Published in final edited form as:

Vaccine. 2009 January 14; 27(3): 396–401. doi:10.1016/j.vaccine.2008.10.083.

Efficient and robust method for comparing the immunogenicity of candidate vaccines in randomized clinical trials

Peter B. Gilbert^{a,*}, Alicia Sato^a, Xiao Sun^b, and Devan V. Mehrotra^b

^a Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, PO Box 19024, Seattle, WA 98109, United States

^b Merck & Co., Inc., UG1CD-44, 351 North, Sumneytown Pike, North Wales, PA 19454, United States

Abstract

In randomized clinical trials designed to compare the magnitude of vaccine-induced immune responses between vaccination regimens, the statistical method used for the analysis typically does not account for baseline participant characteristics. This article shows that incorporating baseline variables predictive of the immunogenicity study endpoint can provide large gains in precision and power for estimation and testing of the group mean difference (requiring fewer subjects for the same scientific output) compared to conventional methods, and recommends the “semiparametric efficient” method described in Tsiatis et al. [Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Stat Med* 2007. doi:10.1002/sim.3113] for practical use. As such, vaccine clinical trial programs can be improved (1) by investigating baseline predictors (e.g., readouts from laboratory assays) of vaccine-induced immune responses, and (2) by implementing the proposed semiparametric efficient method in trials where baseline predictors are available.

Keywords

Immune responses; Statistical analysis; Vaccine trial

1. Introduction

In Phase I and II preventative vaccine trials that randomize subjects to different vaccine regimens, an important objective is to compare the magnitude of vaccine-induced immune responses between groups. Trials of antibody-based vaccines typically compare vaccine-induced neutralization antibody titers among different vaccine formulations, whereas trials of cell-mediated immunity (CMI)-based vaccines typically compare vaccine-induced T cell response levels. For example, Merck Research Laboratories and the HIV Vaccine Trials Network (HVTN) conducted a multi-region, double-blind, dose-escalating study of Merck’s CMI-based replication defective recombinant adenovirus serotype 5 (Ad5) vector vaccine expressing the HIV-1 *gag* gene. A study objective was to evaluate whether recipients of a higher vaccine dose would exhibit greater magnitude T cell responses than recipients of a lower dose, as measured by the HIV-1 Gag-specific interferon-gamma ELISPOT assay, and to estimate the mean difference in response levels. This objective was of particular interest for subjects with pre-existing Ad5 vector immunity, as some studies have suggested that such pre-existing immunity may blunt vaccine-induced CMI responses.

*Corresponding author. Tel.: +1 206 667 7299; fax: +1 206 667 4378. E-mail address: pgilbert@scharp.org (P.B. Gilbert).

For subjects with high baseline Ad5 neutralization titers (defined as titers $> 1:200$), Table 1 shows p -values obtained from four different analytic methods comparing the ELISPOT response levels between the two vaccine dose groups. The p -values vary widely for the different methods, raising the question of what is the most appropriate technique for comparing immune response levels between two randomized vaccine groups? In this report we describe and compare the methods currently in use for HIV-1 vaccine trials, and for vaccine trials in general recommend a new method that is more efficient/powerful and robust.

2. Accounting for baseline predictors improves power and precision compared to methods currently used in HIV-1 vaccine trials

2.1. Methods for comparing immune response levels among vaccine groups

In this section we consider the first three methods in Table 1, deferring discussion of the fourth method, which is more sophisticated, to the next section. The first two methods are currently used in primary analyses to evaluate the immunogenicity of HIV-1 vaccines. Method 1 is a t -test for comparing the “raw readouts” at the Week 8 visit from the assay with Gag antigen present (i.e., the ELISPOTs in response to stimulation with HIV-1 Gag peptide pools), while Method 2 is a t -test for comparing the mock-subtracted readouts, defined as the raw readout minus the readout from the assay performed on the same sample (at Week 8) with Gag antigen absent. Method 3 is an unequal slopes analysis of covariance (ANCOVA), in which a linear regression model is fit with response variable the mean-centered raw readout and independent variables the mean-centered vaccination group indicator, the mean-centered baseline mock readout, and the interaction of these two independent variables. For this method the p -value is obtained based on a Z-statistic, which equals the regression coefficient estimate for the mean-centered vaccination group indicator divided by its square rooted variance estimate. Each method provides a corresponding procedure to estimate the group mean difference in response levels with a confidence interval.

Methods 1, 3, and 4 address a different scientific question than Method 2: the former methods test for a difference in mean values of the raw readout between vaccine groups, whereas the latter method tests for a difference in mean values of the mock-subtracted raw readouts. It is nevertheless still useful to compare their statistical power, as it may inform decisions about the primary objective and analysis method. Moreover, Method 2 addresses the same scientific question as the other methods for cases where the mean value of the mock readout may be assumed equal in the two vaccine groups.

Methods 3 and 4 are motivated by the well-studied fact that accounting for baseline participant characteristics predictive of the response variable can improve power and precision (e.g., [1]). The regression models used in these methods may include any participant characteristics measured at baseline (at or before randomization). However, participant characteristics measured after baseline should usually be excluded, since including them will bias the results if there are unmeasured baseline participant characteristics that predict both the post-baseline covariate and the response variable [2]. Randomization ensures that it is valid to condition on baseline – but not on post-baseline – covariates.

However, in some trials an assumption that the mock readout is independent of randomization assignment may be justified, in which case Methods 3 and 4 may adjust for the mock readout. This may substantially improve power because the mock readout may have a relatively strong association with the raw readout. The mock readout should only be adjusted for after a thorough defense of the independence assumption, which would involve biological knowledge of the assay and vaccines as well as a direct check of the assumption from the trial data. Furthermore, the independence assumption can only be justified if most study volunteers are evaluable for

the primary endpoint and evaluability is plausibly independent of the randomization assignment.

2.2. Comparative power of the methods

To provide a simple comparison of power of Methods 1–3, we suppose the baseline mock (BM), Week 8 mock (M), and Week 8 raw (R) measurements have the same standard deviation (variable notation is listed in Appendix A). The relative power depends on the within-subject correlation of BM and R (ρ_{BMR}) and on the within-subject correlation of M and R (ρ_{MR}). As described mathematically in Appendix A, the variance of the raw readout is less than that of the mock-subtracted readout if and only if $\rho_{MR} < 0.5$. This gives the rule of thumb that Method 2 will tend to provide greater statistical power than Method 1 if M and R are positively correlated at level at least one-half, and less statistical power if their correlation is weaker than one-half. For the Merck/HVTN trial data, the Pearson correlation estimates of ρ_{MR} were 0.32, 0.25, and 0.34 for geographic region sets 1, 2, and 3 in Table 1, explaining why Method 1 tended to give smaller p -values. The “tipping point” for what correlation value ρ_{MR} makes Method 1 superior or inferior to Method 2 is modified if M and R have different variances, as delineated in Appendix A.

In addition, if the baseline mock and raw readouts are correlated ($\rho_{BMR} \neq 0$), then the relevant variance for the ANCOVA method (that of the raw readout conditional on the baseline mock readout) is smaller than the variance for Method 1. Consequently, in general ANCOVA tends to provide at least as much power as Method 1, and greater power if BM and R are positively correlated, with the extent of power gain increasing with the strength of the correlation. In addition, the variance for the ANCOVA method is smaller than that for Method 2 if

$\rho_{BMR}^2 > 2\rho_{MR} - 1$. Therefore even if BM and R are uncorrelated, the ANCOVA method will tend to give greater power than Method 2 if M and R have correlation less than 0.5. Moreover, if M and R are highly correlated, then the ANCOVA method may still be more powerful if BM and R are also highly correlated. For the Merck/HVTN trial the Pearson correlations of BM and R were all positive (0.18, 0.04, and 0.15 for region sets 1, 2, and 3), explaining why ANCOVA gives smaller p -values than Method 1.

Fig. 1 shows the results of a simulation study designed to demonstrate how much power is gained by the ANCOVA method (and the efficient Method 4 described below) as a function of ρ_{BMR} . For this experiment, the values (BM , M , and R) for each vaccine recipient were generated from a normal distribution assuming a common standard deviation of 0.5 for BM , M , and R , to approximately reflect the real ELISPOT data. Each of the correlations ρ_{MR} and ρ_{BMR} were set to 0.3 or 0.7 (weak or moderate). The means of BM and M were set to 1.4 for each vaccine group. The mean of R was set to 2.1 for group 1 and to $2.1 + \delta$ for group 2, with δ ranging from 0 to 0.4. The results are not affected by the choice of means for BM , M , and R ; for these multivariate normal data the power depends only on the standard deviations, correlations, and the effect size δ . Ten thousand vaccine trials were simulated with sample sizes $N = 50$ vaccine recipients per group. The power of each method was estimated by the fraction of trials that rejected the null hypothesis of equal mean response levels R at two-sided $\alpha = 0.05$. Fig. 1 confirms the rule of thumb that ρ_{MR} greater (less) than 0.5 leads to Method 2 beating (losing to) Method 1. It also shows that ANCOVA and Method 4 perform almost identically, and are generally more powerful than Methods 1 and 2, with power gain increasing with ρ_{BMR} .

3. Beyond ANCOVA to a more efficient and robust method

ANCOVA assumes a linear relationship between R and BM , and a normal distribution for R conditional on BM . For trials with large sample sizes ANCOVA provides valid/unbiased

estimation and testing even if the linearity and normality assumptions are violated, but at moderate sample sizes encountered in practice violations of these assumptions can lead to invalid inferences. In addition, for any sample size nonlinearity can make ANCOVA have relatively low power and precision. Moreover, if either the linearity or normality assumptions are violated, then the ANCOVA method can give invalid results if some subjects are missing data on R .

To address these potential limitations, we recommend the “semiparametric efficient” (SP-EFFIC) method developed by Refs. [3–5]. The SP-EFFIC estimator of the mean difference of R is “efficient” in that it has the smallest possible variance in large samples, and “semiparametric” refers to the fact that this optimality is derived using the framework of semiparametric theory. We refer the reader to Appendix A for details on its mathematical implementation, and to Refs. [4,5] for more complete mathematics. In this section we summarize its benefits for vaccine trials and describe how to apply it in non-mathematical terms. The SP-EFFIC method is very similar to the ANCOVA Method 3 if there are no missing response data and all the baseline covariates that are adjusted for are categorical (it gives identical results in large samples), but otherwise the SP-EFFIC method is different and may provide improvements.

3.1. Proposed SP-EFFIC method

The SP-EFFIC method proceeds in two steps:

Step 1: For each vaccine group separately, fit a parametric model relating R to BM and to any other baseline covariates that help predict R .

Step 2: Eqs. (1) and (2) in Appendix A, which incorporate the models built in Step 1, provide the estimate of the group mean difference in R and its estimated variance. These estimates provide a standard Z-statistic based p -value for testing for a group mean difference and a confidence interval for the difference.

In Step 1, any parametric model may be used that provides estimates of the expected value of R given BM and other baseline covariates for each vaccine recipient [5]. For example, if BM is the only baseline covariate, then a simple linear regression model or a polynomial model with quadratic or higher order powers of BM could be used. If a large number of baseline covariates are available, then an automated model selection procedure could be used, such as backwards elimination, all-subsets, or LASSO. The goal of Step 1 is to develop good-fitting models, as better models protect against bias at small sample sizes and can provide greater power. For validity the model building in Step 1 must be performed independently of the results that are obtained in Step 2 [5].

3.2. Comparative power and precision of the methods, continued

For the same simulation design used for Fig. 1, Fig. 2 shows the relative efficiency (ratio of sample variances of the mean difference estimates across the 1000 trials) of Methods 2, 3, and 4 compared to Method 1. The relative efficiency measures the fold-difference in sample size that is needed for the two methods being compared to have the same power to detect the same group difference in mean immunological response levels. For example, when $\rho_{MR} = 0.3$ and $\rho_{BMR} = 0.7$, the relative efficiency of Method 4 versus Method 1 is 0.5, implying that half the number of subjects are needed with Method 4. Fig. 2 shows that the SP-EFFIC method is more powerful than Method 1 if $\rho_{BMR} > 0$ and is more powerful than Method 2 unless ρ_{MR} is considerably greater than ρ_{BMR} . Fig. 2 can be used for selecting the most powerful method for a particular vaccine trial, as it facilitates comparing trial costs under each of the methods, accounting for knowledge of the correlations ρ_{MR} and ρ_{BMR} . Furthermore, the simulations are

informative about the relative precision of estimation, because the square root of each value in Fig. 2 measures the ratio of confidence interval width for Methods 2, 3, and 4 compared to Method 1. As such, Fig. 2 shows that the SP-EFFIC method gives more precise estimates than Method 1 if $\rho_{BMR} > 0$; for example if $\rho_{BMR} = 0.7$, then its confidence intervals are 29% narrower ($0.29 = 1 - 0.5^{1/2}$).

As a rule of thumb, if simple linear regression models are used in Step 1 with a single quantitative baseline covariate, then the SP-EFFIC method requires $(1 - \rho_{BMR}^2)$ -fold fewer subjects than the standard two-sample t -test. In the general case that multiple baseline predictors are used, then the SP-EFFIC method requires approximately $(1 - r^2)$ -fold fewer subjects than the t -test, where r^2 is the coefficient of determination (i.e., the percentage of the variation in the response explained by the baseline covariates).

The simulation study generated data such that the ANCOVA modeling assumptions are true, explaining why the ANCOVA method provides as much power as the SP-EFFIC method (the most favorable case for ANCOVA). In simulation experiments of scenarios where the linearity assumption of ANCOVA fails, Refs. [5–7] demonstrated that the SP-EFFIC method tends to give correct inferences whereas ANCOVA can mislead, giving false positive rates much smaller or larger than the nominal false positive α -level, and the SP-EFFIC method often gives much greater power. We verified this in supplementary simulations (results not shown). This suggests that in practice the SP-EFFIC method is generally preferred to ANCOVA.

3.3. Accommodating missing data

If some vaccine recipients are missing R , then the SP-EFFIC method is especially useful. With missing data ANCOVA is only valid if the normality and linearity assumptions are true, whereas the SP-EFFIC method is valid without requiring these assumptions, as long as a reasonably good model predicting whether R is observed can be constructed (if not, then ANCOVA may perform better with missing data if its assumptions hold). Specifically, the SP-EFFIC method weights subjects by the reciprocal of the estimated probability that R is observed given their baseline covariates, and possibly additional covariates measured after baseline. These estimated probabilities may be obtained, for example, by first fitting a logistic regression model with response variable R . Second, for each subject the estimated probability is taken to be the predicted value from this model based on his/her covariates. For validity, this approach requires the assumption that the data are missing at random, which states that the probability that R is observed depends only on available data.

Davidian et al. [4] described how to use the SP-EFFIC method with missing data. In simulation studies with missing data, Ref. [6] showed that the SP-EFFIC method tends to give correct inferences as long as the missing data model is not badly mis-specified, whereas in contrast the ANCOVA method tends to give misleading results if the linearity assumption fails. As such, for vaccine trials with significant rates of missing immune responses R and with data on participant characteristics that predict whether R is missing, the SP-EFFIC method is generally recommended. For some data sets, however, the SP-EFFIC method may provide unstable estimates and tests because some of the estimated weights are very large. In such cases, which can be diagnosed by examining the estimated weights, ANCOVA is recommended.

4. Conclusions and recommendations

In this note, we considered the choice of statistical method for comparing immune response levels among candidate vaccine groups in Phase I/II randomized vaccine trials. Making this comparison is frequently the primary objective of a Phase I/II vaccine trial. We summarize our findings and recommendations for using the newly proposed SP-EFFIC method for addressing

this objective, and for augmenting Phase I/II vaccine trials with explorations of predictors of immunogenicity:

1. If baseline covariates are collected in the trial that predict the immune response raw readout R , then incorporating them into the analysis through ANCOVA or the SP-EFFIC method will generally provide a more efficient/powerful analysis than standard methods that ignore baseline covariates. The ANCOVA and SP-EFFIC methods provide hypothesis tests for whether there is a difference in mean immune response levels between vaccine groups, and provide point and confidence interval estimates for the group difference in mean response levels (or for the group ratio of geometric means).
2. In general the SP-EFFIC method is recommended over ANCOVA because it provides at least as much statistical power/precision and often more. These two methods have comparable power if the relationship between R and baseline covariates is linear in each vaccine group. If there is nonlinearity in either relationship, then the SP-EFFIC method tends to be more powerful, with the amount of power gain increasing with the degree of nonlinearity. The source of the power gain is more accurate modeling in Step 1 of the method.
3. As a rule of thumb, the SP-EFFIC method requires approximately $(1 - r^2)$ -fold fewer subjects than the t -test, where r^2 is the coefficient of determination of the models fit in Step 1. For example, if baseline covariates explain one-third of the variation ($r^2 = 0.33$), then the SP-EFFIC method requires approximately two-thirds the number of subjects as the t -test.
4. If there are significant rates of missing immune response data R , and there are covariates (baseline and/or post-baseline) that predict whether R is missing, then the SP-EFFIC method is generally recommended. This demonstrates value of investigations into predictors of missed visits, study drop-out, lost specimens, and failure of the immunological assay to measure R . A caveat of the SP-EFFIC method for missing data is that for some data sets the weighting procedure can create instability in the estimation, yielding imprecise inferences. In such cases, which can be diagnosed by assessing whether some of the estimated weights are very large (outliers), ANCOVA is recommended.
5. In the absence of baseline covariates that predict the immune response and the absence of covariates that predict whether the immune response is missing, a standard two-sample test such as the t -test is recommended. The within-vaccine recipient correlation between raw and mock values determines which comparison (raw readouts or mock-subtracted readouts) can be done with greater power.
6. For the SP-EFFIC method the model-building step (Step 1) must be done independently of the estimation and testing step (Step 2). This prevents unscrupulous “fishing” wherein multiple Step 1 models are tried and the one that leads to the smallest p -value in Step 2 is selected and reported. Tsiatis et al. [5] discuss practical approaches to separating the two steps, for example by using independent analysts or by pre-specifying an automated model selection procedure for Step 1 in the protocol or analysis plan. The latter approach ensures that the same result would be obtained by any analyst.
7. A consequence of our findings is that vaccine trial programs may be improved (both in cost-efficiency and in capacity to provide scientific insights) by expanding investigations into predictors of immunogenicity. For example, demographic factors, results of *in vitro* assays, immune responses to the candidate vaccine with the insert removed, immune responses to related vaccines, and host genetics, may be explored

for their ability to predict immunogenicity. Discovery of immunogenicity predictors may also lend biological insights into the process of vaccine development.

Lastly, whereas we focused on the problem of comparing a quantitative immune response variable between vaccine groups, some vaccine trials use a dichotomous endpoint, such as whether a vaccine recipient's response increases at least fourfold compared to pre-vaccination levels. Fortunately, the SP-EFFIC approach is quite general, applying to many endpoint types, and Ref. [7] described its use for a dichotomous endpoint. In simulation studies they showed its superior statistical power/precision compared to standard methods that evaluate the unadjusted odds ratio or that adjust for baseline covariates using logistic regression.

Computer code for implementing the SP-EFFIC method for a quantitative immune response endpoint is available at <http://faculty.washington.edu/peterg/>.

Acknowledgments

The example data used in this manuscript were collected in protocol 018/050 conducted by Merck Research Laboratories and the HIV Vaccine Trials Network. The authors are grateful to the protocol 018/050 study investigators and participants. This work was partially supported by NIH grant 1 U01 AI068635-01.

References

1. Egger MJ, Coleman MC, Ward JR, Reading JC, Williams HJ. Uses and abuses of analysis of covariance in clinical trials. *Control Clin Trials* 1985;6:12–24. [PubMed: 3987300]
2. Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J R Stat Soc, Ser A* 1984;147:656–66.
3. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc* 1994;89:846–66.
4. Davidian M, Tsiatis AA, Leon S. Semiparametric estimation of treatment effect in a pretest–posttest study with missing data. *Stat Sci* 2005;20:261–301. [PubMed: 19081743]
5. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Stat Med* 2007;27:4658–77. [PubMed: 17960577]
6. Davidian, M. Bernard Greenberg Lecture II. Chapel Hill: University of North Carolina; May. 2006 Semiparametric estimation of treatment effect in a pretest–posttest study.
7. Zhang M, Tsiatis AA, Davidian M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics* 2008;64:707–15. [PubMed: 18190618]

Appendix

Appendix A

Variable notation and variance formulas:

BM = baseline mock measurement (immune response readout in assay without the antigen of interest, i.e., negative control response)

M = Week 8 mock measurement

R = Week 8 raw measurement (antigen-specific immune response readout of interest, often the primary endpoint of the Phase I/II trial)

σ_{BM} = standard deviation of BM

σ_M = standard deviation of M

σ_R = standard deviation of R

ρ_{BMR} = linear correlation between BM and R

ρ_{MR} = linear correlation between M and R

Assuming BM , M , and R are jointly normally distributed, the following results attain:

Method 1: $Var(R) = \sigma_R^2$

Method 2: $Var(R - M) = \sigma_M^2 + \sigma_R^2 - 2\rho_{MR}\sigma_M\sigma_R$

Method 3: $Var(R|BM) = \sigma_R^2(1 - \rho_{BMR}^2)$

The “tipping point” at which Methods 1 and 2 provide the same power occurs when $\rho_{MR} = 0.5$ (σ_M/σ_R). To assess which method is expected to be the most powerful for a vaccine trial, each parameter in the above variance formulas can be estimated from available data. The ratio of variance estimates for two methods measures the ratio of the number of subjects needed to achieve the same power for detecting a given difference in mean immune response levels.

Implementation of the SP-EFFIC method

We describe how to implement Steps 1 and 2 described in Section 3.1. This description extracts key results from Tsiatis et al. [5], which is a very readable and accessible supplement. We use the notation of Ref. [5].

Suppose there are n_1 (n_0) subjects randomized to vaccine group 1 (0), and set $n = n_1 + n_0$. Let Z be the randomization group (0 or 1), X be a vector of baseline characteristics, and Y be the response variable ($Y = R$, the raw readout). The i th subject, $i = 1, \dots, n$, contributes the data-points (Z_i , X_i , and Y_i). The goal is to estimate β from these n data-points, where β is the mean difference in Y : $\beta = E[Y|Z = 1] - E[Y|Z = 0]$.

The first key result from Ref. [5] is that all reasonable estimators of β (i.e., regular estimators that are consistent and asymptotically normal) either can be expressed exactly as or equivalent to in large samples, an expression of the form:

$$\bar{Y}^{(1)} - \bar{Y}^{(0)} - \sum_{i=1}^n (Z_i - \bar{Z}) \{n_0^{-1} h^{(0)}(X_i) + n_1^{-1} h^{(1)}(X_i)\} \quad (1)$$

where $\bar{Y}^{(1)}$ ($\bar{Y}^{(0)}$) is the sample average of the $Y^{(1)}$ s for subjects in group 1 (0), $\bar{Z}$ is the sample average of the Z s pooled over the two groups, and $h^{(1)}(X)$ and $h^{(0)}(X)$ are arbitrary scalar functions of X .

The second key result from Ref. [5] is that the optimal estimator of β , that is the estimator with the smallest variance in large samples, is achieved by setting

$$h^{(1)}(X_i) = E[Y_i|Z_i=1, X_i] \text{ and } h^{(0)}(X_i) = E[Y_i|Z_i=0, X_i]$$

these are the true regression relationships of Y on X for each vaccine group separately. If these true relationships are linear and the same in the two groups, then the optimal estimator is

equivalent to the ANCOVA estimator of Method 3. Otherwise the SP-EFFIC estimator is more efficient. Because the true regression relationships are unknown, Step 1 for implementing the SP-EFFIC method is to develop separate parametric models for $E[Y|Z = 1, X]$ and $E[Y|Z = 0, X]$. For each subject i , the first model can be used to compute a predicted Y value, $\hat{f}_{1,i}$ and the second model can be used to compute a predicted Y value, $\hat{f}_{0,i}$

In Step 2, β is estimated by plugging the predicted values into (1):

$$\hat{\beta} = \bar{Y}^{(1)} - \bar{Y}^{(0)} - \sum_{i=1}^n (Z_i - \bar{Z}) \{n_0^{-1} \hat{f}_{0,i} + n_1^{-1} \hat{f}_{1,i}\}$$

Next, the variance of $\hat{\beta}$ is estimated by

$$\begin{aligned} \widehat{Var}(\hat{\beta}) = & C \sum_{i=1}^n [\{n_1^{-1} Z_i - n_0^{-1} (1 - Z_i)\} Y_i - n^{-1} \hat{\beta} - (Z_i - \bar{Z})(n_0^{-1} \hat{f}_{0,i} \\ & + n_1^{-1} \hat{f}_{1,i}) - (Z_i - \bar{Z})\{n_0^{-1} (\bar{Y}^{(0)} - \bar{f}_0) + n_1^{-1} (\bar{Y}^{(1)} - \bar{f}_1)\}]^2 \end{aligned} \quad (2)$$

where $\bar{f}_1$ is the average of the $\hat{f}_1$ s for subjects in group 1, $\bar{f}_0$ is the average of the $\hat{f}_0$ s for subjects in group 0, $C = \{(n_0 - p_0 - 1)^{-1} + (n_1 - p_1 - 1)^{-1}\} / \{(n_0 - 1)^{-1} + (n_1 - 1)^{-1}\}$, and p_1 (p_0) is the number of parameters (not counting the intercept) fit in the regression model for group 1 (0). A 95% confidence interval for β is then computed as $(\hat{\beta} - 1.96 \{\widehat{Var}(\hat{\beta})\}^{1/2}, \hat{\beta} + 1.96 \{\widehat{Var}(\hat{\beta})\}^{1/2})$. The test for $\beta \neq 0$ is conducted by comparing $\hat{\beta} / \{\widehat{Var}(\hat{\beta})\}^{1/2}$ to a standard normal distribution.

Fig. 1.

Power of Methods 1, 2, 3, and 4 versus the mean difference in raw readout R , for true correlations: $\rho_{MR} = 0.3, \rho_{BMR} = 0.3$; $\rho_{MR} = 0.3, \rho_{BMR} = 0.7$; $\rho_{MR} = 0.7, \rho_{BMR} = 0.3$; and $\rho_{MR} = 0.7, \rho_{BMR} = 0.7$.

Fig. 2. Relative efficiencies of Methods 2, 3, and 4 compared to Method 1, for true correlation ρ_{BMR} ranging from 0 to 0.9 and true correlations: $\rho_{MR} = 0.3$ and $\rho_{MR} = 0.7$.

Table 1

p-Values for comparing Week 8 HIV-1 Gag-specific ELISPOT responses between the low and high vaccine dose groups in the Merck/HVTN Phase I collaborative trial^a for the stratum with prior Ad5 vector immunity (defined as baseline neutralization titers > 1:200).

Method	Geographic regions ^b		
	Southern Africa + U.S. + South America + Caribbean (<i>n</i> = 93)	South America + Caribbean (<i>n</i> = 59)	U.S. + South America (<i>n</i> = 57)
1. <i>t</i> -Test raw readout	0.023	0.12	0.10
2. <i>t</i> -Test mock-subtracted readout	0.13	0.34	0.93
3. ANCOVA	0.013	0.11	0.065
4. SP-EFFIC ^c	0.008	0.079	0.039

^aThe study administered three doses of recombinant Ad5 vector HIV-1 gag vaccine at Weeks 0, 4, and 26.

^bSouthern African sites include South Africa and Malawi. South American sites include Brazil and Peru. Caribbean sites include Haiti, Puerto Rico, and the Dominican Republic. The Thailand region was excluded from this illustrative analysis because no association between baseline Ad5 neutralization titers and HIV-1 Gag-specific ELISPOT response was apparent.

^cStep 1 of the SP-EFFIC method used a simple linear regression model relating *R* to *BM* in each of the two vaccine groups.