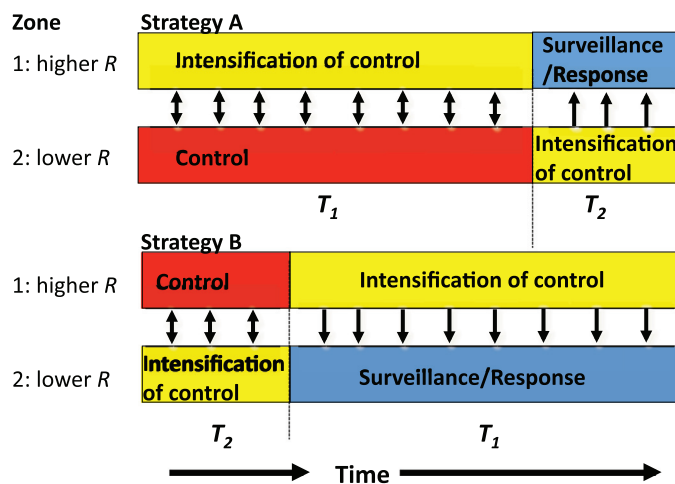

Fig. 1. Strategies being compared. The black arrows indicate the direction of the export/import of infected cases. Here, R denotes the transmission potential and T_i denotes the duration of intensified measures needed to achieve elimination (Table 1).

not available, so that overall elimination can only be achieved by intensified control measures in one area at a time, in each case until local transmission is interrupted.

We label the higher transmission site as $i = 1$, with transmission potential (measured by the reproduction number) R_1 (see Table 1). We label the lower transmission site as $i = 2$ with transmission potential R_2 where $R_2 < R_1$. Both sites are additionally characterised by malaria prevalence at equilibrium, p_i, annual disease burden, B_i, vulnerability, V_i, and required duration of intensified control to eliminate transmission, T_i. More detailed descriptions of these and all other parameters are provided in Table 1. We define elimination here as the lack of sustained local transmission (that is, imported cases may lead to a few secondary cases but each chain of infection dies out and malaria cannot reestablish itself in the population). We use the WHO definition of elimination as the "interruption of local transmission (reduction to zero incidence of indigenous cases)" (World Health Organization, 2016). WHO malaria terminology distinguishes between introduced cases ("first generation local transmission" from an imported case) and indigenous cases ("contracted locally with no evidence of importation and no direct link to transmission from an imported case") (World Health Organization, 2016). Re-establishment of transmission is defined as the "renewed presence of a measurable incidence of locally acquired malaria infection due to repeated cycles of mosquito-borne infections in an area in which transmission had been interrupted" (World Health Organization, 2016). Therefore in a state of elimination, local transmission is possible so secondary (introduced) cases may arise from an imported case, but these should not lead to sustained (re-established) transmission.

There are two possible strategies for elimination across both sites: $J = A$, intensification of control in the higher transmission site ($i = 1$) first, with continued control in lower burden area; and $J = B$, intensification in the lower transmission site ($i = 2$) first, with continued routine control in the higher burden area. These two strategies are illustrated in Fig. 1.

We note that the intensification program must include an effective surveillance response system. After the intensification program has eliminated malaria in the first area and has moved to the second area, the surveillance response system must continue to operate in the first area to prevent reintroduction (re-establishment of transmission), but other control interventions may be withdrawn. We therefore assume that after malaria elimination in the first transmission zone, the transmission potential for secondary cases remains at pre-intervention levels, but there is an effective surveil-

Table 1

Notation and definitions. Here i denotes the site (where $i \in \{1, 2\}$ and $i = 1$ represents the higher transmission site), and J denotes the strategy (where $J \in \{A, B\}$ and $J = A$ represents the strategy of targeting the higher transmission site first).

	Description	Dimensions
R_i	Average transmission potential in area i before intensified control measures. Transmission potential is the control reproduction number, sometimes denoted as R_c, (Smith et al., 2009).	Dimensionless
p_i	Equilibrium prevalence in source area i (before intensified elimination measures).	Proportion
B_i	Pre-elimination burden of disease per unit of time in area i.	Disease events / (person time)
b_i	Average burden of disease per unit time during intensified elimination program in area i.	Disease events / (person time)
B_{max}	Maximum possible burden.	Disease events / (person time)
ψ_B	Saturation factor for burden.	Dimensionless
V_i	Vulnerability of area i following elimination from area i only. Vulnerability is defined as either proximity to a malarious area or frequent influx of infected humans and/or infective mosquitoes (World Health Organization, 2012). Here we assume vulnerability is equivalent to the rate of imported infections.	Infections / (person time)
η	Extent of migration of humans and/or mosquitoes between the populations.	1 / (person time)
V_{max}	Maximum possible vulnerability.	Infections / (person time)
ψ_V	Saturation factor for vulnerability.	Dimensionless
T_i	Duration of intensified measures needed to achieve elimination in area i.	Time
T_{max}	Maximum possible duration of intensified measures needed to achieve elimination.	Time
ψ_T	Saturation factor for time to elimination.	Dimensionless
E_J	Total burden of disease over the whole program, conditional on strategy J.	Disease events / person
ΔE	Net health benefit of strategy A (i.e., the burden averted relative to the alternative strategy B).	Disease events / person
S_J	Total surveillance costs, conditional on strategy J.	Currency
S_0	Fixed cost of a surveillance system.	Currency
β	Health systems cost per primary imported case.	Currency
γ	Health systems cost per secondary case.	Currency
C_J	Total cost of strategy J.	Currency
ΔC	Cost saving with strategy A compared with strategy B.	Currency
λ	Ceiling ratio: the amount of money that the program is willing to pay to avert disease.	Currency / (disease event)
M	Net monetary benefit of strategy A.	Currency

lance system, that prevents re-establishment of transmission from these secondary cases.

2. Assumptions of transmission potential

Determining the preferred strategy to minimise overall disease burden or costs requires additional assumptions about the relationships of the transmission potential, R, with the annual disease burden, B, the vulnerability, V, and the required time for intensified control measures to interrupt transmission, T. These three relationships in turn depend on the relationship between the transmission potential, R, and the equilibrium prevalence, p, during the control phase.

2.1. Relationship between prevalence and transmission potential

The general form of the relationship of the equilibrium prevalence to the reproduction number is different in different models of malaria transmission. When transmission potential is low, infections are sporadic; therefore immunity is negligible, neither sequential nor concurrent co-infections are likely, and simple susceptible-infected-susceptible models (corresponding to the original model of Ross, 1916) represent a reasonable approximation to malaria dynamics. In this case, p_i in the stable state relates to reproduction number via

$$p_i = 1 - \frac{1}{R_i}, \tag{1}$$

with derivative

$$\frac{dp_i}{dR_i} = \frac{1}{R_i^2}. \tag{2}$$

It follows that, in the limiting case of very low transmission, $R_i = 1$,

$$\frac{dp_i}{dR_i} \approx 1, \tag{3}$$

implying that p_i is approximately directly proportional to R_i with an offset of 1 (Fig. 2(a)) and gradient 1, i.e. $p_i \approx R_i - 1$. In high transmission settings where immunity and superinfection play important roles, prevalence saturates at some level, p_{max}, that depends on the extent of transmission heterogeneity but is independent of R_i. Although the value of R_i at which p_i saturates depends on transmission heterogeneity and the dynamics of acquired immunity, the general shape of the relationship between p_i and R_i can be approximated with curves of the form,

$$p_i = p_{max}(1 - \exp(-\psi_P(R_i - 1))), \tag{4}$$

where the saturation factor, ψ_P, is a positive constant (Fig. 2(a)). This assumption fits well with analysis of data relating malaria prevalence to the entomological inoculation rate across different sites in Africa, which considered such an exponential curve amongst others (Smith et al., 2005).

2.2. Relationship between annual disease burden and transmission potential

On the assumption that the populations are initially close to a steady state, pre-elimination annual disease burden, B_i, is related to prevalence, p_i, and hence to R_i by some monotonic function, $B_i = f_B(R_i)$. Since $R_1 > R_2$, it follows that $B_1 > B_2$ and one set of analyses considers the inferences that can be drawn without making any stronger assumptions about transmission-burden relationships (Table 2). In addition, we consider the inferences that follow from assuming that pre-elimination disease burden saturates as transmission potential increases, following the relationship,

$$B_i = B_{max}(1 - \exp(-\psi_B(R_i - 1))), \tag{5}$$

where B_{max} and ψ_B are two positive constants defining the relationship between burden and transmission potential, analogous to the relationship between prevalence and transmission potential (Fig. 2(b)).

Table 2
Scenarios modelled and results for analysis of burden.

	$T_1 = T_2 = T$: *Fixed duration of intensified measures to reduce incidence to negligible levels in area i*	$T_1 > T_2$: *Duration increases with transmission*
$B_1 > B_2$: *Burden increases with transmission*	Strategy A minimises overall burden.	Strategy A minimises overall burden if and only if $B_1/B_2 > T_1/T_2$.

Table 3
Scenarios modelled and results for analysis of costs.

	$T_1 = T_2 = T$: *Fixed duration of intensified measures to reduce incidence to negligible levels in area i*	$T_1 > T_2$: *Duration increases with transmission*
$V_i = \eta(R_j - 1)$ for $i, j \in \{1, 2\}$ and $i \neq j$: *Low transmission*	Strategy A has lower cost	Strategy A has lower cost
$V_1 = V_2$: *High transmission*	Strategy B has lower cost	As R_2 increases, strategy A is more likely to have lower cost; as R_1 increases, strategy B is more likely to have lower cost

2.3. Relationship between vulnerability and transmission potential

V_i, the vulnerability of sink population, i, to source population j (for $j \in \{1, 2\}$, $j \neq i$), depends on the rate at which infections are exported from the source, which is proportional to the prevalence in the source population, p_j. p_j is a function of R_j, implying that V_i can also be expressed as a similar function of R_j: $V_i = \eta p_j = f_V(R_j)$, where the constant of proportionality η captures the extent of migration of humans and/or mosquitoes between the populations.

In low transmission settings, vulnerability, like prevalence, is proportional to the transmission potential of the source patch,

$$V_1 \approx \eta(R_2 - 1), \tag{6a}$$

$$V_2 \approx \eta(R_1 - 1). \tag{6b}$$

At high transmission, vulnerability, like prevalence, saturates so that if both areas are at high transmission, then $V_1 = V_2 = V$ (Table 3).

Similar to the relationship between R and p, the general shape of the relationship between V and R can be approximated with curves of the form,

$$V_1 = V_{max}(1 - \exp(-\psi_V(R_2 - 1))), \tag{7a}$$

$$V_2 = V_{max}(1 - \exp(-\psi_V(R_1 - 1))), \tag{7b}$$

where V_{max} and ψ_V are two positive constants defining the relationship between vulnerability and transmission potential, and where

$$\left.\frac{dV_1}{dR_2}\right|_{V_1=0} = \left.\frac{dV_2}{dR_1}\right|_{V_2=0} = \eta = V_{max}\psi_V,$$

with the low and high transmission examples described above represent limiting cases (Fig. 2(a)).

2.4. Relationship between duration of intensified efforts to interrupt transmission and transmission potential

The duration over which intensified measures must be sustained to interrupt transmission depends on multiple factors affecting intervention effectiveness. The present analyses consider only the effect of transmission potential, R, so that $T_i = f_T(R_i)$, and the areas are treated as equivalent in other respects.

During the Global Malaria Eradication Programme (GMEP) era from 1955 to 1969, it was assumed that an initial attack phase would comprise of maximal scale up of vector control interventions to interrupt transmission (Pampana, 1963); and that this would only need to be sustained for the natural lifetime of the

residual infections, which is approximately independent of the transmission rate, (Macdonald and Göckel, 1964). Therefore, the simplest set of assumptions about the required duration of the elimination measures (Tables 2 and 3) is that they are equal in the two sites,

$$T_1 = T_2 = T.$$

However, it is possible that the required duration of the elimination measures is longer in higher transmission settings, so that T increases with R and $T_1 > T_2$. This assumption underlies a further set of analyses (Tables 2 and 3). In general, as intervention programs are prolonged and intensified, the environmental drivers of malaria transmission (i.e. the determinants of R) may become less important as determinants of program outcomes, relative to operational effectiveness of control programs (see e.g., Mabaso et al., 2007), so that it is to be expected that the function relating T to R will saturate, like those relating B and V to R, and hence can be approximated with

$$T_i = T_{max}(1 - \exp(-\psi_T(R_i - 1))), \tag{8}$$

where T_{max} and ψ_T are two positive constants defining the relationship between the required duration of the elimination program and transmission potential (Fig. 2(c)).

3. Analysis

3.1. Burden of disease

The total burden of disease, E_J, experienced with strategy J, is the sum of the burden during the intensified control phase in the first area to experience intensified control, the pre-elimination burden in the other area, and the burden during intensified control in the second area. For strategy A this is (see Table 1 for notation)

$$E_A = (b_1 + B_2)T_1 + b_2T_2. \tag{9}$$

Similarly, the burden experienced with strategy B is

$$E_B = (b_2 + B_1)T_2 + b_1T_1. \tag{10}$$

The health benefit of the decision for strategy A is

$$\Delta E = E_B - E_A$$
$$= B_1T_2 - B_2T_1, \tag{11}$$

where strategy A minimises burden if $\Delta E > 0$. We determine which strategy minimises overall burden, assuming that B_1 is always greater than B_2, for different assumptions about the dependence of T_i on transmission as indicated in Table 2.

In the special case of $T_1 = T_2 = T$, corresponding to a fixed duration of intensified measures as sometimes achieved by the initial
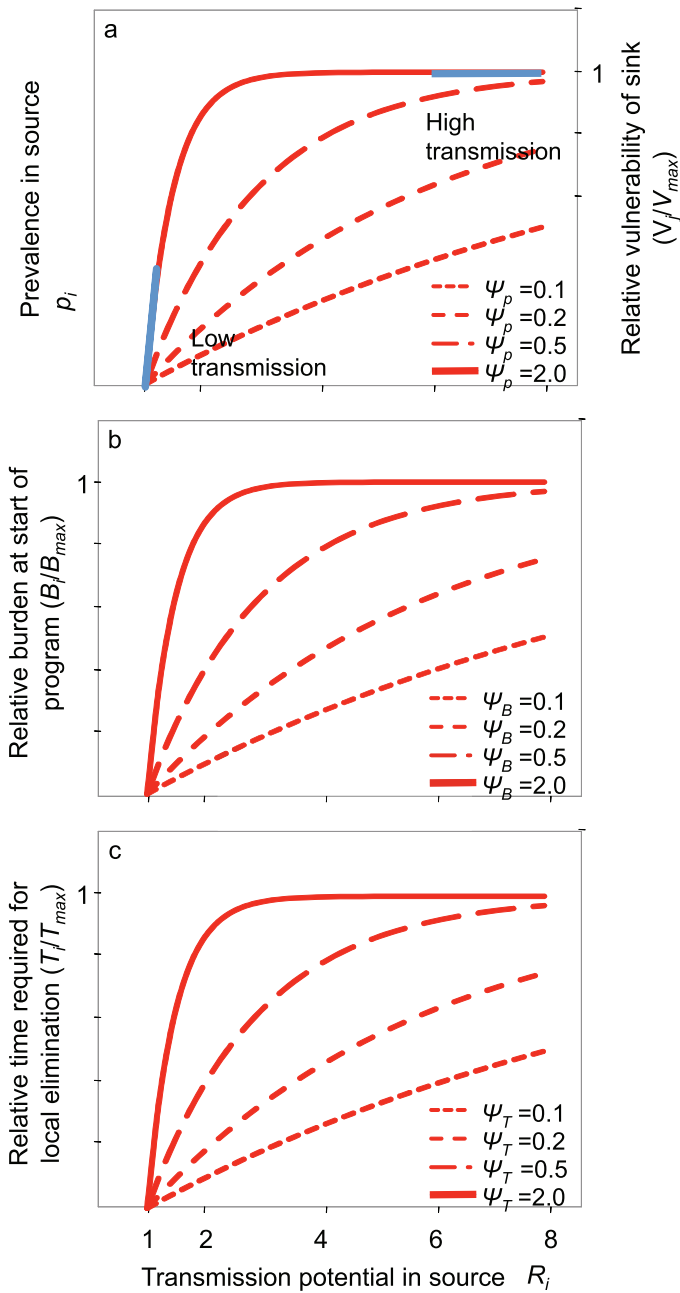
Fig. 2. Proposed relationships of equilibrium prevalence, vulnerability, disease burden and duration of intensified measures needed to achieve elimination with transmission potential. a) Prevalence and vulnerability: vulnerability is proportional to prevalence in the complementary population so they are represented with the same curves. The blue lines correspond to the extreme cases analysed in detail. b) Disease burden. c) Duration of elimination program. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

GMEP strategy (Pampana, 1963), there is always a health benefit in following strategy A (from Eq. (11) because $B_1 > B_2$ — details of this analysis are shown in Appendix B). The health effects are maximised if the program starts in the highest burden population and proceeds in order of decreasing burden. This result generalises to multiple populations, and follows from the assumption that transmission potential is an external, unchanging, factor. This contrasts with models for control of epidemics of directly transmitted disease (Rowthorn et al., 2009) where the default assumption of equal transmission potential in different areas leads to an inverse relationship between susceptibility and initial level of transmission,

so that prioritisation of the initially lower transmission area minimises overall burden.

If the duration of intensified measures varies with transmission potential, $T_1 > T_2$, then Eq. (11) implies that the preferred strategy for minimising the overall burden depends on how the ratio of time to elimination relates to the ratio of pre-elimination burden. Strategy B has lowest overall burden if $T_1/T_2 > B_1/B_2$, while strategy A has lowest overall burden if $B_1/B_2 > T_1/T_2$. Details of this analysis are shown in Appendix C. In summary, to minimise burden over the total duration of the elimination program, if area 1 has a higher burden than area 2, then it is rational to prioritise elimination efforts in area 1, unless the time it would take to achieve elimination in area 1 is *disproportionately* long compared to the time required in area 2.

While at least the relative values of B_i should be known, this is rarely the case for T_i. We therefore represent the general relationships between burden and duration of intensified control with transmission potential, by the models of Eqs. (5) and (8), respectively. Substitution of these formulae into Eq. (11) indicates that strategy A minimises overall burden if and only if $\psi_T > \psi_B$, that is, if the time to elimination saturates faster with the reproduction number than does the initial burden. Details of this analysis are shown in Appendix D.

3.2. Costs

We assume that the overall costs of intensified measures of either strategy are incurred irrespective of the ordering of areas and duration of each phase (implying that the time scales for elimination occur faster than any secular changes such as socio-economic development that would reduce transmission potential over time). Ignoring discounting, these costs therefore cancel in the formulae for the differences in costs between strategies, and the only differences depend on the surveillance costs (see Table 1 for notation),

$$\Delta C = C_A - C_B$$
$$= S_A - S_B,$$
(12)

where strategy A is cost-saving if $\Delta C < 0$.

The total surveillance costs in each area depend on population size, vulnerability, receptivity, and the duration of the program and can be treated as the sum of three components: the per capita base costs of setting up the system of surveillance S_0 (which we assume to be similar across the two zones), the marginal cost of identifying and managing all primary (imported) cases, (equal to the product of the cost per imported case, β, and the importation rate, V), and the costs of responding to any secondary cases (introduced or indigenous), equal to the cost per secondary case, γ, multiplied by the importation rate, V, and the average number of secondary cases, R, per imported case, so that the cost of surveillance with strategy A is

$$S_A = S_0 + \beta V_1 T_2 + \gamma V_1 T_2 R_1$$
$$= S_0 + V_1 T_2 (\beta + \gamma R_1),$$
(13)

and with strategy B is

$$S_B = S_0 + V_2 T_1 (\beta + \gamma R_2).$$
(14)

This leads to incremental cost of strategy A from Eq. (12) of

$$\Delta C = V_1 T_2 (\beta + \gamma R_1) - V_2 T_1 (\beta + \gamma R_2),$$
(15)

which can be expressed as a function of the reproductive numbers,

$$\Delta C = f_V(R_2) f_T(R_2)(\beta + \gamma R_1) - f_V(R_1) f_T(R_1)(\beta + \gamma R_2).$$
(16)

The condition for strategy A being cost saving is then

$$f_V(R_2) f_T(R_2)(\beta + \gamma R_1) < f_V(R_1) f_T(R_1)(\beta + \gamma R_2).$$
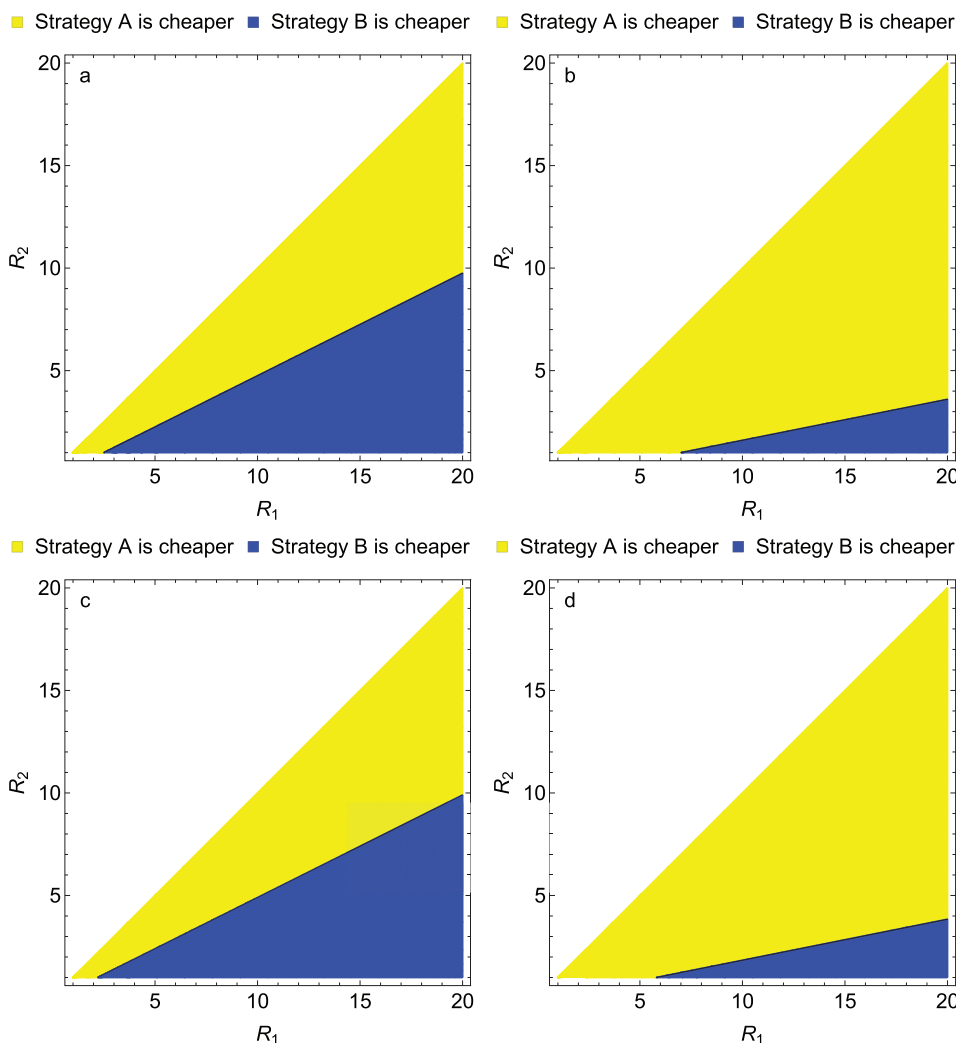(17)

Fig. 3. Indifference curves indicating transmission potentials at which the costs are equivalent when time to elimination differs between areas ($T_1 > T_2$) but vulnerability is saturated ($V_1 = V_2$). The yellow shading indicates the parameter space for which strategy A is cheaper. a) The surveillance costs of a secondary case are twice that of an imported case, $\gamma = 2\beta$. The required duration of the elimination program is twice as long in the higher transmission area, $T_1 = 2T_2$. b) $\gamma = 2\beta$. $T_1 = 5T_2$. c) $\gamma = 5\beta$. $T_1 = 2T_2$. d) $\gamma = 5\beta$. $T_1 = 5T_2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Using this inequality, we consider which strategy minimises overall costs for different combinations of assumptions about T_1, T_2, V_1, and V_2 as indicated in Table 3.

3.2.1. Low transmission

In low transmission areas, where vulnerability is proportional to transmission potential, $V_i = \eta(R_j - 1)$ for $i \neq j$, when the required duration of intensified measures is equal, $T_1 = T_2 = T$, the incremental cost of strategy A from Eq. (15) is

$$\Delta C = \eta(R_2 - 1)T(\beta + \gamma R_1) - \eta(R_1 - 1)T(\beta + \gamma R_2)$$
$$= \eta T(\beta + \gamma)(R_2 - R_1), \tag{18}$$

which is always negative, implying that strategy A has lowest cost (Table (3)). Details of this analysis are shown in Appendix E.

When the required duration of intensified measures increases with transmission, $T_1 > T_2$, the incremental cost (Eq. (15)) is

$$\Delta C = \eta(R_2 - 1)T_2(\beta + \gamma R_1) - \eta(R_1 - 1)T_1(\beta + \gamma R_2), \tag{19}$$

which is even more strongly negative than for the case where $T_1 = T_2 = T$, implying that strategy A again has lowest cost (Table 3). Details of this analysis are shown in Appendix F.

3.2.2. High transmission

If both areas have sufficiently high transmission, vulnerability saturates, $V_1 = V_2 = V$, because the risk of infection no longer increases with R. If the required duration of intensified measures is equal, $T_1 = T_2 = T$, the incremental cost of strategy A from Eq. (15) is

$$\Delta C = VT(\beta + \gamma R_1) - VT(\beta + \gamma R_2)$$
$$= VT\gamma(R_1 - R_2), \tag{20}$$

which is always positive (in contrast to the analysis of low transmission), implying that strategy B has lowest cost (Table 3). Details of this analysis are shown in Appendix G.

When the required duration increases with transmission, $T_1 > T_2$, the incremental cost of strategy A is

$$\Delta C = VT_2(\beta + \gamma R_1) - VT_1(\beta + \gamma R_2), \tag{21}$$

which is negative (favouring strategy A) if and only if

$$R_2 > \frac{T_2}{T_1}\left(R_1 - \frac{\beta}{\gamma}\left(\frac{T_1 - T_2}{T_2}\right)\right). \tag{22}$$

Details of this analysis are shown in Appendix H. Fig. 3 illustrates some of the parts of the parameter space indicating where
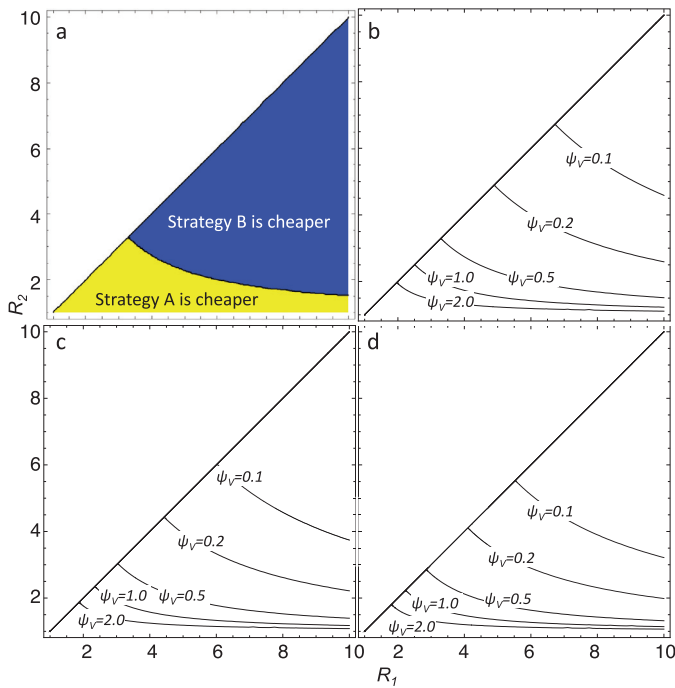
Fig. 4. Indifference curves indicating transmission potentials at which the costs are equivalent for the two strategies assuming equal durations of intensified measures in the two sites, $T_1 = T_2$, and vulnerability follows the general relationship described by (7). a) The R_1 vs. R_2 plane for the vulnerability saturation parameter, $\psi_V = 0.5$ and equal surveillance costs for secondary and imported cases, $\gamma = \beta$. The yellow shading indicates the parameter space for which strategy A is cheaper. b) The location of critical lines for different values of ψ_V with $\gamma = \beta$ (the greater the value of ψ_V, the faster vulnerability saturates with transmission potential). In each case, strategy A is cost saving for values of (R_1, R_2) below the line and strategy B is cost saving for those above the line. c) The location of critical lines for $\gamma = 2\beta$ (costs for secondary cases are twice those of imported cases). d) The location of critical lines for $\gamma = 5\beta$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this condition is satisfied. The parameter values illustrated in Fig. 3(a), in which the surveillance cost per secondary case is twice the cost of a primary case and the time to elimination is twice as high in the higher transmission setting, show a roughly equal division of the reproductive numbers between favouring strategy A and strategy B. Fig. 3(b), in which the time to elimination is five times as high in the higher transmission setting, but with the same surveillance costs as in Fig. 3(a), shows that strategy A is largely preferred to strategy B. Fig. 3(c), in which the surveillance cost per secondary case is five times that of the primary case, but with the same time to elimination as in Fig. 3(a), shows a small increase in situations that favour strategy B when compared to Fig. 3(a) (because strategy A results in a higher ratio of secondary to primary cases than strategy B). Fig. 3(d) shows that when the surveillance cost per secondary case is five times the cost of a primary case and the time to elimination is five times as high in the higher transmission setting, strategy A is largely preferred to strategy B, but there are more settings where strategy B is preferred than in Fig. 3(b). In all four cases, high values of R_2, which are more relevant because vulnerability is more likely to saturate at high transmission potentials, favour strategy A.

3.2.3. Varying transmission with fixed duration of intensified measures

Fig. 4 indicates for which values of R_1 and R_2, the different strategies are to be preferred in terms of costs on the assumption that $T_1 = T_2 = T$ and based on the general relationship of V to R given in Eq. (7). Fig. 4(a) gives the decision rule for one specific set
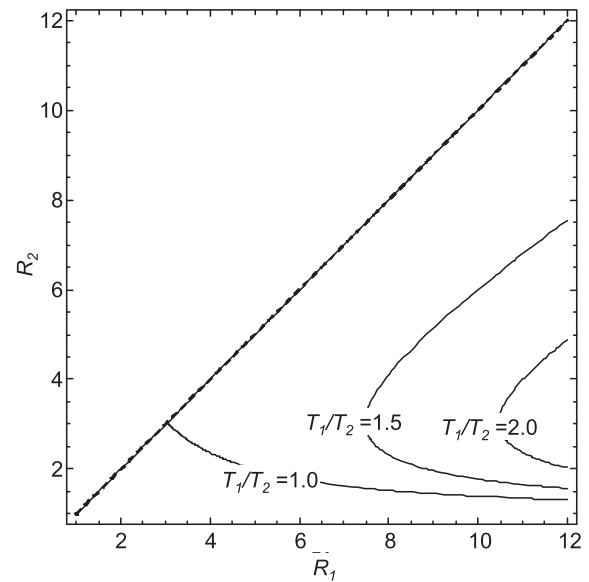

Fig. 5. Indifference curves indicating transmission potentials at which the costs are equivalent for the two strategies assuming the required duration of intensified measures increases with transmission, $T_1 \geq T_2$, and vulnerability follows the general relationship described by (7). The location of critical lines is shown for different values of the ratio of the required durations of intensified measures, T_1/T_2, with the vulnerability saturation parameter, $\psi_V = 0.5$ and equal costs for secondary and imported cases, $\gamma = \beta$. In each case, strategy A is cost saving for values of (R_1, R_2) to the left of the line and strategy B is cost saving for values to the right of the line.

of parameter values. Fig. 4(b)–(d) illustrate the effects of changing the saturation parameter, ψ_V, and the ratio of the surveillance costs of secondary cases to imported cases, $\gamma : \beta$.

In general, at low values of R, strategy A is favoured, and at high values of R, strategy B is favoured. As ψ_V increases, corresponding to a faster saturation of vulnerability with transmission potential, there is an increase in the proportion of settings in which strategy B is cost saving. Increasing the ratio $\gamma : \beta$ makes it more likely that strategy B will be cheaper while decreasing $\gamma : \beta$ makes strategy A more attractive — because the number of secondary cases depends strongly on the receptivity of the sink zone. However these effects are relatively small and the indifference curves are rather insensitive to the ratio of γ to β.

3.2.4. Varying duration of intensified measures

Fig. 5 extends the analysis of costs to the case where the required duration of intensified measures varies with transmission potential with an unspecified function, while vulnerability relates to the transmission potential with the general exponential form in (7). The critical lines in the plot of R_2 against R_1 delimiting parameter sets where strategy A is cost saving (to the left of the critical line) move to the right of the plot as T_1/T_2 increases, indicating that if the required duration of the elimination program is greater in the high transmission area, strategy A is cost-saving in more combinations of transmission settings.

If the relationship between the duration of intensified measures and the transmission potential follows the general relationship given in (8), then the strategy that minimises costs can be established by substituting the formulae for $f_V(R_i)$ and $f_T(R_i)$ from Eqs. (7) and (8) respectively, into Eq. (17) and simplifying. This gives the general condition for strategy A to be cost saving if

$$\frac{(1 - \exp(-\psi_V(R_1 - 1)))(1 - \exp(-\psi_T(R_1 - 1)))(\beta + \gamma R_2)}{(1 - \exp(-\psi_V(R_2 - 1)))(1 - \exp(-\psi_T(R_2 - 1)))(\beta + \gamma R_1)}$$
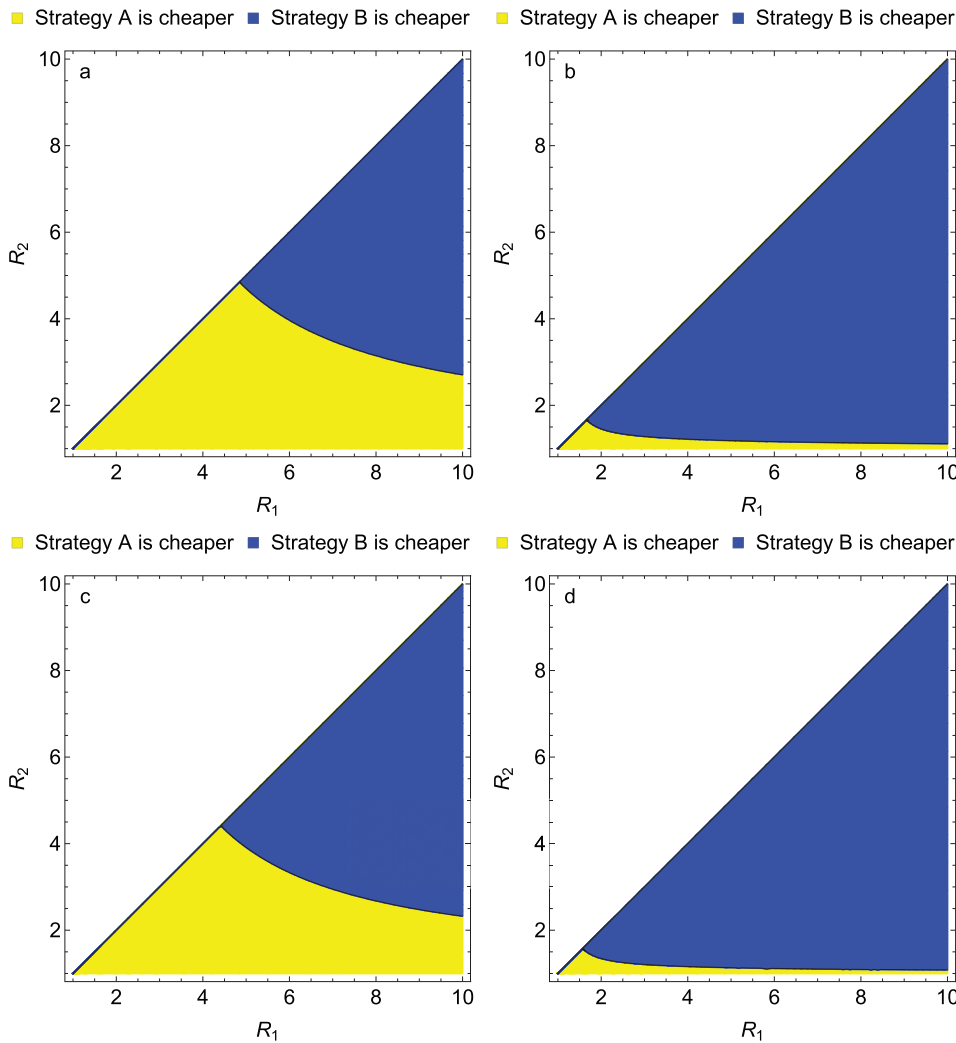$$> 1, \tag{23}$$

Fig. 6. Indifference curves indicating transmission potentials at which the costs are equivalent for the two strategies assuming the required duration of intensified measures, *T*, follows the general relationship given in (8) and vulnerability, *V*, follows the general relationship described by (7). The yellow shading indicates the parameter space for which strategy A is cheaper. a) The costs of secondary cases is equal to that of imported cases, $\gamma = \beta$. The vulnerability saturation parameter $\psi_V = 0.5$. The saturation parameter for the required duration of intensified measures $\psi_T = 0.5$. b) $\gamma = \beta$. $\psi_V = 5$. $\psi_T = 5$. c) $\gamma = 10\beta$ (secondary cases cost ten times as much as imported cases). $\psi_V = 0.5$. $\psi_T = 0.5$. d) $\gamma = 10\beta$. $\psi_V = 5$. $\psi_T = 5$. Higher values for ψ_V (respectively ψ_T) correspond to a faster saturation of vulnerability (respectively duration of intensified measures) with transmission. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as shown in more detail in Appendix I. Unlike the corresponding analysis of burden, this does not simplify to any simple practical rule, but general trends can be extracted from numerical examples (Fig. 6). If the required duration to interrupt transmission saturates quickly with transmission potential, targeting the low transmission site first (strategy B) is favoured (Fig. 6(b) and (d)) compared to Fig. 6(a) and (c)). Similarly, if vulnerability saturates quickly with transmission potential, then targeting the low transmission site is favoured (Fig. 6(b) and (d)) compared to Fig. 6(a) and (c)). This is because as vulnerability and duration of intensified measures saturate with transmission potential, the two settings approach the situation with equal durations of intensified measures and equal vulnerability considered above (Section 3.2.2) where strategy B is favoured. Also, if the relative health systems costs of secondary cases (to imported cases) is high, targeting the low transmission site first (strategy B) is favoured (Fig. 6(c) and (d)) compared to Fig. 6(a) and (b)).

3.3. Net monetary benefit

Where the objectives of minimising disease burden and minimising costs agree on which is the best strategy, this strategy clearly dominates. In the general case, it is possible to combine information on costs and health benefits by assigning a threshold value (λ) corresponding to the amount of money that the program is willing to pay for averting one unit of burden of disease. The net monetary benefit of strategy A, in the absence of discounting is then

$$M = \lambda \Delta E - \Delta C. \tag{24}$$

This provides an unambiguous decision rule, where the strategy A is favoured if $M > 0$. Substituting Eqs. (11) and (12) in (24) provides

$$M = \lambda(B_1 T_2 - B_2 T_1) - V_1 T_2(\beta + \gamma R_1) + V_2 T_1(\beta + \gamma R_2). \tag{25}$$

The condition for this to be positive (i.e. for strategy A to be preferred within the net benefit framework, with initial intensification in the higher burden area) is then

$$\lambda(B_1 T_2 - B_2 T_1) > (V_1 T_2(\beta + \gamma R_1) - V_2 T_1(\beta + \gamma R_2)). \tag{26}$$

In low transmission settings with equal durations of elimination programs, there is no trade-off between reducing burden and costs, since both are minimised by targeting the high transmission site

Table 4

Factors favouring Strategy A, i.e., elimination first in the area with higher transmission instead of the targeting the lower transmission area first.

	Priority is to minimise disease burden over time		Priority is to minimise costs over time
1.	Required duration of intensified control measures until transmission is interrupted is independent of transmission potential, R.	1.	Required duration of intensified control measures until transmission is interrupted is strongly correlated with R.
2.	Annual burden is strongly correlated with R.	2.	Vulnerability is strongly correlated with R.
		3.	Low transmission levels in both sites.
		4.	Low cost of secondary cases relative to primary cases.

first (strategy A), so this strategy dominates. For all other situations the specific values or ranking of the model parameters matter.

The simplest such situation arises with the high transmission case with $T_1 = T_2 = T$, where the health benefits are greater with strategy A, but the costs are also higher, so the preferred strategy depends on trading off health benefits and costs. From Eqs. (20) and (25),

$$M = (\lambda(B_1 - B_2) - \gamma V(R_1 - R_2))T, \tag{27}$$

leading to the conclusion that strategy B should be favoured if

$$\frac{(B_1 - B_2)}{(R_1 - R_2)} < \frac{\gamma V}{\lambda}. \tag{28}$$

If both populations have high transmission initially then, in the limit, burden is independent of transmission intensity, $B_1 \approx B_2$ (Fig. 2), and the left hand side of the equation is close to zero. Since the right hand side is always positive, this implies that strategy B is likely to be the most attractive, i.e., the relatively low transmission population should be addressed first. This is because strategy B reduces the number of secondary cases during the second phase of the program (because the imported cases are in the low transmission area).

3.4. Summary of the results

When all areas involved have low transmission initially, so that burden and vulnerability are proportional to R, and the duration of the intensified measures is independent of initial transmission intensity, the general rule of targeting the higher transmission areas first applies to elimination programs aiming to minimise either total disease burden or costs.

If transmission is initially high in some or all of the zones, or the duration of the intensified measures depends on R, then the analysis becomes more complex, and neither strategy dominates. Many specific parameter sets lead to trade-offs between minimising burden and costs, but overall the results can be summarised by the general statements in Table 4.

4. Discussion and conclusions

Currently there are large investments and efforts to further control malaria burden and move towards elimination in many geographic areas. The strategies to achieve elimination will depend on intense control, focussed efforts to halt and prevent transmission, and improved surveillance and response in existing health systems. Economic analysis can be useful to guide strategic thinking concerning ordering of zones targeted for elimination when the zones have different levels of disease and transmission potential.

In this paper we developed and analysed a mathematical model to understand the implications of prioritisation of different strategies of elimination of neighbouring zones with malaria. The results, summarized in Table 4, indicate general principles that may be applicable at different spatial scales across the whole range of malaria transmission intensities. If all areas involved have low transmission initially, and the duration of intensified control needed to interrupt transmission is the same in all areas, the area

with the highest initial level of transmission should be prioritised to minimise total disease burden and/or costs. The mathematics is scale-free and applies whether the areas are small hamlets or islands, districts or countries. The key results extend to any number of different areas, but the models would need significant extension to allow for arbitrary patterns of case exportation and importation.

When some of the areas have high transmission, or the equal duration assumption does not hold, the decision is more complicated. In such cases the strength of the relationship between T (time to reduce incidence to negligible levels) and R (transmission potential) affects the ordering preference based on burden or on costs in converse ways (Table 4), potentially leading to the need to trade-off these considerations against each other.

As a general rule, we allowed the prevalence (and consequently vulnerability, V), the burden, B, and the time to elimination, T to saturate exponentially with the transmission potential, R. This is a reasonable approximation for prevalence and burden, as seen for example in bifurcation diagrams for standard susceptible-infectious-susceptible (SIS) ordinary differential equation (ODE) models. We additionally assumed that the saturation parameters for vulnerability and burden are the same for both areas. This is also reasonable but there may be small differences in the saturation parameters if the degree of seasonality is substantially different between the two areas (for example if one area is highly seasonal while the other is relatively flat) (Stuckey et al., 2014).

However, assuming an exponentially saturating relationship between T and R remains a simplifying assumption, considering the heterogeneity of contexts and strategies in elimination programs. We use an exponential function because the relationship should be monotonic and if elimination is feasible, it is likely that the time to elimination will saturate as a function of transmission potential. We expect that T is considerably longer than the time required for residual infections to clear once transmission has been interrupted (Smith and Hay, 2009). In a stable environment, the maximal impact on transmission of a malaria intervention program occurs shortly after scale-up (Smith and Schapira, 2012) but the trajectories of most recent elimination programs do not correspond to the GMEP model with fixed T (Macdonald and Göckel, 1964; Pampana, 1963). Many factors work to make T longer and more irregular in contemporary programs: tropical areas in most cases have entomological characteristics slowing or blunting the effectiveness of available vector control measures; the populations needing coverage may be large; human ecology, for example population movement, may constrain intervention effectiveness; and long lasting insecticidal nets, if used as the main transmission-reducing intervention, may achieve more gradual scale-up of impact than indoor residual spraying. Furthermore, health system heterogeneities may greatly increase the time needed to interrupt transmission, which depends not only on vector control, but also on surveillance as part of the intensified measures package. Accounting for the uncertainty in the T may be in incorporated into program planning but would require substantially more complicated stochastic models. Nonetheless, given the profound influence of variations in T, demonstrated here, it would be rational for elimination programs to assess T as carefully as possible in different zones, based on

experience and knowledge about determinants of intervention effectiveness, to better clarify the relationship between T and R.

Other factors, especially costs, also clearly depend on local information. When initial transmission and prevalence are high, surveillance costs may be minimised if the lower reproduction number zone is targeted first, since vulnerability saturates with level of transmission so that introduced cases are equally frequent whichever ordering is chosen, while secondary infection rates depend on the local reproduction number. The ratio between surveillance costs for primary and secondary cases is then relevant to the choice of strategy, and a field-determined value of this would be needed for a full quantitative analysis of any specific situation. Population sizes and per capita costs may also vary between locations. However, to identify important principles, we also made the simplifying assumption that the two zones have similar population sizes and the same base costs of setting up the surveillance system.

In reality, there is also likely to be temporal overlap between intensification of control in different zones. Temporal overlap reduces the marginal cost of surveillance proportionately to the reduction in overall duration, but this should not change the preferred ordering, because the mathematics does not change if the import/export of cases during the overlap period is small. In the limiting case, when transmission is interrupted everywhere in the world (and complete eradication is achieved) simultaneously, the surveillance period post-elimination and its costs are zero.

The most critical assumption here is that elimination is feasible. There is always more uncertainty about a strategy with a specific endpoint (elimination) than to a less specific one (control), and there is a need for decision rules allowing for this uncertainty (which go beyond the scope of this paper). Spatially progressive elimination can be an appropriate eradication strategy for infections with patchy distributions and long generation times (like onchocerciasis) or where intervention strategies lead to long-term protection of humans from infection (as with smallpox vaccination); but it is unclear under which circumstances this will work for malaria, where neither of these conditions hold. In most endemic settings, the spatial distribution of malaria is widespread, malaria parasites have a short generation time (about two months in the case of *Plasmodium falciparum* (Huber et al., 2016) and the effects of most interventions are of a relatively short duration, so transmission will recover rapidly if intervention deployment ceases.

In many settings, the rationale for spatially progressive elimination therefore depends critically on the assumption that if transmission is interrupted, it is unlikely to re-establish itself (i.e., the malaria free state is asymptotically stable — or 'sticky' (Smith et al., 2013). This can be the case if an effective surveillance system is in place and infections become more detectable as transmission is brought down because of loss of clinical immunity, and/or the vectorial capacity is low from the beginning. It may also be unlikely for transmission to resume if malaria elimination is accompanied by changes in housing or lifestyles that permanently reduce vectorial capacity. If such changes do not occur, then the risk of reintroduction can be contained by indefinite maintenance of proactive vector control. In most situations, the costs of this would be unacceptably high, so 'stickiness' depends on changes in the human and vector ecology that are contingent on vector biology, economic development and equity, and hence outside the control of the health system. However, analysis of countries that are either on the path to elimination or have already achieved elimination has shown that while resurgence is likely to occur in countries that have not achieved elimination, it is unlikely to occur after elimination has been reached (Cohen et al., 2012; Feachem et al., 2009; Smith et al., 2013). Smith et al. (2013) state:

What is surprising, however, is that of the 50 elimination programmes identified as successful, only four (8%) were found to have experienced resurgence, despite continued importation over many years. Moreover of these four, two eliminated malaria and are malaria-free once again. These 50 countries still have competent vectors and few have ongoing transmission-lowering activities.

Therefore, it is likely that surveillance programs will be able to maintain elimination after it has been achieved.

Rational decision-making about prioritising low or high transmission areas in an elimination programme may be straightforward under certain conditions: if transmission is initially relatively low in all areas, if the disease burden is strongly correlated with level of transmission, while expected time to elimination is not, then the high transmission area should be prioritised. Otherwise, strategic planning would benefit from a detailed assessment of patterns of vulnerability and surveillance costs that relaxes the assumptions made in this analysis.

Acknowledgements

Appendix A. Preliminary analysis

Lemma 1. *Let*

$$f(x) = (e^{ax} - 1)(e^{a/x} - 1) \tag{A.1}$$

with $a > 0$ and domain $\mathcal{D} = \{x \in \mathbb{R} | x > 0\}$. Then $f(x)$ has a global minimum at 1, that is,

$$f(1) \leq f(x) \quad \forall x \in \mathcal{D}. \tag{A.2}$$

Proof. $f(x) = (e^{ax} - 1)(e^{a/x} - 1)$ has a global minimum at $x = 1$ if and only if

$$f(x) - 1 = (e^{ax} - 1)(e^{a/x} - 1) - 1 \tag{A.3a}$$

$$= e^{ax + a/x} - e^{ax} - e^{a/x}, \tag{A.3b}$$

has a global minimum at $x = 1$. For notational convenience we redefine

$$f(x) = e^{ax + a/x} - e^{ax} - e^{a/x}. \tag{A.4}$$

To prove that $f(x)$ has a global minimum at $x = 1$, we will show that:

(i) $f'(1) = 0$
(ii) $f'(x) > 0$ for $x > 1$
(iii) $f'(x) < 0$ for $0 < x < 1$,

where $f'(x)$ denotes the derivative of $f(x)$ with respect to x,

$$f'(x) = \left(a - \frac{a}{x^2}\right) e^{ax + a/x} - a e^{ax} + \frac{a}{x^2} e^{a/x}. \tag{A.5}$$

Substituting $x = 1$ shows that $f'(1) = 0$ so condition (i) is satisfied. To prove conditions (ii) and (iii), we use the change of variables,

$$u = \ln(x), \tag{A.6}$$

with $u \in \mathbb{R}$ and $x = e^u$, to define

$$g(u) = f(e^u) = e^{a(e^u + e^{-u})} - e^{ae^u} - e^{ae^{-u}}. \tag{A.7}$$

Since u is a monotonically increasing function of x, with $x = 1$ cor-

responding to $u = 0$, statement (ii) is equivalent to showing that $g'(u) > 0$ for $u > 0$ and statement (iii) is equivalent to showing that $g'(u) < 0$ for $u < 0$, where $g'(u)$ denotes the derivative of $g(u)$ with respect to u,

$$g'(u) = ae^u\left(e^{a(e^u+e^{-u})} - e^{ae^u}\right) - ae^{-u}\left(e^{a(e^u+e^{-u})} - e^{ae^{-u}}\right). \tag{A.8}$$

We introduce the function

$$h(u) = \frac{ae^u\left(e^{a(e^u+e^{-u})} - e^{ae^u}\right)}{ae^{-u}\left(e^{a(e^u+e^{-u})} - e^{ae^{-u}}\right)}$$

$$= e^{2u}\frac{e^{ae^u}\left(e^{ae^{-u}} - 1\right)}{e^{ae^{-u}}\left(e^{ae^u} - 1\right)}, \tag{A.9}$$

so that $h(u) > 1$ is equivalent to $g'(u) > 0$ and $h(u) < 1$ is equivalent to $g'(u) < 0$.

We note here that

$$e^\xi - 1 = \xi e^{\xi - c(\xi)}, \tag{A.10}$$

for

$$c(\xi) = \ln(\xi) - \ln\left(1 - e^{-\xi}\right). \tag{A.11}$$

Then using the above substitution (A.11), with $\xi = ae^{-u}$ in the numerator of $h(u)$ and $\xi = ae^u$ in the denominator of $h(u)$, and simplifying,

$$h(u) = \frac{e^{-c(ae^{-u})}}{e^{-c(ae^u)}}$$

$$= e^{c(ae^u) - c(ae^{-u})}. \tag{A.12}$$

Introducing the function

$$j(u) = c(ae^u) - c(ae^{-u}), \tag{A.13}$$

we see that if $j(u) = 0$, then $h(u) = 1$; if $j(u) > 0$, then $h(u) > 1$; and if $j(u) < 0$, then $h(u) < 1$. We expand

$$j(u) = 2u + \ln\left(1 - e^{-ae^{-u}}\right) - \ln\left(1 - e^{-ae^u}\right), \tag{A.14}$$

see that $j(0) = 0$, and differentiate

$$j'(u) = 2 - \frac{v_- e^{-v_-}}{1 - e^{-v_-}} - \frac{v_+ e^{-v_+}}{1 - e^{-v_+}}, \tag{A.15}$$

where

$$v_- = ae^{-u}, \tag{A.16a}$$

$$v_+ = ae^u. \tag{A.16b}$$

Noting that, $\xi < e^\xi - 1$, for $\xi \neq 0$,

$$j'(u) > 2 - \frac{(e^{v_-} - 1)e^{-v_-}}{1 - e^{-v_-}} - \frac{(e^{v_+} - 1)e^{-v_+}}{1 - e^{-v_+}} \tag{A.17}$$

$$= 0. \tag{A.18}$$

Since $j(0) = 0$ and $j(u)$ is monotonically increasing for all $u \in \mathbb{R}$, for $u < 0$, $j(u) < 0$, and for $u > 0$, $j(u) > 0$. Therefore, for $u > 0$, $h(u) > 1$, $g'(u) > 0$, and correspondingly, $f'(x) > 0$ for $x > 1$ so condition (ii) is satisfied. Similarly, for $u < 0$, $h(u) < 1$, $g'(u) < 0$, and correspondingly, $f'(x) < 0$ for $0 < x < 1$ so condition (iii) is satisfied. □

Lemma 2. For all $\varphi > 0$, $\theta > 0$, $x > 0$ and $y > 0$,

$$e^{\varphi x + \theta y} + e^{-\varphi y} + e^{-\theta x} > e^{-(\varphi y + \theta x)} + e^{\varphi x} + e^{\theta y}. \tag{A.19}$$

Proof. We rearrange the inequality to be proved (A.19) as

$$e^{\varphi x + \theta y} - e^{\varphi x} - e^{\theta y} > e^{-\varphi y - \theta x} - e^{-\varphi y} - e^{-\theta x}. \tag{A.20}$$

We introduce the change of variables,

$$s = \varphi y + \theta x, \tag{A.21a}$$

$$\varepsilon = \frac{\varphi y}{\varphi y + \theta x}, \tag{A.21b}$$

$$t = \frac{\varphi x}{\theta y}, \tag{A.21c}$$

with $s > 0$, $t > 0$, and $0 < \varepsilon < 1$. Then

$$\varphi x = \sqrt{\varepsilon(1 - \varepsilon)}s\sqrt{t}, \tag{A.22a}$$

$$\varphi y = \varepsilon s, \tag{A.22b}$$

$$\theta x = (1 - \varepsilon)s, \tag{A.22c}$$

$$\theta y = \sqrt{\varepsilon(1 - \varepsilon)}s\frac{1}{\sqrt{t}}. \tag{A.22d}$$

Using the change of variables, (A.21), the inequality to be proved (A.20) becomes

$$e^{\sqrt{\varepsilon(1-\varepsilon)}s(\sqrt{t}+1/\sqrt{t})} - e^{\sqrt{\varepsilon(1-\varepsilon)}s\sqrt{t}} - e^{\sqrt{\varepsilon(1-\varepsilon)}s(1/\sqrt{t})}$$
$$> e^{-s} - e^{-\varepsilon s} - e^{-(1-\varepsilon)s}, \tag{A.23}$$

which is equivalent to

$$\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s\sqrt{t}} - 1\right)\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s(1/\sqrt{t})} - 1\right)$$
$$> \left(e^{-\varepsilon s} - 1\right)\left(e^{-(1-\varepsilon)s} - 1\right). \tag{A.24}$$

Applying Lemma 1 to the left hand side of (A.24) with $a = \sqrt{\varepsilon(1-\varepsilon)}s$ and $x = \sqrt{t}$, implies

$$\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s} - 1\right)^2 \leq \left(e^{\sqrt{\varepsilon(1-\varepsilon)}s\sqrt{t}} - 1\right)\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s(1/\sqrt{t})} - 1\right). \tag{A.25}$$

Therefore to prove inequality (A.24), it is sufficient to show that

$$\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s} - 1\right)^2 > \left(e^{-\varepsilon s} - 1\right)\left(e^{-(1-\varepsilon)s} - 1\right), \tag{A.26}$$

which is equivalent to proving

$$\frac{\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s} - 1\right)^2}{\left(e^{-\varepsilon s} - 1\right)\left(e^{-(1-\varepsilon)s} - 1\right)} > 1. \tag{A.27}$$

Using the inequalities

$$e^x - 1 > x + (1/2)x^2 \quad \text{for } x > 0, \tag{A.28a}$$

$$e^x - 1 < x + (1/2)x^2 \quad \text{for } x < 0, \tag{A.28b}$$

we see that

$$\frac{\left(e^{\sqrt{\varepsilon(1-\varepsilon)}s} - 1\right)^2}{\left(e^{-\varepsilon s} - 1\right)\left(e^{-(1-\varepsilon)s} - 1\right)}$$

$$> \frac{\left(\sqrt{\varepsilon(1-\varepsilon)}s + (1/2)\varepsilon(1-\varepsilon)s^2\right)^2}{\left(-\varepsilon s + (1/2)\varepsilon^2 s^2\right)\left(-(1-\varepsilon)s + (1/2)(1-\varepsilon)^2 s^2\right)}$$

$$= \frac{\varepsilon(1-\varepsilon)s^2 + (1/4)\varepsilon^2(1-\varepsilon)^2 s^4 + (\varepsilon(1-\varepsilon))^{3/2}s^3}{\varepsilon(1-\varepsilon)s^2 + (1/4)\varepsilon^2(1-\varepsilon)^2 s^4 - (1/2)\varepsilon(1-\varepsilon)s^3}$$

$$> 1. \tag{A.29}$$

□

Appendix B. Analysis 1: Equal time and higher burden

The main assumptions are:

$$T_1 = T_2, \tag{B.1a}$$

$$B_1 > B_2. \tag{B.1b}$$

The reduction in burden due to targeting site 1 first is

$$\begin{aligned}
\Delta E &= B_1 T_2 - B_2 T_1 \\
&= (B_1 - B_2) T \\
&> 0.
\end{aligned} \tag{B.2}$$

Therefore, Strategy A always reduces burden.

Appendix C. Analysis 2: Higher time and higher burden

The main assumptions are:

$$T_1 > T_2, \tag{C.1a}$$

$$B_1 > B_2. \tag{C.1b}$$

The reduction in burden due to targeting site 1 first is

$$\Delta E = B_1 T_2 - B_2 T_1. \tag{C.2}$$

Therefore, Strategy A reduces burden if and only if

$$B_1 T_2 > B_2 T_1, \tag{C.3}$$

or equivalently,

$$\frac{B_1}{B_2} > \frac{T_1}{T_2}. \tag{C.4}$$

Appendix D. Analysis 3: Saturating time and saturating burden

The main assumptions are:

$$T_i = T_{\max}\left(1 - e^{-\psi_T(R_i - 1)}\right) \quad \text{for} \quad i \in \{1, 2\}, \tag{D.1a}$$

$$B_i = B_{\max}\left(1 - e^{-\psi_B(R_i - 1)}\right) \quad \text{for} \quad i \in \{1, 2\}. \tag{D.1b}$$

The reduction in burden due to targeting site 1 first is

$$\begin{aligned}
\Delta E &= B_1 T_2 - B_2 T_1 \\
&= B_{\max} T_{\max}\Big[\left(1 - e^{-\psi_B(R_1 - 1)}\right)\left(1 - e^{-\psi_T(R_2 - 1)}\right) \\
&\quad - \left(1 - e^{-\psi_B(R_2 - 1)}\right)\left(1 - e^{-\psi_T(R_1 - 1)}\right)\Big].
\end{aligned} \tag{D.2}$$

We first assume $\psi_B > \psi_T$, and use the change of variables,

$$y = R_2 - 1, \tag{D.3a}$$

$$x = R_1 - R_2, \tag{D.3b}$$

$$\varphi = \psi_T, \tag{D.3c}$$

$$\theta = \psi_B - \psi_T, \tag{D.3d}$$

with $x > 0$ (since $R_1 > R_2$), $y > 0$, $\varphi > 0$, and $\theta > 0$, to rewrite (D.2) as

$$\begin{aligned}
\Delta E &= B_{\max} T_{\max}\Big[\left(1 - e^{-(\varphi + \theta)(x + y)}\right)\left(1 - e^{-\varphi y}\right) \\
&\quad - \left(1 - e^{-(\varphi + \theta)y}\right)\left(1 - e^{-\varphi(x + y)}\right)\Big].
\end{aligned} \tag{D.4}$$

We rewrite (D.4) after some algebraic manipulations as

$$\begin{aligned}
\Delta E &= B_{\max} T_{\max} e^{-(\varphi x + \varphi y + \theta y)}\Big[\left(e^{-(\varphi y + \theta x)} + e^{\varphi x} + e^{\theta y}\right) \\
&\quad - \left(e^{\varphi x + \theta y} + e^{-\varphi y} + e^{-\theta x}\right)\Big].
\end{aligned} \tag{D.5}$$

Applying Lemma 2 (see Appendix A) to (D.5) we see, that for $\psi_B > \psi_T$,

$$\Delta E < 0, \tag{D.6}$$

so strategy B reduces burden.

A similar calculation for $\psi_T > \psi_B$, with the change of variables,

$$y = R_2 - 1, \tag{D.7a}$$

$$x = R_1 - R_2, \tag{D.7b}$$

$$\zeta = \psi_B, \tag{D.7c}$$

$$\nu = \psi_T - \psi_B, \tag{D.7d}$$

with $x > 0$, $y > 0$, $\zeta > 0$, and $\nu > 0$, shows that

$$\begin{aligned}
\Delta E &= B_{\max} T_{\max} e^{-(\zeta x + \zeta y + \nu y)}\Big[\left(e^{\zeta x + \nu y} + e^{-\zeta y} + e^{-\nu x}\right) \\
&\quad - \left(e^{-(\zeta y + \nu x)} + e^{\zeta x} + e^{\nu y}\right)\Big],
\end{aligned} \tag{D.8}$$

where by Lemma 2,

$$\Delta E > 0, \tag{D.9}$$

so strategy A reduces burden.

Appendix E. Analysis 4: Equal time and linear vulnerability

The main assumptions are:

$$T_1 = T_2, \tag{E.1a}$$

$$V_i = \eta(R_j - 1) \quad \text{for} \quad i, j \in \{1, 2\} \text{ with } i \neq j. \tag{E.1b}$$

The additional costs of strategy A are

$$\begin{aligned}
\Delta C &= V_1 T_2 (\beta + \gamma R_1) - V_2 T_1 (\beta + \gamma R_2) \\
&= \eta(R_2 - 1)T(\beta + \gamma R_1) - \eta(R_1 - 1)T(\beta + \gamma R_2) \\
&= \eta T (\beta + \gamma)(R_2 - R_1) \\
&< 0.
\end{aligned} \tag{E.2}$$

Therefore, Strategy A is always cheaper.

Appendix F. Analysis 5: Higher time and linear vulnerability

The main assumptions are:

$$T_1 > T_2, \tag{F.1a}$$

$$V_i = \eta(R_j - 1) \quad \text{for} \quad i, j \in \{1, 2\} \text{ with } i \neq j. \tag{F.1b}$$

The additional costs of strategy A are

$$\begin{aligned}
\Delta C &= V_1 T_2 (\beta + \gamma R_1) - V_2 T_1 (\beta + \gamma R_2) \\
&= \eta(R_2 - 1)T_2(\beta + \gamma R_1) - \eta(R_1 - 1)T_1(\beta + \gamma R_2).
\end{aligned} \tag{F.2}$$

We use the change of variables,

$$T_d = T_1 - T_2, \tag{F.3}$$

with $T_d > 0$, and $T_1 = T_d + T_2$ to rewrite (F.2) as

$$\begin{aligned}
\Delta C &= \eta T_2 (\beta + \gamma)(R_2 - R_1) - \eta(R_1 - 1)T_d(\beta + \gamma R_2) \\
&< 0.
\end{aligned} \tag{F.4}$$

Therefore, Strategy A is always cheaper.

Appendix G. Analysis 6: Equal time and equal vulnerability

The main assumptions are:

$$T_1 = T_2, \tag{G.1a}$$

$$V_1 = V_2. \tag{G.1b}$$

The additional costs of strategy A are

$$\Delta C = V_1 T_2 (\beta + \gamma R_1) - V_2 T_1 (\beta + \gamma R_2)$$
$$= V T \gamma (R_1 - R_2)$$
$$> 0. \tag{G.2}$$

Therefore, Strategy B is always cheaper.

Appendix H. Analysis 7: Higher time and equal vulnerability

The main assumptions are:

$$T_1 > T_2, \tag{H.1a}$$

$$V_1 = V_2. \tag{H.1b}$$

The additional costs of strategy A are

$$\Delta C = V_1 T_2 (\beta + \gamma R_1) - V_2 T_1 (\beta + \gamma R_2)$$
$$= V T_2 (\beta + \gamma R_1) - V T_1 (\beta + \gamma R_2)$$
$$= V \beta (T_2 - T_1) + V \gamma (T_2 R_1 - T_1 R_2). \tag{H.2}$$

The sign of (H.2) depends on parameter values, and we can show with some algebraic manipulations that strategy A is cheaper ($\Delta C < 0$) when

$$R_2 > \frac{T_2}{T_1} \left(R_1 - \frac{\beta}{\gamma} \left(\frac{T_1 - T_2}{T_2} \right) \right). \tag{H.3}$$

Appendix I. Analysis 8: Saturating time and saturating vulnerability

The main assumptions are:

$$T_i = T_{\max} \left(1 - e^{-\psi_T (R_i - 1)} \right) \quad \text{for} \quad i \in \{1, 2\}, \tag{I.1a}$$

$$V_i = V_{\max} \left(1 - e^{-\psi_V (R_j - 1)} \right) \quad \text{for} \quad i, j \in \{1, 2\} \text{ with } i \neq j. \tag{I.1b}$$

The additional costs of strategy A are

$$\Delta C = V_{\max} \left(1 - e^{-\psi_V (R_2 - 1)} \right) T_{\max} \left(1 - e^{-\psi_T (R_2 - 1)} \right) (\beta + \gamma R_1)$$
$$- V_{\max} \left(1 - e^{-\psi_V (R_1 - 1)} \right) T_{\max} \left(1 - e^{-\psi_T (R_1 - 1)} \right) (\beta + \gamma R_2)$$
$$= V_{\max} T_{\max} \beta \left[\left(1 - e^{-\psi_V (R_2 - 1)} \right) \left(1 - e^{-\psi_T (R_2 - 1)} \right) \left(1 + \frac{\gamma}{\beta} R_1 \right) \right.$$
$$\left. - \left(1 - e^{-\psi_V (R_1 - 1)} \right) \left(1 - e^{-\psi_T (R_1 - 1)} \right) \left(1 + \frac{\gamma}{\beta} R_2 \right) \right]. \tag{I.2}$$

We do not have an explicit analytical expression that relates the sign of ΔC to R_1, R_2 and the parameters so we rely on the numerical simulations as shown in the main text.

References

Aylward, R.B., Acharya, A., England, S., Agocs, M., Linkins, J., 2003. Global health goals: lessons from the worldwide effort to eradicate poliomyelitis. Lancet 362 (9387), 909–914.

Chiyaka, C., Tatem, A.J., Cohen, J.M., Gething, P.W., Johnston, G., Gosling, R., Laxminarayan, R., Hay, S.I., Smith, D.L., 2013. The stability of malaria elimination. Science 339 (6122), 909–910.

Cohen, J.M., Smith, D.L., Cotter, C., Ward, A., Yamey, G., Sabot, O.J., Moonen, B., 2012. Malaria resurgence: a systematic review and assessment of its causes. Malar. J. 11, 122.

R.G.A., Phillips, A.A., Targett G.A. (Eds.), Shrinking the malaria map: A prospectus on malaria elimination, The Global Health Group, Global Health Sciences. University of California, San Francisco, San Francisco.

Gosling, R.D., Whittaker., M., Gueye, C.S., Fullman, N., Baquilod, M., Kusriastuti, R., Feachem, R.G., 2012. Malaria elimination gaining ground in the asia pacific. Malar. J. 11, 346.

Gueye, C.S., Newby, G., Hwang, J., Phillips, A.A., Whittaker, M., MacArthur, J.R., Gosling, R.D., Feachem, R.G.A., 2014. The challenge of artemisinin resistance can only be met by eliminating plasmodium falciparum malaria across the Greater Mekong subregion. Malar. J. 13, 286.

Hemami, M.R., Sari, A.A., Raeisi, A., Vatandoost, H., Majdzadeh, R., 2013. Malaria elimination in Iran, importance and challenges. Int. J. Prev. Med. 4 (1), 88–94.

Henderson, D.A., 2009. Smallpox: The Death of a Disease — The Inside Story of Eradicating a Worldwide Killer. Prometheus Books, Amherst, New York, USA.

Huber, J.H., Johnston, G.L., Greenhouse, B., Smith, D.L., Perkins, T.A., 2016. Quantitative, model-based estimates of variability in the generation and serial intervals of plasmodium falciparum malaria. Malar. J. 15, 490.

Kelly, G.C., Tanner, M., Vallely, A., Clements, A., 2012. Malaria elimination: moving forward with spatial decision support systems. Trends Parasitol. 28 (7), 297–304.

Lines, J., Schapira, A., Smith, T., 2008. Tackling malaria today. BMJ 337, a869.

Mabaso, M.L., Kleinschmidt, I., Sharp, B., Smith, T., 2007. El Nino Southern Oscillation (ENSO) and annual malaria incidence in Southern Africa. Trans. R. Soc. Trop. Med. Hyg. 101 (4), 326–330.

Macdonald, G., Göckel, G.W., 1964. The malaria parasite and interruption of transmission. Bull. World Health Organ. 31, 369–387.

Moonen, B., Cohen, J.M., Snow, R.W., Slutsker, L., Drakeley, C., Smith, D.L., Abeyasinghe, R.R., Rodriguez, M.H., Maharaj, R., Tanner, M., Targett, G., 2010. Operational strategies to achieve and maintain malaria elimination. Lancet 376 (9752), 1592–1603.

Pampana, E., 1963. A Textbook of Malaria Eradication. Oxford University Press, London.

Roberts, L., 2010. Elimination meets reality in hispaniola. Science 328 (5980), 850–851.

Ross, R., 1916. An application of the theory of probabilities to the study of a priori pathometry, part i. Proc. R. Soc. London, Ser. A 92, 204–230.

Rowthorn, R.E., Laxminarayan, R., Gilligan, C.A., 2009. Optimal control of epidemics in metapopulations. J. R. Soc. Interface 6 (41), 1135–1144.

Ruiz-Tiben, E., Hopkins, D.R., 2006. Dracunculiasis (guinea worm disease) eradication. Adv. Parasitol. 61, 275–309.

Schapira, A., Zamani, G., 2012. Assignment Report on a Mission to Review the Malaria Programme in Oman: 3–7 March 2012. Technical Report. World Health Organization Regional Office for the Eastern Mediterranean, Cairo.

Shah, N.K., 2010. Assessing strategy and equity in the elimination of malaria. PLoS Med. 7 (8), e1000312.

Smith, D.L., Cohen, J.M., Chiyaka, C., Johnston, G., Gething, P.W., Gosling, R., Buckee, C.O., Laxminarayan, R., Hay, S.I., Tatem, A.J., 2013. A sticky situation: the unexpected stability of malaria elimination. Philos. Trans. R. Soc. London. B, Biol. Sci. 368, 20120145.

Smith, D.L., Cohen, J.M., Moonen, B., Tatem, A.J., Sabot, O.J., Ali, A., Mugheiry, S.M., 2011. Solving the Sisyphean problem of malaria in Zanzibar. Science 332 (6036), 1384–1385.

Smith, D.L., Dushoff, J., Snow, R.W., Hay, S.I., 2005. The entomological inoculation rate and Plasmodium falciparum infection in African children. Nature 438, 492–495.

Smith, D.L., Hay, S.I., 2009. Endemicity response timelines for plasmodium falciparum elimination. Malar. J. 8, 87.

Smith, D.L., Smith, T.A., Hay, S.I., 2009. Measuring malaria for elimination. In: Feachem, R.G.A., Phillips, A.A., Targett, G.A. (Eds.), Shrinking the Malaria Map: A Prospectus on Malaria Elimination. The Global Health Group, San Francisco, pp. 108–126.

Smith, T., Schapira, A., 2012. Reproduction numbers in malaria and their implications. Trends Parasitol. 28 (1), 3–8.

Southern African Development Community, 2013. Five SADC countries rewarded for commitment in eliminating malaria. http://www.sadc.int/news-events/news/five-sadc-countries-rewarded/. Date accessed: 11 June 2018.

Stuckey, E.M., Smith, T.A., Chitnis, N., 2014. Seasonally dependent relationships between indicators of malaria transmission and disease provided by mathematical model simulations. PLoS Comput. Biol. 10 (9), e1003812.

World Health Organization, 2012. Disease surveillance for malaria elimination: operational manual. http://www.who.int/malaria/publications/atoz/9789241503334/en/.

World Health Organization, 2015. Global technical strategy for malaria 2016–2030. http://www.who.int/malaria/publications/atoz/9789241564991/en/.

World Health Organization, 2016. WHO malaria terminology. http://www.who.int/malaria/publications/atoz/malaria-terminology/en/. (updated March 2018).

World Health Organization, 2017. A framework for malaria elimination. http://www.who.int/malaria/publications/atoz/9789241511988/en/.

World Health OrganizationGlobal Malaria Programme, University of California, San Francisco, 2012. Preventing reintroduction in mauritius. Eliminating malaria: case study, 4. World Health Organization, Geneva.