Web Material

Web Appendix 1

Methods

1. Theoretical model

Our initial theoretical analysis is based on final size and outbreak probability calculations rather than simulation, but otherwise the trial population has the same structure as in the simulation model. In particular, we assume that a proportion of communities receive a single disease importation, and any outbreak that arises from an importation runs until there are no longer any infectious individuals.

For a community in which there are no vaccinees, the standard final size equation (1) applies for the cumulative incidence, namely CI solves $\mathbf{CI = 1 - e^{-R_0 CI}}$, when R_0>1. Similarly, the proportion of communities with importations in which an outbreak will occur, x, solves the same equation. For a community in which a proportion *p* of the individuals are vaccinated with vaccine efficacy VE, the equations for the CI among the vaccinated and unvaccinated, CI_V and CI_U respectively, are

$$\mathbf{CI_V = 1 - e^{-R_0(1-VE)[(1-p)CI_U + pCI_V]}},$$

$$\mathbf{CI_u = 1 - e^{-R_0[(1-p)CI_U + pCI_V]}}.$$

The outbreak probability in a community in which a proportion *p* of the individuals are vaccinated, x_V, solves the equation $\mathbf{x_V = 1 - e^{-R_0^V x_V}}$, where $\mathbf{R_0^V = (1 - pVE)R_0}$. Sample size calculations were based on a hazard rate analysis, with vaccine effect estimated in both trial designs as $\boldsymbol{VE = 1 - \frac{\ln(1-CI_V)}{\ln(1-CI_U)}}$ (2). Specifically, number of individuals needed to achieve 90% power to detect vaccine effect was given by $\boldsymbol{S = \frac{(1.96+1.28)^2}{4*CI_O*\ln(1-VE)^2}}$ (3), where CI_O is the cumulative incidence of infection in the trial population. For the cRCT, this sample size is multiplied by the design effect as defined in the main text, with ICC calculated using the ANOVA method (4). We calculate necessary sample size to achieve 90% power for an iRCT (in which half of the vaccinees in each study cluster are vaccinated) and for a cRCT (in which half of the study clusters have all participants vaccinated, and the other half are given control), and plot the ratio of the necessary sample size for a cRCT compared to an iRCT. Areas of parameter space in which this ratio is less than 1 are indicative of parameters for which the cRCT is theoretically more efficient at detecting the total effect than the iRCT is at detecting the direct effect.

When $R_0<1$, the size of an outbreak in a large population is given by $\boldsymbol{\frac{1}{1-R_0}}$, but this formula does not apply to small communities, especially when R_0 is close to 1. Therefore, we restrict the theoretical analyses to parameter combinations when R_0 in vaccinated communities in the cRCT is greater than 1, assuming that any qualitative results we saw in this parameter space would be maintained as R_0 crosses 1.

2. Simulation

The main population model is a standard deterministic susceptible-exposed-infectious-removed (SEIR) compartmental model, with three exposed and three infectious compartments to yield

gamma-distributed incubation and infectious periods. We assumed a time-varying transmission rate in the main population, so that the importation rate into the communities is proportional to the prevalence of infection in the main population, and disease natural history parameters representative of the 2014-2015 Ebola epidemic in Liberia (5).

The disease model in the communities is a stochastic susceptible-exposed-infectious-removed (SEIR) model. Each susceptible individual has a daily hazard of becoming infected and moving into the exposed compartment from two sources: the daily hazard of infection from each infectious *neighbor* is β, and the daily hazard of infection for an individual in community i from the main population is F_iI, where I is the prevalence of infectious individuals in the main population and F_i is a proportionality constant reflecting the degree of contact between the main population and the i^{th} community.

The hazard rate of introduction into the study population is time-varying with the progression of the epidemic in the main population, and we calibrate the constant of proportionality in each cluster F_i using an assumed rate of importation events, M_i cases/year. The formula that connects these two quantities is $F_i = -\frac{\ln(1-M_i*T)}{f/\mu}$, where f is the final size of the epidemic in the main population, μ is the mean infectious period, and T is the length of the epidemic in years. We model the relationship between importation rate and community size in two ways. Firstly, for community i we assume $M_i = a\sqrt{N_i}$, where N_i is the community size (6), and the *per capita* importation rate in community i is $\frac{a}{\sqrt{N_i}}$, where the constant a determines the magnitude of the importation rate. Secondly, we assume $M_i = a'N_i$, so that the *per capita* importation rate in community i is a'. The values for a and a' were chosen so that a community of size N_i=100 had on average between 0.25 and 1 introductions over the course of a two-year epidemic.

The transmission rate β in the main population varied with time using the formula $\beta(t) = \hat{\beta}(1 - \frac{\alpha_2}{1+ e^{-\alpha_1(t-\alpha_\tau)}})$. Parameters were chosen to give a reasonable fit to weekly Ebola incidence data from Liberia. Specifically, $\hat{\beta}$ = 0.94, α_1 = 0.19, α_2 = 0.6, α_τ = 27.79. The average incubation/latent period is 7.14 days and the average infectious period is 3 days.

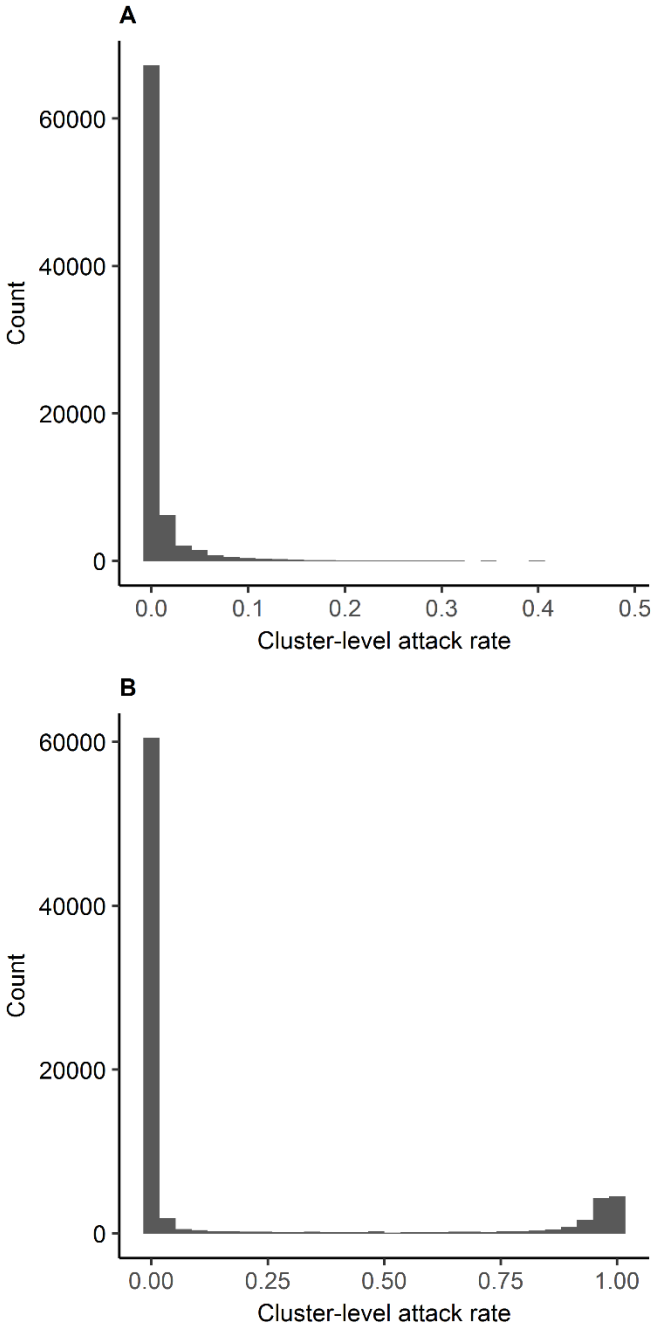
We assume that the incubation and latent periods are concurrent, meaning that symptom onset occurs when infectiousness begins. Once infected, individuals spend a number of days in the exposed compartment drawn from a gamma distribution with mean 9.7 days and SD 5.5 days before moving into the infectious compartment (7). They spend a number of days in the infectious compartment drawn from an independent gamma distribution with mean 5 and SD 4.7 based on data on the time to hospitalization (7), after which they move into the removed compartment. For simplicity and to generalize away from the Ebola epidemic, we assume no *post mortem* transmission, meaning that whether an individual dies or recovers does not affect the estimated efficacy or power of the trial.

Once enrolled, individuals are followed for a number of days and, for infected individuals, time from enrollment to symptom onset is recorded. Individuals who never develop symptoms are censored at the end of the study; there are no other sources of censoring. The vaccine is multiplicative leaky (8), reducing susceptibility to infection by a factor (1-VE) and having no effect on those who are already exposed or infectious when vaccinated, and no effect on the progression or infectiousness of vaccinated individuals who become infected. We assume the protective efficacy of the vaccine starts on the day of vaccination.

REFERENCES

1. Diekmann OH, J.A.P. *Mathematical Epidemiology of Infectious Diseases*. Chichester, UK: John Wiley & Sons, Ltd; 2000.
2. Haber M, Longini IM, Jr., Halloran ME. Measures of the effects of vaccination in a randomly mixing population. *Int J Epidemiol* 1991;20(1):300-310.
3. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983;39(2):499-503.
4. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemp Clin Trials* 2012;33(5):869-880.
5. Kucharski AJ, Camacho A, Flasche S, et al. Measuring the impact of Ebola control measures in Sierra Leone. *PNAS* 2015;112(46):14366-14371.
6. Keeling MJ, Rohani P. The Importance of Imports. In: Keeling MJ, Rohani P, eds. *Modeling Infectious Diseases in Humans and Animals*. Princeton, New Jersey: Princeton University Press, 2007:209-212.
7. Team WHOER. Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014;371(16):1481-1495.
8. Smith PG, Rodrigues LC, Fine PEM. Assessment of the protective efficacy of vaccines against common disease using case-control and cohort studies. *Int J Epidemiol* 1984;13(1):87-93.

Web Figure 1. Relationship between R_0 and distribution of cluster-level attack rates. Histogram of cluster-level attack rate for R_0=0.6 (**A**) and R_0=3 (**B**).



Web Figure 2. Ratio of necessary sample size for 90% power to detect vaccine effect for a cRCT (total effects) relative to an iRCT (direct effect) with a hazard rate-based analysis, varying R_0 and true vaccine efficacy. Final size equations apply only when $R_0^V > 1$.