



Published in final edited form as:

*J Infect Dis.* 2009 August 15; 200(4): 609–613. doi:10.1086/600891.

## Power to Detect HIV Vaccine Effects in Repeated Low-Dose Challenge Experiments

Michael G. Hudgens<sup>1</sup>, Peter B. Gilbert<sup>2</sup>, John R. Mascola<sup>3</sup>, Chih-Da Wu<sup>1</sup>, Dan H. Barouch<sup>4</sup>, and Steven G. Self<sup>2</sup>

<sup>1</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC

<sup>2</sup>Fred Hutchinson Cancer Research Center and Department of Biostatistics, University of Washington, Seattle, WA

<sup>3</sup>Vaccine Research Center, NIAID, NIH, Bethesda, MD

<sup>4</sup>Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

### Abstract

Simulation studies were conducted to estimate the statistical power of repeated low-dose challenge experiments in non-human primates to detect a candidate HIV vaccine's effect. The effect of various design parameters on power was explored. Simulation results indicate repeated low-dose challenge studies with total sample size 50 (25 per arm) typically provide adequate power to detect a 50% reduction in the per-exposure probability of infection due to vaccination. Power generally increases with the maximum number of allowable challenges per animal, the per-exposure risk of infection in controls, and the proportion susceptible to infection.

### Keywords

HIV; Macaque; Vaccine

## 1 Introduction

A preventive vaccine against HIV-1 would be a valuable tool in curbing the pandemic. Evaluation of vaccine candidates in the SIV non-human primate model is an important step in the assessment of the potential efficacy of analogous HIV-1 vaccines [1,2]. Historically, vaccine regimens have been tested in non-human primates by administering a single high dose intravenous or mucosal inoculation of the challenge virus, typically resulting in infection of all animals under study after one exposure. Recently, evaluation of candidate HIV vaccines (and other preventive interventions) has entailed repeated low-dose mucosal challenge studies [3–6] that may more closely mimic typical exposure in natural human transmission settings. A primary objective of these studies is to assess vaccine efficacy for prevention of infection.

Since the repeated low-dose challenge study design has only recently been implemented in evaluation of candidate HIV vaccines, the literature on design considerations of these studies

---

To whom correspondence should be addressed: Michael G. Hudgens, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC 27599 mhudgens@bios.unc.edu.

Potential conflicts of interest: none

Presented at: AIDS Vaccine 2008 Meeting, Cape Town, South Africa, October 2008

is limited. Recent investigations [7,8] demonstrated such challenge studies can be adequately powered to test for vaccine efficacy to prevent infection with feasible samples sizes of non-human primates. However, the effect of various design parameters such as the challenge dose, the percent of animals susceptible to infection, or the unequal allocation of animals to vaccine and placebo arms have not been systematically investigated. Given results from a recent HIV vaccine efficacy trial [9], the possibility that vaccination may increase the probability of infection needs to be entertained. Direct comparisons of different statistical tests employed in the analysis of repeated low-dose challenge studies are also needed.

In this paper, we describe simulation studies which were conducted to better understand how the design of challenge experiments can affect the statistical power to detect a candidate vaccine's effect. Simulated scenarios were varied according to the total sample size, the per-exposure risk of infection in controls, the magnitude and direction of the vaccine effect, the fraction allocated to vaccine or control, the proportion of susceptible animals, and the maximum number of challenges per animal. While this work is motivated by the development of an HIV vaccine, our results can be used to inform the design of challenge studies for evaluation of other vaccines and other preventive interventions.

## 2 Methods

Simulation studies were conducted to assess the statistical power of challenge experiments to detect a candidate vaccine's effect on the per-exposure risk of infection. Power was estimated by simulating multiple challenge studies and calculating the proportion of simulated data sets where various statistical tests (described below) rejected the null hypothesis of no vaccine effect. Simulations were also conducted under the null hypothesis to evaluate type I error, i.e., the probability a statistical test incorrectly rejects the null. All statistical tests were two-sided and conducted at the  $\alpha=0.05$  significance level. For each scenario described below, 10,000 challenge studies were simulated.

### 2.1 Assumptions

Simulations were conducted under the following four key assumptions. First, challenge studies have two arms, with a specified fraction of non-human primates randomly assigned to receive the candidate vaccine and the remaining animals serving as controls. Second, the vaccine has a leaky mechanism of protection in animals that are susceptible to infection, i.e., the vaccine decreases (or increases) the per-exposure probability of infection by the same multiplicative amount for all susceptible animals. The power is expected to be at least as high in trials where the vaccine has an all-or-none mechanism of protection [8]. Third, challenges are ceased if an animal remains uninfected after  $c_{max}$  exposures. Fourth, given an animal is susceptible, the probability of infection is independent of the number of prior exposures.

### 2.2 Statistical tests and power

Let  $p_1$  ( $p_0$ ) denote the probability of infection from a single exposure in a vaccinated (unvaccinated) animal. We are interested in testing the null hypothesis  $H_0 : p_0 = p_1$  or equivalently  $H_0 : RR = p_1/p_0 = 1$ , where RR denotes the relative risk of infection per-exposure. Three different tests of  $H_0$  were considered in the simulations studies: the logrank test (as in [3]), Fisher's exact test (as in [7]), and a likelihood ratio test (as in [4]). The likelihood ratio test is described in the Appendix. For all tests, we evaluated power as a function of RR. Values of RR less than one indicate vaccination has a protective effect. For example,  $RR = 0.4$  corresponds to 60% vaccine efficacy where vaccine efficacy equals  $(1 - RR) \times 100\%$ . Conversely, values of RR greater than one indicate a harmful effect of vaccination. For instance,  $RR = 1.5$  corresponds to the vaccine increasing the per-exposure probability of infection by 50%.

### 3 Results

Unless stated otherwise, all results are based on simulations where  $p_0 = 0.5$ , equal numbers of animals are allocated to the vaccine and control arms, the maximum number of challenges per animal is  $c_{max} = 10$ , and all animals are susceptible to infection. Simulated power is presented only for the logrank test unless the results for the likelihood ratio or Fisher's exact tests differ markedly.

#### 3.1 Sample size

First we considered the effect of the total sample size  $N$  on power. The results depicted in Figure 1A indicate a repeated low-dose challenge study with  $N = 50$  animals has at least 80% power to detect  $RR = 0.5$ . Smaller studies of size  $N=40$  and  $N=30$  have only 74% and 61% power respectively to detect  $RR = 0.5$ . If the vaccine enhances the probability of infection,  $N = 50$  also provides at least 80% power to detect  $RR = 1.7$ .

#### 3.2 Challenge dose

Figure 1B shows power for different values of  $p_0$ . These results demonstrate that power increases with  $p_0$ , suggesting challenge doses should be chosen such that the per-exposure probability of infection is approximately 0.5. Values of  $p_0$  greater than 0.5 were not considered as such high doses can preclude the ability to determine whether vaccination enhances the probability of infection. For instance, suppose  $p_0 = 1$  corresponding to a high-dose challenge study and the vaccine has no protective effect. Then all animals will become infected after a single challenge, making it impossible to detect any increase in the probability of infection due to vaccination that might occur at lower doses. Additional disadvantages of high-dose challenge studies have been described elsewhere [4, 7].

#### 3.3 Unequal allocation

The effect of varying the relative number of animals allocated to the vaccine ( $n_v$ ) and control ( $n_c$ ) arms is depicted in Figure 1C. In general, there is a modest diminution of power with an unbalanced design. Given a secondary objective of challenge studies can entail identifying immunological correlates of risk, allocating more than half of the non-human primates to vaccine may be preferred.

#### 3.4 Maximum number of challenges per animal

Figure 1D shows power increases with the maximum number of challenges per animal ( $c_{max}$ ). There is an appreciable increase in power moving from  $c_{max} = 1$  to  $c_{max} = 3$ , indicating the importance of allowing for repeated exposures when using a low-dose challenge. For example, when  $RR = 0.5$  there is 47% power for  $c_{max} = 1$  compared to 75% power for  $c_{max} = 3$ . Gains in power are less pronounced when increasing  $c_{max}$  from 3 to 10. For example, the power to detect  $RR = 0.5$  is 84% when  $c_{max} = 10$ . Increasing  $c_{max}$  to 20 yields no appreciable increase in power over  $c_{max} = 10$  (results not shown) since so few animals are expected to escape infection after 10 challenges when  $p_0=0.5$  and  $RR=0.5$ .

#### 3.5 Heterogeneous susceptibility

Finally, we consider the possibility some fraction of animals is not susceptible to infection. For example, using repeated low-dose rectal challenges of SHIV, Garcia-Lerma et al. [3] reported 18 untreated macaques became infected after a median of two challenges, yet one macaque remained uninfected after 14 exposures. We assume the probability of an animal not being susceptible is independent of randomization assignment. Figure 2 shows the power of the three tests assuming different fractions of susceptible animals (100%, 90%, 80%). Unlike

previous results (Figure 1), power and type I error of the logrank, Fisher's exact and likelihood ratio tests differ for this set of simulations.

There are three important results to be gleaned from Figure 2. First, Fisher's exact test has a type I error rate substantially above the nominal  $\alpha = 0.05$  significance level. Second, power of the logrank test decreases as the fraction of susceptible animals decreases. This decrease in power is consistent with the results of Regoes et al. [7] when they consider the effect of heterogeneity in the per-exposure probability of infection on power. Third, the likelihood ratio test maintains the appropriate type I error rate and has only minimal loss of power as the fraction of susceptible animals decreases.

These results have three implications. First, if some animals are not susceptible to infection, the standard logrank test is not recommended due to diminished power. Rather, tests designed to provide greater power in the presence of an immune or "cured" fraction should be considered. If the modeling assumptions are justified, the likelihood ratio test will be optimal. However, if the model is incorrect, the likelihood ratio test may not have the correct type I error rate or may be less powerful than nonparametric tests designed for this setting (e.g., a weighted logrank test [10]). Second, Fisher's exact test is not recommended if some animals are not susceptible due to inflated type I error. Similar results (not shown) were obtained from simulations allowing for heterogeneity in susceptibility between animals by randomly sampling individual transmission probabilities from beta distributions [7]. Third, provided the appropriate statistical test is employed, repeated low-dose challenge studies can still achieve adequate power even if 10–20% of animals are not susceptible to infection. For example, a challenge study with  $N = 50$  animals has over 80% power (using the likelihood ratio test) to detect  $RR = 0.45$  when 90% of the animals are susceptible.

## 4 Conclusions

Despite limited sample size, repeated low-dose challenge studies can reliably detect the protective effect of a vaccine candidate. For example, 50 non-human primates (25 per arm) will generally provide sufficient power to detect a 50% reduction in the per-exposure probability of infection due to vaccination. In other words, repeated low-dose challenge studies can achieve the same power to detect vaccine efficacy as much larger phase IIB or III clinical trials and can produce roughly the same number of endpoints (i.e., infections) as a screening test-of-concept (TOC) trial [11,12].

Power of repeated low-dose challenge studies generally increases with the per-exposure risk of infection in controls, suggesting titration studies should strive for identifying challenge doses where the per-challenge probability of infection is approximately 0.5. Randomizing animals to vaccine in 2:1 or 3:1 ratios results in modest diminution of power compared to 1:1 allocation. Power tends to decrease with the proportion not susceptible to infection, although adequate power can still be achieved provided the proportion of animals not susceptible does not exceed 20% and the appropriate statistical test is employed. Finally, power decreases with the maximum number of challenges per animal, with repeated low-dose challenge designs having substantially more power than single challenge studies.

The results presented here can aid individual investigators in determining an efficient study design for their particular setting. Additionally, we have developed a web-based calculator (accessible from <http://www.bios.unc.edu/~mhudgens>) to estimate the power of repeated low-dose challenge studies for various study designs.

## Acknowledgments

Financial support: National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant R01 AI054165-04)

## Appendix

### Appendix

The likelihood ratio test is based on the following discrete time survival model. Let  $\theta$  denote the probability an animal is not susceptible to infection. For animal  $i$ , let  $z_i$  denote treatment assignment (1 vaccine, 0 control),  $t_i$  denote the number of exposures until infection (up to  $c_{\max}$ ), and  $\delta_i$  denote whether the animal is infected by the end of the study. Assume we observe  $N$  independent, identically distributed copies of  $(z_i, t_i, \delta_i)$  and that, conditional on whether an animal is susceptible, the probability of infection is independent of the number of prior exposures. Then the likelihood is

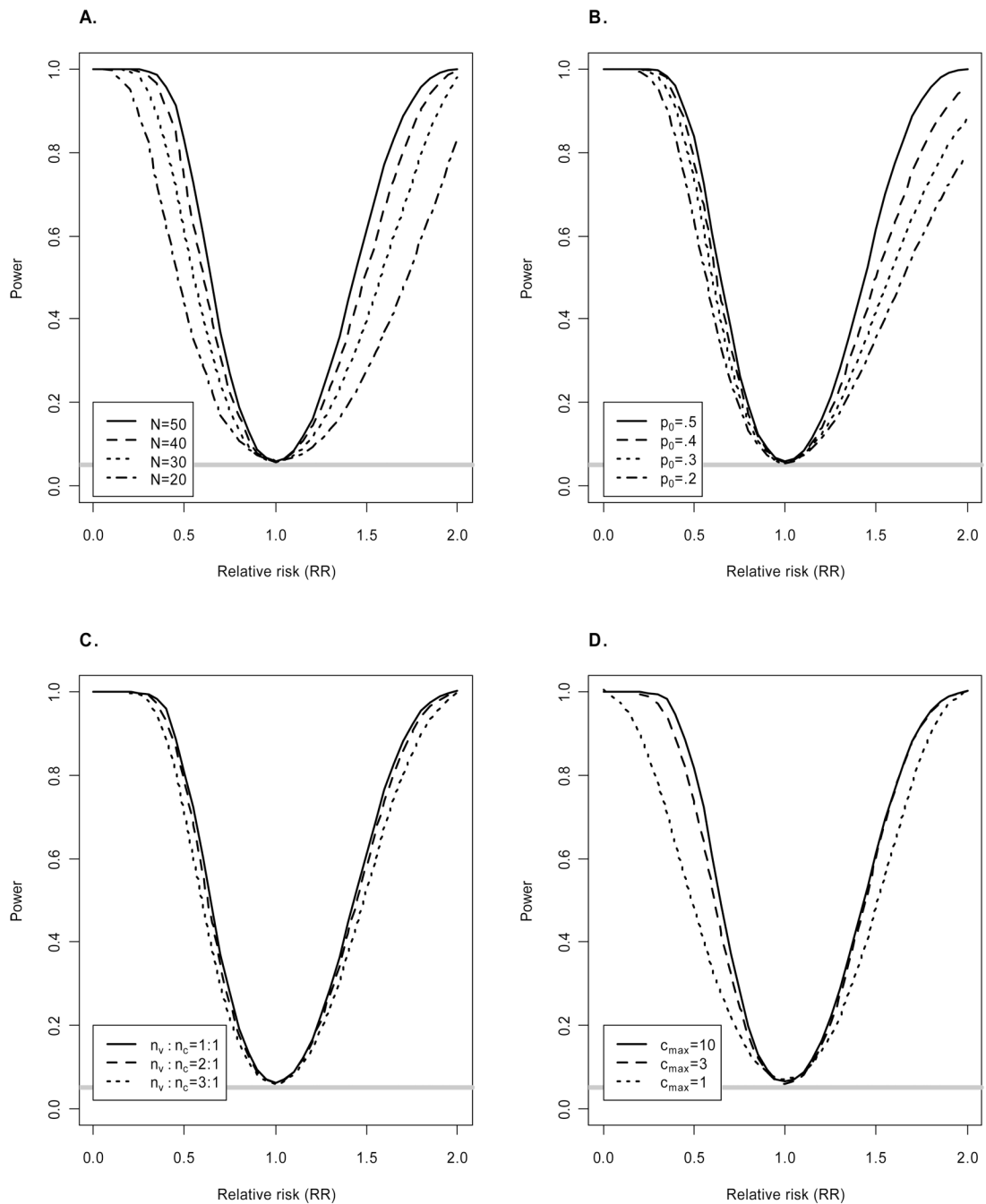
$$\prod_{i=1}^N \prod_{z=0}^1 \left\{ \theta + (1-\theta)(1-p_z)^{t_i} \right\}^{I_{z_i=z, \delta_i=0}} \left\{ (1-\theta)(1-p_z)^{t_i-1} p_z \right\}^{I_{z_i=z, \delta_i=1}}$$

The likelihood ratio test equals twice the difference in the maximum value of the likelihood computed with and without the constraint  $p_0 = p_1$ .

## References

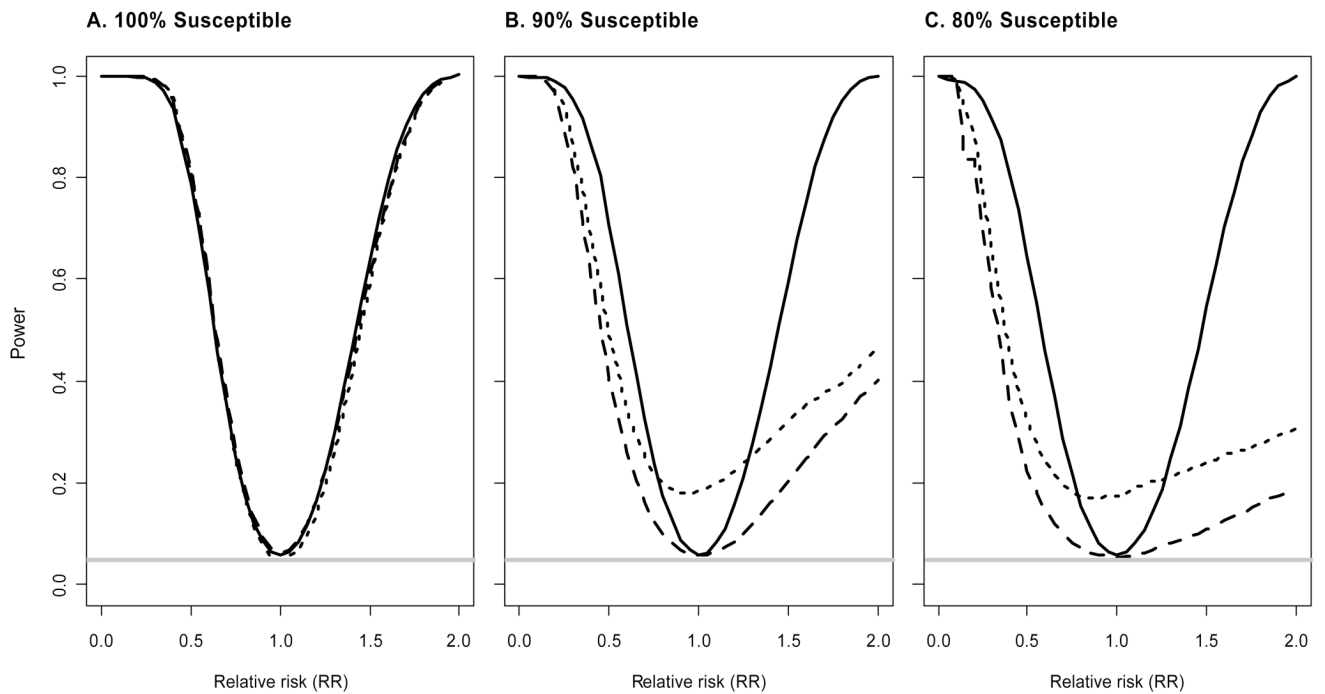
1. Morgan C, Marthas M, Miller C, Duerr A, Cheng-Mayer C, et al. The use of nonhuman primate models in HIV vaccine development. *PLoS Med* 2008;5(8):e173. [PubMed: 18700814]
2. Fauci AS, Johnston MI, Dieffenbach CW, Burton DR, Hammer SM, Hoxie JA, Martin M, Overbaugh J, Watkins DI, Mahmoud A, Greene WC. HIV vaccine research: the way forward. *Science* 2008;321:530–532. [PubMed: 18653883]
3. Garca-Lerma JG, Otten RA, Qari SH, Jackson E, Cong ME, Masciotra S, Luo W, Kim C, Adams DR, Monsour M, Lipscomb J, Johnson JA, Delinsky D, Schinazi RF, Janssen R, Folks TM, Heneine W. Prevention of rectal SHIV transmission in macaques by daily or intermittent prophylaxis with emtricitabine and tenofovir. *PLoS Med* 2008;5:e28. [PubMed: 18254653]
4. Ellenberger D, Otten RA, Li B, Rodriguez V, Sariol CA, Martinez M, Monsour M, Wyatt L, Hudgens MG, Kraiselburd E, Moss B, Robinson H, Folks T, Butera S. HIV-1 DNA/MVA vaccination reduces the per exposure probability of infection during repeated mucosal SHIV challenges. *Virlogy* 2006;352:216–225.
5. Subbarao S, Otten R, Ramos A, Jackson E, Monsour M, Bashirian S, Kim C, Johnson J, Soriano V, Hudgens MG, Butera S, Janssen R, Paxton L, Greenberg A, xxFolks T. Chemoprophylaxis with Tenofovir Disoproxil Fumarate provided partial protection against Simian Human Immunodeficiency Virus infection in macaques given multiple virus challenges. *Journal of Infectious Diseases* 2006;194:904–911. [PubMed: 16960777]
6. Otten RA, Adams DR, Kim CN, Jackson E, Pullium JK, Lee K, Grohskopf LA, Monsour M, Butera S, Folks TM. Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *J. Infect. Dis* 2005;191:164–173. [PubMed: 15609225]
7. Regoes RR, Longini IM, Feinberg MB, Staprans SI. Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Medicine* 2005;2(8):e249. [PubMed: 16018721]
8. Hudgens MG, Gilbert PB. Assessing vaccine effects in repeated low-dose challenge experiments. *Biometrics*. 2009In press

9. Buchbinder SP, Mehrotra DV, Duerr A, et al. for the Step Study Protocol Team. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet* 2008;372:1881–1893. [PubMed: 19012954]
10. Ewell M, Ibrahim JG. The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis* 1997;3:5–12. [PubMed: 9384622]
11. WHO/UNAIDS/IAVI International Expert Group. Executive summary and recommendations from the WHO/UNAIDS/IAVI expert group consultation on 'Phase IIB-TOC trials as a novel strategy for evaluation of preventive HIV vaccines', 31 January-2 February 2006, IAVI, New York, USA. *AIDS* 2007;21:539–546. [PubMed: 17301582]
12. Excler JL, Rida W, Priddy F, Fast P, Koff W. A strategy for accelerating the development of preventive AIDS vaccines. *AIDS* 2007 Nov;21:2259–2263. [PubMed: 18090273]



**Figure 1.**

Power of logrank test as a function of relative risk (RR). Unless indicated otherwise, simulations assumed equal numbers of animals allocated to the vaccine ( $n_v$ ) and control ( $n_c$ ) arms, total sample size  $N = 50$ , maximum of  $c_{max} = 10$  challenges per animal, per-challenge probability of infection for controls  $p_0 = 0.5$ , and all animals susceptible to infection. The horizontal gray line corresponds to the significance level  $\alpha=0.05$ . Power of likelihood ratio test and Fisher's exact test is similar (results not shown).



**Figure 2.**

Power of likelihood ratio, logrank, and Fisher's exact tests as a function of relative risk (RR) for different fractions of susceptible animals. Solid lines correspond to the likelihood ratio test, long dashes to the logrank test, and short dashes to Fisher's exact test. Simulations assume equal allocation to the vaccine and control arms, total sample size  $N = 50$ , maximum of  $c_{max} = 10$  challenges per animal, and per-challenge probability of infection for controls  $p_0 = 0.5$ . The horizontal gray line corresponds to the significance level  $\alpha = 0.05$ .