Published in final edited form as:
AIDS. 2012 January 14; 26(2): 123–126. doi:10.1097/QAD.0b013e32834e1ce7.

Apparent Declining Efficacy in Randomized Trials: Examples oftheRV144 HIV Vaccine and CAPRISA 004 Microbicide Trials

Justin J. O'HAGAN[1,2], **Miguel A. HERNÁN**[1,3], **Rochelle P. WALENSKY**[4,5,6], and **Marc LIPSITCH**[1,2,7]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[2]Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, MA, USA

[3]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

[4]Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[5]Center for AIDS Research, Harvard Medical School, Boston, MA, USA

[6]Division of Infectious Disease, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[7]Department of Immunology, Harvard School of Public Health, Boston, MA, USA

Recent HIV prevention trials have given hope that a suite of interventions which effectively reduce individuals' risk of HIV infection will soon be widely available [1–6]. In two studies, the RV144 and CAPRISA 004 trials, the relative risk of infection increased towards the null value of one over time. The RV144 and CAPRISA investigators interpreted these trends as evidence that the interventions' effects declined over the study period and suggested that their respective findings may be explained by waning vaccine efficacy and decreasing adherence [2,7,8]. Here, we discuss these trends in the trials' results and note that, in addition to the possible mechanisms cited by the investigators, their apparent waning efficacy may be explained in part by selection bias due to heterogeneity in infection risk, an explanation that has not been considered previously. This bias arises when study participants vary in their susceptibility to infection, e.g., because of differences in immune systems or exposure to infection. This can lead to increasing differences in the composition of the study population in each trial arm over time as those at highest risk become infected, and can occur despite comparability between arms at baseline. This issue is termed "frailty" in statistics and demography, where a large body of literature addresses the matter [9–14]. We also discuss several methods that can improve understanding of the effects of infectious disease interventions and risk factors by assessing the impact of frailty on results.

Variation in frailty among study participants likely creates trends in the incidence of infection over the course of a study. To understand this phenomenon, consider a theoretical placebo-controlled randomized trial in which the risk of infection varies among participants. The highest risk individuals in such a trial are expected to become infected earlier, leaving a

Corresponding author, address & email: Justin J. O'HAGAN, 677 Huntington Avenue, Dept. of Epidemiology, Boston, MA 02115, USA johagan@hsph.harvard.edu.
Request for reprints: Justin J. O'HAGAN, 677 Huntington Avenue, Dept. of Epidemiology, Boston, MA 02115, USA

Conflicts of Interest
JOH and MAH report no conflicts of interest.

pool of lower risk individuals at later time points. If the intervention being tested is effective, the decline in the incidence of infection over time will be larger in the placebo arm because these individuals experience no direct protection from the intervention, and so those at high-risk will be quickly depleted, thereby lowering the infection rate over follow-up. However, high-risk individuals in the active arm may remain uninfected due to the protection conferred by the intervention, so the active arm's infection rate will be less affected. Consequently, the time-specific rate ratio (RR) for treatment vs. placebo will increase over time from a value of less than one initially to a value that may exceed one later. This phenomenon has also been termed "survivor bias", "survivor cohort effect", "crossing of hazards" and "depletion of susceptibles", and is observed in both chronic and infectious disease research [15–19].

Many randomized trials report a weighted average of the time-specific RR over the entire follow-up, but, because of the phenomenon described above, the average RR will become increasingly attenuated as follow-up time increases. This attenuation will occur even if all risk factors were balanced between study arms at baseline, as expected in a large randomized trial, and if the biological efficacy of the intervention is constant over time, which can lead investigators to reject an efficacious intervention. Because the results of any rate analysis can be affected by frailty, it can also cause investigators to overlook a risk factor for infection that is in fact harmful.

To illustrate these issues we assessed whether the results of the RV144 and CAPRISA trials are consistent with an attenuation of the average RR due to frailty [1,2]. Though we focused on the widely discussed RR estimates, the RV144 team also presented Kaplan-Meier estimates [7]. Tables 1–2 show average incidence rates of infection and RRs calculated from baseline up to several time points in each trial [1,2]. Both studies show decreasing incidence rates in the placebo arms and generally increasing RRs. The incidence rate of infection in the intervention arm of the CAPRISA microbicide trial decreased slightly [2]. The incidence rate of infection in the vaccine arm of the RV144 trial increased from 0–12 months to 0–24 months; it then decreased slightly but remained below that of the placebo arm.

To test for variation in the time-specific RRs, we fit a Cox model with a product ("interaction") term between the intervention arm and time (on the logarithmic scale) on study [20]. A positive coefficient for this product term would indicate that the time-specific RR increases over time, in the context of prevention trials where the average RR is less than one, this implies that the average RR becomes attenuated over time. Figure 2C of the RV144 publication provides information on the number of infections and loss to follow-up in each arm for various time points [1]. Using these data we reconstructed an interval-censored dataset of the trial for analysis with a Cox model. The coefficient of the product term was 0.42, indicating that the time-specific RR increased over time, but the 95% confidence interval was wide (−0.09, 0.93). However, as such tests have limited power, failure to reject the null hypothesis cannot be taken as strong evidence that there was no attenuation over time [21]. Because any possible trend in estimates may not be linear, we conducted several other tests of the proportional hazards assumption (e.g., using quadratic forms of time), and each produced similar results. There was insufficient information to perform a similar test on the CAPRISA data.

The attenuation of the average RR suggested by our analysis of the RV144 trial may have been due to frailty, waning efficacy, differential loss to follow-up, random variability, or any combination of the above. Furthermore, we note that frailty will have a large impact on results only when disease incidence is high in at least one subgroup. The RV144 investigators report that immune responses among the vaccinated waned during the first 6 months after vaccination; therefore vaccine efficacy may have similarly waned [7]. As

waning efficacy and frailty are not mutually exclusive mechanisms of attenuation, both may have contributed to the increasing trend in the RR. The CAPRISA investigators report that some of the increase in the RR may be explained by decreasing adherence to microbicide use [2]. While waning vaccine efficacy and trends in adherence may help to explain the increase in RR observed over the course of both trials, neither mechanism can explain the decreasing incidence in the placebo arms. In contrast, frailty parsimoniously explains both observations.

Because RR estimates are less affected by frailty during early follow-up, this may partly explain why studies that stop early because a considerable benefit is found can give larger estimates than those that continue. This is an alternative, and non-mutually exclusive, explanation to random variation [22] that has been insufficiently addressed in the literature on biases involved in stopping trials early [22–25]. Similarly, frailty has not been considered as a candidate explanation for null results in HIV prevention trials [26–30], or for variation in estimated effects across studies in meta-analyses [31,32].

Three approaches for dealing with frailty have been suggested. First, data on known risk factors for infection, including direct or proxy data on exposure to infection, could be collected and used in sensitivity analyses. This approach typically adjusts for baseline risk factors only, and uses standard methods, such as Cox models [9,10,32,33]. However, it may not be feasible to measure all risk factors for infection, and some risk factors may not even be known. When the goal is to estimate a biological measure of the direct effect of an intervention or risk factor, frailty effects due to heterogeneity in exposure to infection can be minimized by censoring individuals when they are no longer at risk of infection. Data on exposure to sexually transmitted infections can be obtained from serodiscordant couples studies, from partner report, or when participants report having no partners during a study period, and so could not have been exposed to HIV via sexual routes [4,34–38]. Survival curves, which use risks instead of rates, are not subject to the same frailty effects discussed above and can be estimated for observational studies as well as trials [15,39]. Second, some have suggested a type of crossover design to reduce the magnitude of attenuation in trials that examine reversible interventions through limiting the depletion of those most susceptible to infection by alternately giving both study arms access to the protective intervention [30]. Third, alternative analytical methods have been proposed, such as frailty models, which can account for heterogeneity by modeling the distribution of risk in the population based on parametric assumptions, and which have been used to analyze several vaccine trials [40–42]. Bayesian hierarchical models have been used to estimate the efficacy of a previous HIV vaccine candidate, and permit the incorporation of data on contact patterns and the prevalence of infection [43]. Frailty models have been widely used, particularly when the assumption of proportional hazards is considered unlikely to hold due to frailty effects [14]. In summary, accounting for frailty may improve our understanding of study results and facilitate decision-making on the use of interventions and avoidance of risk factors.

Acknowledgments

Sources of Funding:

ML reports having received consulting fees from Novartis, Pfizer/Wyeth Vaccines, Roche and AIR Worldwide. RPW reports accepting an invitation to present expert testimony for a law firm.

JOH, ML and RPW were supported by Award Number U54GM088558 from the National Institute of General Medical Sciences. MAH was supported by NIH grant R01 AI073127. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

JOH drafted the manuscript and conducted the analysis. All authors contributed to the interpretation of the data and critically revised the article for important intellectual content.

References

1. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med. 2009; 361:2209–2220. [PubMed: 19843557]
2. Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. Science. 2010; 329:1168–1174. [PubMed: 20643915]
3. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010; 363:2587–2599. [PubMed: 21091279]
4. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. N Engl J Med. 2011; 365:493–505. [PubMed: 21767103]
5. Centers for Disease Control Prevention. [Accessed July 20, 2011] TDF2 Press Release. Available from: http://www.cdc.gov/nchhstp/newsroom/PrEPHeterosexuals.html
6. [Accessed July 20, 2011] Partners in Prevention PrEP Press Release. Available from: http://depts.washington.edu/uwicrc/research/studies/files/PrEP_PressRelease-UW_13Jul2011.pdf
7. Michael, N. RV 144 Update: Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thai Adults. 17th Conference on Retroviruses & Opportunistic Infections; February 16–19, 2010; Available from: http://app2.capitalreach.com/esp1204/servlet/tc?c=10164&cn=retro&e=12354&m=1&s=20431&&espmt=2&mp3file=12354&m4bfile=12354
8. Abdool Karim, Q. Safety and Effectiveness of 1% Tenofovir Vaginal Microbicide Gel in South African Women: Results of the CAPRISA 004 Trial. AIDS; XVIII International AIDS Conference; July 18–23, 2010; 2010. Available from: http://globalhealth.kff.org/AIDS2010/July-20/Safety-and-Effectiveness.aspx
9. Lagakos SW, Schoenfeld DA. Properties of proportional-hazards score tests under misspecified regression models. Biometrics. 1984; 40:1037–1048. [PubMed: 6534407]
10. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1986; 71:431–444.
11. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979; 16:439–454. [PubMed: 510638]
12. Vaupel JW, Yashin AI. Heterogeneity's ruses: some surprising effects of selection on population dynamics. Am Stat. 1985; 39:176–185. [PubMed: 12267300]
13. Yashin AI, Iachine IA. How frailty models can be used for evaluating longevity limits: taking advantage of an interdisciplinary approach. Demography. 1997; 34:31–48. [PubMed: 9074830]
14. Govindarajulu US, Lin H, Lunetta KL, D'Agostino RB Sr. Frailty models: Applications to biomedical and genetic studies. Stat Med. 2011; 30:2754–2764. [PubMed: 21786277]
15. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010; 21:13–15. [PubMed: 20010207]
16. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004; 15:615–625. [PubMed: 15308962]
17. Rothman, KJ.; Greenland, S.; Lash, TL. Modern Epidemiology, Third Edition. Philadelphia: Lippincott Williams & Wilkins; 2008.
18. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. Epidemiology. 1994; 5:189–196. [PubMed: 8172994]
19. Moride Y, Abenhaim L, Yola M, Lucein A. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. J Clin Epidemiol. 1994; 47:731–737. [PubMed: 7722586]
20. Cox DR. Regression models and life-tables. J R Stat Soc Series B Methodol. 1972; 34:187–220.

21. Boutitie F, Gueyffier F, Pocock SJ, Boissel JP. Assessing treatment-time interaction in clinical trials with time to event data: a meta-analysis of hypertension trials. Stat Med. 1998; 17:2883–2903. [PubMed: 9921608]
22. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA. 2010; 303:1180–1187. [PubMed: 20332404]
23. Bassler D, Ferreira-Gonzalez I, Briel M, Cook DJ, Devereaux PJ, Heels-Ansdell D, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. J Clin Epidemiol. 2007; 60:869–873. [PubMed: 17689802]
24. Montori VM, Devereaux PJ, Adhikari NKJ, Burns KEA, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005; 294:2203–2209. [PubMed: 16264162]
25. Pocock S, White I. Trials stopped early: too good to be true? Lancet. 1999; 353:943–944. [PubMed: 10459899]
26. Graham SM, Shah PS, Aesch ZC-von, Beyene J, Bayoumi AM. A systematic review of the quality of trials evaluating biomedical HIV prevention interventions shows that many lack power. HIV Clin Trials. 2009; 10:413–431. [PubMed: 20133272]
27. Padian NS, McCoy SI, Balkus JE, Wasserheit JN. Weighing the gold in the gold standard: challenges in HIV prevention research. AIDS. 2010; 24:621–635. [PubMed: 20179575]
28. Hayes R, Kapiga S, Padian N, McCormack S, Wasserheit J. HIV prevention research: taking stock and the way forward. AIDS. 2010; 24(Suppl 4):S81–S92. [PubMed: 21042056]
29. Gray RH. Methodologies for evaluating HIV prevention intervention (populations and epidemiologic settings). Curr Opin HIV AIDS. 2009; 4:274–278. [PubMed: 19532064]
30. Auvert B, Sitta R, Zarca K, Mahiane SG, Pretorius C, Lissouba P. The effect of heterogeneity on HIV prevention trials. Clinical Trials. 2011; 8:144–154. [PubMed: 21478327]
31. Ford I, Norrie J, Ahmadi S. Model inconsistency, illustrated by the Cox proportional hazards model. Stat Med. 1995; 14:735–746. [PubMed: 7644855]
32. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? Control Clin Trials. 1998; 19:249–256. [PubMed: 9620808]
33. Hernández AV, Eijkemans MJ, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? Ann Epidemiol. 2006; 16:41–48. [PubMed: 16275011]
34. Celum C, Wald A, Lingappa JR, Magaret AS, Wang RS, Mugo N, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med. 2010; 362:427–439. [PubMed: 20089951]
35. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. Cochrane Database Syst Rev. 2002; 1:CD003255. [PubMed: 11869658]
36. Warner L, Newman DR, Austin HD, Kamb ML, Douglas JM Jr, Malotte CK, et al. Condom effectiveness for reducing transmission of gonorrhea and chlamydia: the importance of assessing partner infection status. Am J Epidemiol. 2004; 159:242–251. [PubMed: 14742284]
37. Niccolai LM, Rowhani-Rahbar A, Jenkins H, Green S, Dunne DW. Condom effectiveness for prevention of Chlamydia trachomatis infection. Sex Transm Infect. 2005; 81:323–325. [PubMed: 16061540]
38. Warner L, Ghanem KG, Newman DR, Macaluso M, Sullivan PS, Erbelding EJ. Male circumcision and risk of HIV infection among heterosexual African American men attending Baltimore sexually transmitted disease clinics. J Infect Dis. 2009; 199:59–65. [PubMed: 19086815]
39. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed. 2004; 75:45–49. [PubMed: 15158046]
40. Longini IM, Halloran ME. A frailty mixture model for estimating vaccine efficacy. J R Stat Soc Ser C Appl Stat. 1996; 45:165–173.
41. Halloran ME, Longini IM, Struchiner CJ. Estimability and interpretation of vaccine efficacy using frailty mixing models. Am J Epidemiol. 1996; 144:83–97. [PubMed: 8659489]
42. Haber M, Ndikuyeze A. Estimation of individual and population vaccination effectiveness from time-to-event data. Stat Med. 1998; 17:2617–2623. [PubMed: 9839352]

43. Yang Y, Gilbert P, Longini IM, Halloran ME. A Bayesian Framework for Estimating Vaccine Efficacy per Infectious Contact. Ann Appl Stat. 2008; 2:1409–1431. [PubMed: 19169384]

Table 1

RV144 HIV vaccine trial modified intent-to-treat (mITT) data[1]

	12 months		24 months		36 months		42 months	
	Vaccine	**Placebo**	**Vaccine**	**Placebo**	**Vaccine**	**Placebo**	**Vaccine**	**Placebo**
HIV Infections	12	30	32	50	45	65	51	74
Follow-up time (person-years)	7803	7790	15478	15443	22956	22892	26631	26556
Average IR (Cases/1000 PY)	1.54	3.85	2.07	3.24	1.96	2.84	1.92	2.79
mITT Average HR (95% CI)		0.40 (0.21, 0.78)		0.64 (0.41, 1.0)		0.69 (0.47, 1.01)		0.69 (0.48, 0.98)

IR = incidence rate (number of HIV infections/person-time)

PY = person-years

mITT = modified intent-to-treat

VE = vaccine efficacy

HR = hazard ratio

CI = confidence interval

The mITT analysis excluded 7 individuals who were already infected with HIV at baseline and were inadvertently enrolled due to a delay between screening for infection and vaccination.[1] As it was expected that some individuals would be enrolled with prevalent HIV infection, the mITT analysis was pre-specified as being the primary analysis in the trial protocol.[1] All other participants were analyzed according to their randomization group.

The dataset used to perform analyses was created using information provided in Figure 2C of the RV144 publication.[1] Individuals lost to follow-up are censored at their last visit and we assume that infections happened at the midpoint of the time interval in which they occurred.

All incidence rates, hazard ratios, number of cases and person-time data are estimated from baseline to the time indicated.

Table 2

CAPRISA HIV microbicide trial intent-to-treat (ITT) data[2]

	6 months		12 months		18 months		24 months		30 months	
	M	**P**	**M**	**P**	**M**	**P**	**M**	**P**	**M**	**P**
Average IR (Cases/100 PY)	6.0	11.2	5.2	10.5	5.3	10.2	5.6	10.2	5.6	9.1
ITT RR	0.53		0.50		0.53		0.60		0.61	
P-value	0.064		0.007		0.004		0.013		0.017	

M = microbicide arm

P = placebo arm

IR = incidence rate (number of HIV infections/person-time)

PY = person-years

ITT = intent-to-treat

RR = rate ratio

All incidence rates and incidence rate ratios are estimated from baseline to the time indicated.

Case and person-time data for each arm are not provided in the CAPRISA publication.[2]

AIDS. Author manuscript; available in PMC 2013 January 14.