Statistical Interpretation of the RV144 HIV Vaccine Efficacy Trial in Thailand: A Case Study for Statistical Issues in Efficacy Trials

Peter B. Gilbert,[1,2] James O. Berger,[4] Donald Stablein,[5] Stephen Becker,[3] Max Essex,[7] Scott M. Hammer,[9] Jerome H. Kim,[6] and Victor G. DeGruttola[8]

[1]Vaccine Infectious Disease Division, Fred Hutchinson Cancer Research Center, [2]Department of Biostatistics, University of Washington, [3]Bill and Melinda Gates Foundation, Seattle, Washington; [4]Department of Statistical Science, Duke University, Durham, North Carolina; [5]The EMMES Corporation and [6]Department of Molecular Virology and Pathogenesis, Division of Retrovirology, Walter Reed Army Institute of Research, US Military HIV Research Program, Rockville, Maryland; [7]Harvard School of Public Health AIDS Initiative, Department of Immunology and Infectious Diseases and [8]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; and [9]Division of Infectious Diseases, Columbia University, College of Physicians and Surgeons, New York, New York

Recently, the RV144 randomized, double-blind, efficacy trial in Thailand reported that a prime-boost human immunodeficiency virus (HIV) vaccine regimen conferred ~30% protection against HIV acquisition. However, different analyses seemed to give conflicting results, and a heated debate ensued as scientists and the broader public struggled with their interpretation. The lack of accounting for statistical principles helped flame the debate, and we leverage these principles to provide a more scientific interpretation. We first address interpretation of frequentist results, including interpretation of *P* values, synthesis of results from multiple analyses (ie, intention-to-treat versus per-protocol/fully immunized), and accounting for external efficacy trials. Second, we address how Bayesian statistics, which provide clearly interpretable statements about probabilities that the vaccine efficacy takes certain values, provide more information for weighing the evidence about efficacy than do frequentist statistics alone. Third, we evaluate RV144 for completeness of end point ascertainment and integrity of blinding, necessary tasks for establishing robustly interpretable results.

On 24 September 2009 the primary result of the RV144 randomized, placebo-controlled, efficacy trial of a prime-boost human immunodeficiency virus (HIV) vaccine regimen in Thailand was reported: borderline significant evidence that the vaccine reduced the rate of HIV acquisition ($P = .04$); modest vaccine efficacy (VE) estimated at 31% (95% confidence interval [CI], 1%–51%) [1]. Controversy ensued as scientists and the broader public struggled with interpreting these results, with fervor rekindled from the pretrial controversy about whether the trial should have taken place [2, 3]. Debate also centered on the importance of other, less encouraging assessments of VE in 2 other, overlapping study populations (eg, [4, 5]). Our goal is to foster more rigorous scientific interpretation of HIV vaccine efficacy trial results by deeper consideration of statistical principles.

Our discussion has 4 parts. The first addresses interpretation of frequentist results from efficacy trials, including (1) interpretation of *P* values and CIs, (2) synthesis of results from multiple populations, and (3) placement of results in context. The second illustrates the use of a complementary, Bayesian framework, which was applied in exploratory analyses of past efficacy trials [6]. For either statistical framework,

Received 8 June 2010; accepted 15 November 2010.

Potential conflicts of interest: S.M.H. has served as a scientific advisor to Merck and Progenics, has served as a member of a Data Monitoring Committee for Bristol-Myers Squibb, and is on the Board of Directors of SIGA. All other authors: no conflicts.

Presented in part: The HIV Vaccine Trials Network Full Group Meeting, Washington, DC, May 2010; at the Harvard School of Public Health Merck Workshop on Vaccines and the Control of Infectious Diseases, Boston, Massachusetts, June 2010; and at the Harvard School of Public Health Symposium in Honor of Steve Lagakos, Boston, Massachusetts, October 2010.

Reprints or correspondence: Peter Gilbert, PhD, Vaccine Infectious Disease Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, PO Box 19024, Seattle, WA 98109 (pgilbert@scharp.org).

The Journal of Infectious Diseases 2011;203:969–75

1537-6613/2011/2037-0001$15.00
DOI: 10.1093/infdis/jiq152

high rates of primary end point ascertainment and participant blinding are critical for obtaining valid inferences about VE, and the third part evaluates these issues and presents a simple sensitivity analysis. The fourth part addresses implications for improving future efficacy trials.

Interpretation of Frequentist Results from HIV Vaccine Efficacy Trials

A Brief History. Five efficacy trials have been conducted, 4 of which were completed (Table 1). The first 2 evaluated bivalent recombinant gp120 envelope protein-based vaccines in North America [7] and Bangkok, Thailand [8]. These phase 3 trials were designed to test whether VE was >30% and demonstrated no efficacy; failure appears to have resulted from inadequate antibody responses to exposing HIV variants [9]. Difficulties in developing efficacious HIV vaccines led to a switch from phase 3 licensure trials to phase 2b test-of-concept trials, which screen for VE >0% rather than VE >30% [10, 11]. Phase 2b trials are intended to weed out ineffective vaccines while advancing promising ones to further testing and require approximately one-third as many infection events as do phase 3 trials. Two Phase 2b trials have been conducted, both of a T cell–based vaccine. Results from the "Step trial" in the Americas indicated that the vaccine was ineffective and may have increased the rate of HIV acquisition for some subgroups [12], whereas the "Phambili trial" in South Africa did not yield definitive results, because the announcement of potential vaccine-harm in Step led to very early unblinding [13].

The US Military HIV Research Program, in collaboration with the Thai Ministry of Health and the National Institutes of Health conducted the RV144 efficacy trial [1]. Although this trial had far greater enrolment than the VaxGen Phase 3 trials (16,400 subjects, compared with 5400 and 2500), it is a large phase 2b trial, because it observed only 125 infections overall (the most relevant size measure), compared with 368 and 225 infections in the phase 3 trials.

Interpretation of Frequentist Statistics. P values have nonintuitive interpretations [14]. A common misinterpretation of a P value of .04 is that there is a 4% chance that the vaccine has no efficacy (VE = 0%). This may reflect what people really want to know; however, a P value does not support a statement about the probability of certain VE values, but rather provides the probability that the estimate of VE would be as far or farther from 0 than the observed estimate of 31% if the truth is that VE equals 0%. Understanding the true meaning of a P value requires a thought experiment: imagine repeated trials of the same design, but with random samples of individuals drawn from the same population under identical conditions. Although not feasible in reality, the thought experiment illustrates that the variability captured by P values refers only to that arising from sampling, not to other sources of variability. Thus, in the context of the RV144 trial and assuming that VE is 0%, a 1-sided P value of .02 implies that, in 100 repetitions of the trial, an average of 2 trials would show results with an estimated VE > 31%.

Similar misinterpretation may arise for the CI, where a 95% CI of 1%–51% is thought to imply a 95% probability that the true VE lies between 1% and 51%. However, a CI has a less intuitive interpretation: in the thought experiment above, we expect 95% of the repeated trials to yield a CI that includes the true VE. Statistically incorrect interpretations of the P value and CI contributed to a misunderstanding of the RV144 efficacy results. Another statistical approach− the Bayesian framework described below− provides more readily interpretable inferences about VE, complementing the frequentist framework.

Analysis of Multiple Populations. The RV144 paper reported analyses of VE for 3 study populations: the intention-to-treat (ITT) population, which included all randomized subjects; the modified ITT (MITT) population, which excluded subjects found to be HIV positive at the time of randomization; and the per-protocol (PP) population, a subset of the MITT population who remained HIV negative at the week 26 study visit and were

Table 1. Summary of Trial Results for Evaluating Human Immunodeficiency Virus Vaccine Efficacy

Efficacy trial	HIV risk group	Population	Nv (nv)	Np (np)	Estimated VE, %	(95% CI)	2-Sided P value
RV144[1]	General Population	ITT	8202 (56)	8200 (76)	26	(−4 to 48)	.08
Thailand	Mostly at	MITT	8197 (51)	8198 (74)	31	(1−51)	.04
	Heterosexual risk	PP	6176 (36)	6366 (50)	26	(−13 to 52)	.16
	61% Men						
Vax004[7]	MSM	MITT	3598 (241)	1805 (127)	6	(−17 to 24)	.59
North America	and Women; 94% Men	PP	3330 (191)	1679 (98)	4	(−23 to 24)	.77
Vax003[8]	Injection Drug	MITT	1267 (106)	1260 (105)	0	(−31 to 24)	.99
Bangkok	Users; 93% Men	PP	1193 (86)	1167 (79)	−8	(−46 to −21)	.64
Step[12]	MSM;	MITT	914 (49)	922 (33)	−50	(−141 to 5)	.07
Americas	100% Men	PP	835 (41)	840 (23)	−60	(−160 to 1)	.05

NOTE. Vaccine Efficacy is [1 – RH] × 100%, where RH is the relative hazard rate of Human Immunodeficiency Virus (HIV) infection in the vaccine versus placebo group. Nv (nv) is the number of subjects (number of these diagnosed with HIV infection) in the indicated population and assigned vaccine. Np (np) is similar for subjects assigned placebo. For the Step trial, women were also enrolled, but because only one woman acquired HIV infection, the efficacy analyses were restricted to men who have sex with men (MSM). CI, confidence interval; ITT, intention-to-treat; MITT, modified ITT; PP, per-protocol.

adherent to the protocol in a prespecified way [1]. The press release reported only results for the MITT population, which showed $P < .05$, delaying to the paper the reporting of the ITT and PP results, both of which showed $P > .05$ (Table 1). Among the scientific and lay communities, opinions varied on the interpretation of the differences in results, the overall meaning of the results, and the appropriate material to present in the initial report. However, for the 5 reasons listed below, we believe that the MITT analysis reasonably represents the study and that the observed differences among the analyses, although contributing to our understanding, are nonetheless of only minor scientific importance.

First, despite the custom in the scientific literature to enshrine the P value cut-off of .05 as the arbiter of whether an effect is likely to be real, small differences in P values have only minor impact on the probability that the VE equals 0%. Taken together, the 3 analyses provide modest evidence of a low-level protective efficacy—an important observation for the vaccine field and an interpretation that is not sensitive to which analyses are reported.

Second, the ITT analysis was conducted only to follow the protocol that, in hindsight, could reasonably have omitted this analysis. In general, an ITT analysis is prioritized for randomized, double-blind trials because it ensures that all prognostic factors are evenly distributed between the treatment groups on average, thereby ensuring a valid (unbiased) assessment of the effect of treatment assignment [15]. However, because in RV144 the baseline HIV infection status was ascertained through blinded procedures, the MITT analysis is equally valid as the ITT analysis. The published analyses of the other HIV vaccine efficacy trials reported only planned MITT and PP analyses (Table 1) [7, 8, 12].

Third, the PP analysis had less statistical power than the MITT analysis as a result of the 31% reduction in the number of end points, which would make the P value larger even if the VE levels are the same. Fourth, the standard analysis of VE is on shakier scientific footing for the PP population than for the MITT population because the comparator groups in the PP analysis are only subsets of randomized subjects, resulting in possible confounding [16–18]. Specifically, the PP analysis included only the subset of randomized subjects who tested HIV negative at the week 26 visit and adhered to the protocol, resulting in a 24% reduction of the analyzed population (Table 2). To improve on the standard analysis of VE in the PP population, an analytic method that adjusts for measured confounding factors should be applied (eg, like those in [19–23], which are different from standard regression models relating outcome to randomized group and prognostic factors), which in addition to correcting for bias, can improve statistical power by leveraging prognostic factors. Moreover, because some confounding factors may be unmeasured, the sensitivity of results to such factors should also be investigated (eg, [24, 25]).

Fifth, the MITT analysis was prespecified as primary in the final analysis plan prior to study unblinding. This prespecification of the details of the primary analysis is standard practice in clinical trials for ensuring objectivity, and it is not unusual to initially report only primary analyses. Therefore, the consideration of statistical principles resolves the initial confusion about the apparently conflicting RV144 study results.

Interpreting Results Accounting for Other Efficacy Trials. Individual efficacy trials are designed to avoid false-positive results, typically controlling the risk that the results will indicate benefit or harm of a truly useless vaccine (with VE equaling 0%) at 5%. However, if 10 similar efficacy trials are performed, and if all the vaccines are truly identical to placebo, then there is a 40% chance that at least 1 trial will produce a P value $<.05$. This occurs because each trial has a 5% risk of a false-positive result, and the chance of $\geqslant 1$ false-positive result cumulates with the number of trials. For the HIV vaccine field with 4 completed efficacy trials, if in truth the 3 different tested vaccines all had no effect, then (from the negative binomial distribution) there is a 19% chance that $\geqslant 1$ of the trials would yield a P value $<.05$. Therefore, taking the history of trials into account may lead to placing less confidence in the results of a single positive study, and having multiple positive efficacy trials provides more compelling evidence than does a single such trial. The most valuable result of the RV144 study is to encourage future trials.

Bayesian Analysis of Vaccine Efficacy

Unlike the frequentist approach, the Bayesian framework of statistics provides estimates of the probabilities that the VE takes certain values [26, 27]. This approach is intuitively interpretable because it allows for statements such as, "the probability that VE $>$ 0% (ie, that the vaccine has some beneficial effect) is 80%." In the frequentist framework, a proposition is either true or false; in the Bayesian framework, we can speak of the probability that it is true. To produce the latter, the Bayesian approach uses all of the information in the observed data but also requires specification of a prior distribution of VE, which specifies how likely each possible value of VE is based on any beliefs and information one has from outside the experiment at hand. This prior distribution can, alternatively, be set to a default distribution, so that inference is driven only by the internal data. Studies of sensitivity to the choice of prior are a typical component of Bayesian analysis; thus, we also consider a range of priors that are consistent with equipoise [28] and that could reflect the views of the different stake-holders, including the vaccine manufacturer, the sponsor, and study team, and expert scientists with no apparent interest in the trial outcome.

Because in HIV vaccine efficacy trials the null hypothesis (of no efficacy) is scientifically plausible, the Bayesian analysis assigns a prior probability Pr(VE = 0%) to this hypothesis. An obvious choice is Pr(VE = 0%) = .5, so that there is an even chance of zero efficacy and of nonzero efficacy. The remaining

Table 2. Culling of the Modified Intention-to-Treat Population to Form the Per-Protocol Population in the RV144 Trial

Reason for exclusion from the PP population	MITT vaccine (n = 8197)	MITT placebo (n = 8198)
Diagnosed with HIV infection by week 26	5 (0.06%)	10 (0.12%)
Dropped out by week 26 while HIV negative	237 (2.9%)	210 (2.6%)
Reached week 26 visit while HIV negative but was nonadherent to vaccinations (protocol-specified)	1779 (21.7%)[a]	1612 (20.4%)[b]
Total culled out	2021 (24.7%)	1842 (22.5%)

NOTE. Group-imbalances in prognostic factors for human immunodeficiency virus (HIV) infection could arise due to differences (by treatment assignment) in probabilities of any of the events (1) infection, (2) dropout, or (3) nonadherence by week 26. MITT, Modified Intention-to-Treat; (PP) Per-Protocol.

[a] One thousand twenty-nine subjects received <4 doses of vaccine; 742 received all 4 doses, with receipt of ⩾1 dose occurring outside of the window; and 8 were nonadherent for other reasons.

[b] Nine hundred forty-one subjects received <4 doses of vaccine; 670 received all 4 doses, with receipt of ⩾1 dose occurring outside of the window; and 1 was nonadherent for other reasons.

probability of 1-Pr(VE = 0%) is distributed among the nonzero values of VE in some way (examples given below). Note that it is not always scientifically plausible to assign a prior probability to a particular null hypothesis. For instance, if the trial is comparing a radically new vaccine with an existing vaccine with known efficacy of 47%, it may be implausible that the new vaccine would also have precisely 47% efficacy.

After specifying the prior distribution, Bayes theorem (see the online supplement) is applied to convert the prior beliefs about VE into posterior beliefs about VE, given the information in the data ("posterior" means "after" seeing the data). For RV144, the posterior distribution has 2 components:

- Pr(VE = 0% | RV144 data), the posterior probability that the vaccine has no effect; and
- p(VE | RV144 data), the posterior density (likelihood) of the different nonzero values of VE, indicating the likely level of VE.

Bayesian Analysis of RV144

Study Team Prior. The lead statistician of RV144 (Donald Stablein) suggested that, before conducting the trial, the study team members had quite different opinions about Pr(VE = 0%) but had a rough consensus concerning the magnitude of VE if the vaccine were to have an effect. In particular, their prior beliefs were roughly that each *nonzero* value of VE between −20% and 60% was equally likely. For this prior on the nonzero values of VE and supposing Pr(VE = 0%) = .5, Bayes theorem yields Pr(VE = 0%|RV144 data) = .20 (ie, the chance that the vaccine is completely ineffective is 20%). Recall that the P value was .04, which is often erroneously interpreted as strong evidence against the null hypothesis. The fact that there remains a 20% chance that the vaccine is ineffective must be factored into any scientific decisions based on the study results.

The posterior density P(VE|RV144 data) for the nonzero values of VE suggest that, if efficacious, the vaccine efficacy is most likely to be ∼30% (Figure 1). This posterior density can be summarized with a 95% Bayesian CI (typically called a credible interval)– here the interval from 3% to 52% – but this should not be reported in isolation. The overall Bayesian summary is that there is a 20% chance that VE equals 0% (no efficacy) but that, if efficacious, VE lies between 3% and 52% with 95% probability.

Because the opinions concerning Pr(VE = 0%) were quite varied, we present the Bayesian conclusions for a variety of choices of this prior probability in Table 3. Thus the skeptic who assigns prior chance of 90% that the vaccine is ineffective (eg, [2]) will conclude after seeing the RV144 data that there remains a 70% chance that the vaccine is ineffective. Table 3 also shows that the posterior probability that the vaccine is harmful (ie, that VE is <0%) is negligible.

Accounting for Other Efficacy Trials. Primary analyses of trials like RV144 make use of data only from the individual study; however, as mentioned above, more can be learned by placing results in a broader context of other trials of similar agents with similar goals. In Bayesian analysis, one controls for multiple testing by considering that each trial has unknown prior probability Pr(VE = 0%), and then one learns from the trials about this unknown prior probability. Assume that RV144

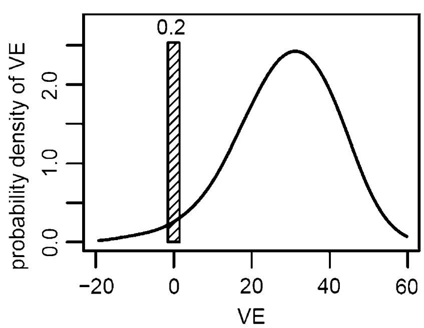

Figure 1. Pr(VE = 0%|RV144 data) and the density P(VE|RV144 data) for nonzero values of vaccine efficacy (VE) when the assumed prior is Pr(VE = 0%) = .5 and Pr(−20% < VE < 60%) = .5 with equal likelihood of all *nonzero* VE values between −20% and 60%.

Table 3. Various Prior and Resulting Posterior Probabilities that the Vaccine Has No Effect (Vaccine Efficacy [VE], 0%), Is Efficacious (VE, >0%), or Is Harmful (VE, <0%)

Pr(VE = 0%) [Prior]	0.1	0.2	0.3	0.4	0.5	0.6	0.61	0.7	0.8	0.9
Pr(VE = 0%|data)	0.03	0.06	0.10	0.14	0.20	0.28	**0.29**	0.37	0.50	0.70
Pr(VE > 0%|data)	0.96	0.93	0.89	0.85	0.79	0.72	**0.71**	0.63	0.50	0.30
Pr(VE < 0%|data)	0.01	0.01	0.01	0.01	0.01	<.01	**<0.01**	<.01	<.01	<.01

NOTE. The posterior probability combines the information from the prior and the data from RV144. The prior probability Pr(VE = 0%) = .61 from the bolded column was estimated based on the 3 previous human immunodeficiency virus vaccine trials.

is viewed as the fourth in a series of related relevant HIV vaccine trials – the first 2 being the efficacy trials of VaxGen's envelope subunit protein with results as reported in Table 1, and the third 'trial' formed by pooling HIV incidence data from the 5 placebo-controlled randomized phase 1/2 trials of prime–boost HIV vaccine regimens containing canarypox (28 infected persons of 1497 enrolled) [29]. Then the unknown Pr(VE = 0%) is estimated to be .61. From Table 3, it would follow that the chance that the RV144 vaccine has some efficacy is 71%.

Using results of other trials to inform the prior for a new trial is complicated by any differences among the trial designs or tested products. RV144 departed from the VaxGen trials in the vaccine regimen, exposure route, balance of male and female participants, and the magnitude of HIV exposure (Table 1). The extent to which these differences affect the Bayesian likelihood of vaccine efficacy introduces uncertainty into the computation of Pr(VE = 0%) for RV144. Nevertheless, this Bayesian analysis provides additional insight by giving one way to account explicitly for the past trial results.

Sensitivity to the Prior. The most arbitrary feature of the study team choice of prior above was constraining VE between −20% and 60%. To study the sensitivity of conclusions to this choice, we consider instead the prior that constrains VE between -VE*/3 and VE* (the largest plausible efficacy before seeing the data) and assumes all nonzero values of VE in this interval are equally likely. We also choose Pr(VE = 0%) to be the adjusted estimate arising from considering the 3 previous relevant vaccine trials; its expression depends on VE* and is omitted here. For VE* varied from 0% to 100%, the resulting posterior distribution is graphed in Figure 2, showing that there is at least a 22% chance that VE = 0% regardless of the choice of VE*.

In conclusion, the Bayesian analysis provides additional information for weighing the evidence about VE than the frequentist analysis alone. Although the frequentist *P*-value of .04 does not inform about the chance that the vaccine had some efficacy, the Bayesian posterior probabilities do, indicating at most a 78% chance that the vaccine is efficacious (.78 equals one minus the smallest posterior probability of no efficacy in Figure 2). In addition, although frequentist statistics only assess data internal to the trial, Bayesian statistics facilitate integration of the internal data with external data, knowledge, and beliefs.

Evaluation of Study Integrity

Some HIV infection events may be unobserved due to missing data on scheduled HIV tests caused by dropout, missed visits, or processing errors. The assessment of VE may be biased by such missed infections even if the missingness rate is the same in the randomized groups, but most severely if the rate differs. A differential rate during the immunization series could stem from vaccine-reactogenicity, and during all periods of follow-up, it could stem from participant unblinding. Moreover, in general, participant unblinding may introduce bias by leading to group-imbalances in HIV exposure. Therefore, assessing rates of HIV tests and of participant blinding are important components of evaluating study validity. We assess these factors for the MITT population of RV144 and use the results in a simple sensitivity analysis of VE.

End Point Ascertainment. After completion of the immunization series, 7212 (88.0%) of 8197 vaccine recipients and 7227 (88.2%) of 8198 placebo recipients were ascertained for HIV infection, either by having a week 26 HIV test result or by having a previous HIV positive infection diagnosis. At the last scheduled visit, the rates of end point ascertainment were 7398 (90.3%) of 8197 vaccine recipients and 7399 (90.3%) of 8198 placebo recipients. Thus, there was a high rate of HIV ascertainment that was not differential between the arms. Furthermore, of the 28,511 possible follow-up years for the vaccine

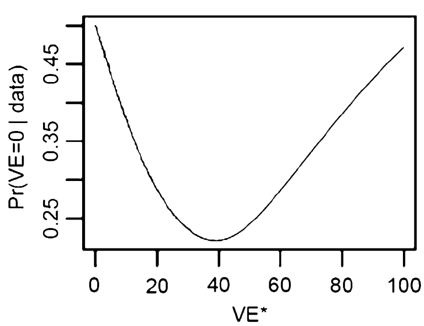

Figure 2. Pr(VE = 0%|RV144 data) as one varies the prior upper limit, VE* (the largest efficacy one might expect before seeing the data), on vaccine efficacy (VE) when the assumed prior has Pr(VE = 0%) adjusted for the three previous HIV vaccine trials and has equal likelihood of all nonzero VE values between –(VE*/3)% and VE*%.

group and 28,434 possible follow-up years for the placebo group, 92.7% were observed for each group. Assuming the arm-pooled 0.24% annual HIV incidence observed in the trial, we expect that 10 infections were missed. An imbalance with 8 in one arm and 2 in the other would lead to VE estimates of 22% or 35%, illustrating the degree of sensitivity of the estimates to the unobserved infections.

Unblinding Ascertainment. Biannual behavioral questionnaires asked about thoughts as to receipt of candidate vaccine/placebo/don't know. At the last visit, 13,495 participants answered 'don't know' and 1301 (7.9%) provided a treatment choice, with 495 (78.8%) of 628 vaccine recipients guessing correctly and 179 (26.6%) of 673 placebo recipients guessing correctly. Therefore, the number of MITT subjects who correctly perceived their treatment assignment is estimated to be $(0.788 \times (628/1301) + 0.266 \times (673/1301) - 0.50) \times 16{,}395 = 296$, a rate of 1.8%. Of the 296 unblinded subjects, we expect that at most 3 became infected, potentially slightly altering the VE estimate to 29%–32%. The estimated correct treatment perception rates were similarly low at other visits, supporting a high rate of blinding. High quality of the blind is particularly important in trials where the end point is caused by behavioral-associated exposures.

CONCLUSIONS

Interpretation of the RV144 results benefits from consideration of statistical principles: the meaning of *P* values and CIs; the distinctions among analyses of VE in the 3 study populations (especially the validity of the MITT analysis versus the bias-prone PP analysis); the impact of data from other efficacy trials; and the uses of Bayesian assessment of probabilities that VE takes certain values. The Bayesian analyses are presented to help understand the RV144 data, but we are not proposing that the particular choices of prior distributions would be used in future trials.

These considerations lead to our conclusion that the RV144 data provide moderate evidence of low-level positive VE – with ≥22% chance remaining for no efficacy under a range of prior assumptions− an inference that reflects greater uncertainty than has much of the discussion about this trial. This uncertainty about the signal, and the fact that multiple positive trials provide more compelling evidence for positive VE than does a single positive trial, support conduct of more efficacy trials of prime – boost candidate vaccines. These trials would benefit from conduct of Bayesian analyses of VE to complement the frequentist analyses, and from conduct of sensitivity analyses to demonstrate how the inferences about VE could be biased due to incomplete ascertainment of HIV infection end points and/or to participant unblinding. Moreover, sensitivity analyses should be included in the assessment of PP VE, and the trial design should seek to minimize the differences in the MITT and PP populations—for example through a pre-randomization run-in period during which subjects demonstrate their ability to adhere.

Supplementary Data

Supplementary data are available at http://jid.oxfordjournals.org/online.

Funding

P.B.G. and S.M.H. were supported by the National Institutes of Health (2R37AI054165-08 and UO1 AI069470, respectively). J.O.B. was supported by National Science Foundation (DMS-1007773). USMHRP (J.H.K., D.M.S.) is supported in part by an Interagency Agreement Y1-AI-2642-15 between U.S. Army Medical Research and Materiel Command (USAMRMC) and the National Institutes of Allergy and Infectious Diseases. In addition this work was supported by a cooperative agreement (W81XWH-07-2-0067) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the U.S. Department of Defense (DOD).

References

1. Rerks-Ngarm S., Pitisuttithum P., Nitayaphan S., et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med **2009**; 361:2209–20.
2. Burton D.R., Desrosiers R.C., Doms R.W., et al. A sound rationale needed for phase III HIV-1 vaccine trials. Science **2004**; 303:316.
3. McNeil J.G., Johnston M.I., Birx D.L., Tramont E.C. HIV vaccine trial justified. Science **2004**; 303:961.
4. Cohen J. Massive AIDS vaccine study a "modest" success. Sci Insider. 24 September 2009.
5. Cohen J. Unrevealed analysis weakens claim of AIDS vaccine "success." Sci Insider. 5 October 2009.
6. Yang Y., Gilbert P., Longini I.M. Jr., Halloran M.E. A bayesian framework for estimating vaccine efficacy per infectious contact. Ann Appl Stat **2008**; 2:1409–31.
7. Flynn N.M., Forthal D.N., Harro C.D., Judson F.N., Mayer K.H., Para M.F. the rgp120 HIV Vaccine Study Group. Placebo-controlled trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J Infect Dis **2005**; 191:654–65.
8. Pitisuttithum P., Gilbert P., Gurwith M., et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. J Infect Dis **2006**; 194:1661–71.
9. Moore J.P., Ho D.D. HIV-1 neutralization: the consequences of viral adaptation to growth on transformed T cells. AIDS **1995**; 9:S117–S136.
10. Rida W., Fast P., Hoff R., Fleming T.R. Intermediate-size trials for the evaluation of HIV vaccine candidate: a workshop summary. J Acquir Immune Defic Syndr Hum Retrovirol **1997**; 16:195–203.
11. Mehrotra D.V., Li X., Gilbert P.B. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial. Biometrics **2006**; 62:893–900.
12. Buchbinder S.P., Mehrotra D.V., Duerr A., et al. 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**; 372:1881–93.
13. Gray G.E., Bekker L., Churchyard G.J., et al. Did unblinding affect HIV risk behaviour and risk perception in the HVTN503/Phambili study? Retrovirology **2009**; 6:P209.

14. Cohen J. Mission improbable: a concise and precise definition of p-value. ScienceNOW Daily News. 30 October 2009.
15. Lachin J.M. Statistical considerations in the intent-to-treat principle. Control Clin Trials **2000**; 21:167–89.
16. Rosenbaum P.R. The consequences of adjustment for a concomitant variable that has been affected by the treatment. J Roy Stat Soc Ser A **1984**; 147:656–66.
17. Robins J.M., Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology **1992**; 3:143–55.
18. Frangakis C.E., Rubin D.B.. Principal stratification in causal inference. Biometrics **2002**; 58:21–9.
19. Tsiatis A.A., Davidian M., Zhang M., Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. Stat Med **2008**; 27:4658–77.
20. Zhang M., Tsiatis A.A., Davidian M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Biometrics **2008**; 64:707–15.
21. Lu X., Tsiatis A.A.. Improving the efficiency of the log-rank test using auxiliary covariates. Biometrika **2008**; 95:679–94.
22. Moore K.L., van der Laan M.J. Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. Stat Med **2009**; 28:39–64.
23. Zhang M., Gilbert P. Increasing the efficiency of prevention trials by incorporating baseline covariates. Stat Commun Infect Dis **2010**; 2: DOI: 10.2202/1948-4690.1002 Available at: http://www.bepress.com/scid/vol2/iss1/art1. Accessed January 31, 2011.
24. Scharfstein D.O., Rotnitzky A., Robins J.M. Adjusting for nonignorable drop-out using semiparametric nonresponse models. J Am Stat Assoc **1999**; 94:1096–146.
25. Shepherd B., Gilbert P., Lumley T. Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. J Am Stat Assoc **2007**; 102:573–82.
26. Spiegelhalter D.J., Abrams K.R., Myles J.P. Bayesian approaches to clinical trials and health-care evaluation. New York: John Wiley & Sons, 2004.
27. Berger J. Could Fisher, Jeffreys and Neyman have agreed on testing (with discussion)? Stat Sci **2003**; 18:1–32.
28. Lilford R.J. Ethics of clinical trials from a bayesian and decision analysis perspective: whose equipoise is it anyway? BMJ **2003**; 326: 980–1.
29. Lee D., Graham B.S., Chiu Y.L., et al. Breakthrough infections during phase 1 and 2 prime-boost HIV-1 vaccine trials with canarypox vectors (ALVAC) and booster dose of recombinant gp120 or gp160. J Infect Dis **2004**; 190:903–7.