SOME STATISTICAL ISSUES IN HIV VACCINE TRIALS*

WASIMA N. RIDA
Biostatistics Research Branch, Division of AIDS, National Institute of Allergy and Infectious Diseases,
6003 Executive Blvd., Bethesda, MD 20892, U.S.A.

AND

DALE N. LAWRENCE
Vaccine Trials and Epidemiology Branch, Division of AIDS, National Institute of Allergy and Infectious Diseases,
6003 Executive Blvd., Bethesda, MD 20892, U.S.A.

SUMMARY

Efficacy trials of prophylactic HIV vaccines will be among the most difficult clinical trials ever attempted. Not only will there be challenges with the recruitment and retention of high-risk uninfected individuals, there will be many statistical challenges to the design, conduct, analysis, and interpretation of these trials. General features of an efficacy trial are described, including choice for the primary endpoint and testing for and estimating vaccine efficacy. Secondary objectives of trials are also discussed. These include determining the correlates of protective immunity, assessing the impact of HIV genetic variation on vaccine efficacy, and using biological markers such as viral load and CD4+ lymphocyte cell count to gain insight on a vaccine's ability to prevent or delay disease. The use of biological markers as surrogates for disease outcome is discussed. Last, trial designs for studying several candidate vaccines or other HIV prevention strategies in a single trial are examined.

1. INTRODUCTION

More than 11 million human immunodeficiency virus (HIV) infections are thought to have occurred worldwide since the beginning of the epidemic.1 If present trends continue, 30 to 40 million cumulative infections will take place by the year 2000. Ninety per cent of these infections will have occurred in developing countries. In the United States, a minimum of 40,000 new infections are expected to occur each year.2 While education and other interventions are critical, a safe and effective vaccine may be the best hope for controlling the pandemic.

Experiments in chimpanzees and monkeys demonstrate that immunization can protect these animals from viral challenge under certain experimental conditions. Furthermore, preliminary safety and immunogenicity studies in humans show candidate vaccines to be well tolerated and capable of inducing immune responses. With these encouraging results, public health agencies are now planning for the conduct of large-scale efficacy trials in high-risk uninfected individuals in the United States and in developing countries.

This paper discusses some of the many challenges to be faced in the design, conduct and analysis of these efficacy trials. Section 2 discusses approaches to HIV vaccine development and

* Adapted from a chapter entitled 'Prophylactic HIV vaccine trials' which shall appear in AIDS Clinical Trials,
Finkelstein, D. and Schoenfeld, D. (eds), Wiley-Liss

CCC 0277-6715/94/202155-23
© 1994 by John Wiley & Sons, Ltd.
how one might decide when a candidate vaccine is ready for large-scale efficacy testing. Results from primate studies and phase I and II safety and immunogenicity trials are briefly reviewed. In Section 3, the general features of a phase III efficacy trial are described. These include choice for the primary endpoint, sample size considerations, and testing for and estimating vaccine efficacy. Section 4 discusses secondary objectives of vaccine trials. These include determining the correlates of protective immunity, assessing the impact of HIV variation on vaccine efficacy, and using virologic and immunologic measures to gain insight on a vaccine's ability to prevent or delay disease. The last section examines other trial designs for studying several candidate vaccines or other prevention strategies in a single trial.

2. HIV VACCINE DEVELOPMENT

2.1. Approaches to vaccine development

Table I lists the various types of vaccines currently under study in primates and humans. They range from live attenuated vaccines, which present the entire virus (minus certain regulatory genes) to the immune system, to peptide vaccines which are small pieces of viral proteins. In between are whole inactivated vaccines, pseudovirions, live recombinant vector vaccines, and recombinant subunit vaccines. Other vaccines include such approaches as DNA vaccines.

While a vaccine that contains most of the virus is likely to be the most immunogenic, there are serious safety concerns with the use of a live attenuated or whole inactivated vaccine in humans. A live attenuated HIV vaccine may revert to a virulent form in the vaccinee by natural mutation of the virus or by recombination with pathogenic HIV acquired after immunization. With a whole inactivated vaccine, the inactivation process may not kill all viruses, thus exposing the vaccinee to live virus. Such was the case with some polio vaccines distributed shortly after the conclusion of the 1954 Salk vaccine trial. Of the 400,000 people inoculated with certain preparations, 79 contracted polio. Another 125 individuals became infected through contact with those vaccinated. Three-quarters of these cases involved paralysis and 11 cases were fatal. This unfortunate experience, widely known as the Cutter incident, led to new techniques of viral inactivation and provided a sobering lesson in exercising extraordinary care with inactivation procedures.

Because of safety concerns, ‘first generation’ vaccines have focused on recombinant subunits. These are genetically engineered viral proteins that are mostly based on the envelope glycoprotein precursor, gp160, or the glycoprotein gp120. These products contain no viral RNA and thus are not infectious to the vaccinee. Live recombinant vector vaccines are also among first generation vaccines. Here, the genes that code for certain HIV proteins, such as envelope proteins, are inserted into viruses such as vaccinia or canarypox. HIV viral proteins are produced when the vector infects human cells. The vector itself may or may not replicate. Canarypox, for example, does not replicate in human cells. Vaccinia, however, replicates briefly until the immune system clears it. Another approach to immunization combines recombinant subunits with recombinant vectors. A live recombinant vector is used to prime the vaccinee's immune system while a recombinant subunit is used to boost it.

‘Second generation’ vaccines include pseudovirions and peptide vaccines. Pseudovirions are virus-like particles that contain no genetic material. Envelope and core proteins are genetically engineered and then combined in such a way that the resulting particles closely approximate the three-dimensional structure of HIV. Peptide vaccines, on the other hand, are genetically engineered amino acid sequences that correspond to important epitopes of viral proteins. One such epitope is the third hyper-variable (V3) region of gp120 which contains the principal neutralizing
Table I. HIV vaccines in development

<table>
<thead>
<tr>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live attenuated virus</td>
</tr>
<tr>
<td>Whole inactivated virus</td>
</tr>
<tr>
<td>Pseudovirions</td>
</tr>
<tr>
<td>Live recombinant vectors</td>
</tr>
<tr>
<td>Recombinant subunits</td>
</tr>
<tr>
<td>Peptide vaccines</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

domain of HIV. A combination of peptides based on different strains of HIV also represents the first attempt at a multivalent vaccine.⁴

Vaccines in subsequent generations may include such novel approaches to HIV immunization as DNA vaccines. Recently, Fynan et al.⁵ demonstrated the ability of plasmid DNAs expressing influenza virus haemagglutinin glycoproteins to induce protective immunity in mice and chickens against lethal influenza challenges of the same subtype. Liu and Vogel⁶ discuss the potential use of DNA vaccines against HIV infection in humans. However, as promising as DNA vaccines may be, there are safety concerns with injecting naked DNA into humans. Theoretically, vaccine DNA might integrate into the host cell chromosome and cause the expression of oncogenes which could lead to tumour formation.

The most rational approach to vaccine development relies on the development of an animal model of HIV infection and disease.⁷ Investigators would then identify vaccines that are effective in preventing infection in animals and determine which immune responses correlate with protection. Analogous vaccines would then be screened in humans and those producing immune responses that correlate with protection in animals would be candidates for efficacy trials.

Work to date in monkeys, chimpanzees, and other animals has been encouraging, but investigators have not found an ideal animal model. Chimpanzees, for example, can be infected with HIV-1, but they do not develop disease. Monkeys, however, show promise as a model. While rhesus macaques cannot be infected with HIV-1, they can be infected with simian immunodeficiency virus (SIV). SIV is genetically similar to HIV and causes a disease in these animals that is similar to HIV disease in humans. Also, recent work with pigtail macaques has shown that these monkeys can be infected with HIV-1, raising hope for the direct evaluation of HIV vaccines in macaques.⁸ Another recent development has been the successful insertion of the HIV-1 envelope gene into an SIV genetic background to produce a chimeric virus that can replicate in macaques.⁹ With this approach, one could study the breadth of protection in vaccinated animals by substituting the many variant HIV-1 envelope genomes into the SIV genetic background.

Given the urgency of the AIDS pandemic, one may not wait until animal studies determine the correlates of protective immunity. In this case, one might accelerate an accelerated clinical trials programme. Candidate vaccines shown to be safe in humans and capable of inducing broadly reactive immune responses in vitro against a wide spectrum of HIV strains would then be considered for efficacy testing. However, one may want to strike a balance between the need for more information on the mechanisms of protection and the urgency of the growing pandemic. Systematic vaccine development might be stressed based on the currently available information on HIV pathogenesis and animal protection studies. Phase I and II clinical trials of the safety and immunogenicity of candidate vaccines would be conducted while
animal studies are going on. Thus, the primary goal would be to obtain as quickly as possible the critical information from animal and human studies to prioritize candidate vaccines for efficacy testing.

It may be difficult to devise strict criteria for deciding when a candidate vaccine is ready for phase III efficacy testing. At a minimum, it seems reasonable to require evidence that the vaccine prevents infection in animals to some degree and that it induces immune responses in humans similar to those induced in animals. One would not necessarily know the correlates of protection before starting phase III trials. Furthermore, phase I and II trials would need to demonstrate adequate safety of the candidate vaccine.

2.2. Primate studies

Vaccines have been found that prevent infection in chimpanzees and monkeys when the animals are challenged with live virus (see Schultz and Hu\textsuperscript{10} and the references cited therein). Chimpanzees, immunized with subunit vaccines based on gp120 or gp160, have been protected against challenge with HIV-1.\textsuperscript{11,12} Moreover, protection appears to be associated with a higher titre of neutralizing antibodies induced in the chimpanzee at the time of challenge. Further evidence of the protective role of neutralizing antibodies stems from the passive immunization of chimpanzees with HIV-1 antibodies derived from human sera.\textsuperscript{13}

Rhesus macaques, on the other hand, have been protected against challenge with SIV by whole inactivated SIV vaccines. Early studies of whole inactivated vaccines were problematic because the vaccine as well as the virus used to challenge the monkeys was grown in human cells. Consequently, these preparations contained human cellular antigens to which the animals responded. Protection was thus partially attributable to the immune recognition of the human antigens found in the vaccine and challenge virus. However, recent studies with whole inactivated SIV vaccines which are not confounded by the cellular antigen problem have demonstrated protection. In one study, all four control monkeys became infected after challenge with cells from a moribund SIV infected macaque while four of eight immunized animals were protected.\textsuperscript{14,15} In another study, no protection was observed.\textsuperscript{16} However, a significantly larger challenge dose of SIV was used.

Rhesus macaques have also been protected by live vaccinia virus engineered to contain the gene that expresses SIV envelope protein followed by injection with purified envelope protein to further promote an immune response.\textsuperscript{17,18} Peptide vaccines have also afforded some protection in rhesus macaques,\textsuperscript{19} and recent experiments with live attenuated vaccines have demonstrated impressive protection against SIV challenge. In particular, non-pathogenic SIV in which the regulatory gene nef has been deleted causes an infection in macaques that induces a serological response resembling that of active infection.\textsuperscript{20} Furthermore, four of four immunized animals were protected from challenge with pathogenic SIV.

Studies involving macaques have also suggested that vaccines may prevent or delay disease. In one experiment, five macaques immunized with whole, inactivated SIV became infected when challenged with pathogenic SIV, but three remained healthy 18 months later.\textsuperscript{21,22} In contrast, six unvaccinated controls became infected and developed serious opportunistic infections within a few months. Similar observations were made by Shafferman and colleagues\textsuperscript{23} in macaques immunized with peptide vaccines. Furthermore, the amount of virus in blood of immunized monkeys that became infected appears to have been reduced. In studies of 23 such monkeys, 'viral load' was found to be 10 to 20 per cent of the amount of SIV found in control animals.\textsuperscript{24} It remains to be shown, however, that reduced viral load can delay or prevent the onset of disease. Unfortunately, many animals were sacrificed before clinical outcome
could be determined because of limited space and funds to house these animals for long-term study.

2.3. Preliminary trials in humans

To date, more than 20 candidate HIV vaccines have been entered into phase I clinical trials to assess their safety and immunogenicity in low-risk uninfected individuals (see Fast and Walker\textsuperscript{24} and the references cited therein). These randomized, controlled studies have ranged in size from a handful to 128 volunteers. They have investigated different doses of the vaccines, different schedules of vaccine administration, and different adjuvants to enhance the immune response. Recently, a phase II trial of two gp120 subunit products was begun in which high-risk as well as low-risk individuals were enrolled (L. Corey, personal communication). This trial, among other things, will compare the immunologic responses between high- and low-risk individuals. Injecting drug users, for example, might not respond as well as healthy, low-risk individuals. Although larger in size than a phase I trial, a phase II study does not have adequate sample size to evaluate a candidate vaccine's ability to prevent infection or disease. Such an evaluation must await a phase III efficacy trial.

The side-effects of candidate HIV-1 vaccines have been typical of vaccines in general. These include local pain and tenderness at the site of injection, low-grade fever and minor systemic symptoms. With some of the newer adjuvants, some volunteers have experienced more severe malaise and fever, occasionally leading to one or two days of missed work or school. Overall, the vaccines appear to be well tolerated. No haematologic, renal, hepatic, or neurologic toxicity has been reported. Also, CD4\(^+\) T lymphocyte cell counts have been monitored in vaccines receiving gp120 or gp160 vaccines because of the concern that these subunits might cause immune suppression or loss of CD4 cells through binding to CD4 on lymphocytes. CD4 counts, however, have remained stable in these vaccines.

Humoral immune responses have been detected in most vaccine recipients. Antibodies which bind to HIV-1 have been induced, frequently after the first or second immunization. Type-specific neutralizing antibodies, on the other hand, are detectable after the third or fourth immunization. Antibody titres also appear to be dose-dependent with larger doses inducing higher titres in general. Antibody titres, however, appear to be an order of magnitude lower on average than titres observed in natural infection. Furthermore, antibodies persist for months rather than years. Whether the magnitude and duration of antibody response is adequate to support the entry of any of these candidate vaccines into efficacy testing is presently being debated.

Cellular responses to immunization include \textit{in vitro} proliferation of lymphocytes, a measure of T-cell memory. More difficult to induce are cytotoxic T lymphocytes (CTLs) which can kill HIV-infected cells. No CD8\(^+\) CTLs have been detected in vaccinees immunized with subunit vaccines to date. However, priming vaccinees with recombinant vaccinia expressing gp160 and then boosting with purified gp160 protein has induced CD8\(^+\) CTLs in some volunteers. No data are presently available on the CTL responses of peptide vaccines which contain CTL inducing epitopes.

In the U.S., the two subunit products currently in phase II clinical testing have been the best candidates so far to consistently induce neutralizing antibodies against laboratory isolates of HIV-1. French studies of gp160 purified protein in combination with live vector or V3 peptides have demonstrated comparable levels of neutralizing antibodies. However, the ability of vaccine-induced antibodies to neutralize the many different field isolates to which individuals may be exposed remains to be determined. Initial assays to determine the breadth of the immune response have proven controversial.\textsuperscript{25}
3. GENERAL FEATURES OF VACCINE EFFICACY TRIALS

3.1. Outcomes of immunization

Immunization of high-risk uninfected individuals with a candidate HIV vaccine may result in a number of outcomes. In the best case, the vaccine would induce 'sterilizing immunity' that blocks the initial infection of susceptible cells by HIV. Immunization, however, might not induce sterilizing immunity, but it might prevent the establishment of a chronic infection. There might be a brief period of viral replication indicating an acute infection, but then the immune system clears the virus completely from the host. In either situation, infection is not established and the immunized individual does not develop AIDS.

Even if immunization does not prevent infection, it might prevent disease. For example, proviral DNA may be present in some cells, but it remains latent or when expressed it triggers an immune response which kills the infected cell. Another outcome of immunization may be that neither infection nor disease is prevented, but the onset of AIDS is somehow delayed. In this situation, it will be important to determine if immunization also reduces HIV transmission. If immunized individuals who become infected are infectious for longer periods and continue to expose themselves to others, incidence of HIV infection could increase. In this case, the vaccine would be a medical success, but a public health failure.

Conceivably, there could be some negative consequences of immunization. Immunization might induce enhancing antibodies which instead of preventing infection enable the virus to infect susceptible cells.25 As a result, immunization might facilitate infection or cause a more rapid course of disease. This phenomenon has been noted with earlier measles vaccines and has been seen in infants with residual maternal antibodies to dengue virus.27 There is also concern that immunization might cause autoimmune disease as certain epitopes of HIV proteins may mimic host proteins. Careful monitoring of trial participants, therefore, will be required. Furthermore, the possibility of a negative outcome of immunization necessitates the use of a two-sided statistical test of vaccine efficacy.

Traditionally, most vaccines have prevented disease rather than infection. Poliovirus, for example, can multiply actively in the intestinal tract of a person immunized with a killed poliomyelitis vaccine, but central nervous disease is prevented by the immune barriers induced by the vaccine. With HIV, the goal of preventing disease will be difficult to study since the median incubation period for AIDS is estimated to be on the order of 10 years and no reliable short-term clinical endpoints have yet been identified. Furthermore, the use of biological markers to document disease progression is problematic. Markers, such as CD4+ cell counts, are not only highly variable, but they may not correlate strongly with clinical outcome in immunized individuals who do become infected. CD4 count might be maintained for some time after infection in vaccine recipients, but this may not imply clinical benefit to the vaccinee. Also, immunization might have clinical benefit without significantly affecting CD4 count. Thus, at least in the U.S., trials are being designed with prevention of infection as the primary objective. In developing countries, where HIV incidence rates may be higher and the median incubation period for disease may be shorter, it may be feasible to study a candidate vaccine’s ability to prevent or delay disease.

3.2. Detection of HIV infection

One challenge at the moment is to develop assays that distinguish reliably between infected individuals and those uninfected persons with an immune response to the vaccine. Ordinarily, infection is detected by the presence of antibody to HIV.28 However, an immunized individual is
likely to produce antibodies in response to the vaccine and consequently to test positive for HIV by standard antibody assays such as the enzyme-linked immunosorbent assay (ELISA), whether or not he is actually infected. If the vaccine is based on a limited array of antigens, such as envelope proteins, a follow-up Western blot with antibody responses to other antigens should identify an infected person, but a more complex vaccine, such as a whole inactivated virus, would produce a Western blot pattern indistinguishable from actual infection. To detect infection in this situation, one would need to test for the presence of virus by such methods as an antigen capture assay, viral culture, or the polymerase chain reaction (PCR), a technique for viral RNA or complementary DNA amplification. Live attenuated vaccines would present further difficulties with the need to distinguish between the attenuated and wild-type viruses. There are also cost considerations in the selection of a method for HIV detection. PCR and viral culture are considerably more expensive than an antibody or antigen capture assay.

Reliable methods for HIV detection are also needed to screen individuals for eligibility in vaccine trials. Trial participants will be screened at entry for HIV infection, and only those who test negative will be eligible for study, but because infected individuals can take more than 6 months to develop antibodies, an HIV screen that relies on the presence of antibodies will miss some early infections. Thus, if an ELISA is used to screen participants at entry, some infected individuals are likely to be enrolled. It will be important in this situation to store blood samples on all participants at baseline. Those individuals who are found to be infected, say within 6 months of enrollment, would have their stored blood sample screened with a more definitive test such as PCR. Those individuals found to have been infected at entry could be excluded from the data analysis provided infected vaccinees and controls are equally likely to be detected in this manner. However, if immunization delays the development of antibodies to natural infection, infected vaccinees may be less likely to be detected than infected controls. Excluding these individuals could then introduce bias into the data analysis.

Because cumulative infection rates are expected to be low in these trials (less than 10 per cent), a high false-positive rate associated with the method of HIV detection could clearly obscure the trial results. For unvaccinated individuals in the U.S., estimates of the specificity for the standard ELISA with confirmatory Western blot exceed 0.999. However, the specificity of this approach in immunized individuals is not likely to be as good. For example, if the vaccine is based on envelope proteins and the vaccinee has a Western blot with positive bands for both envelope and core antibodies, he would be considered infected by the standard definition, but some individuals are expected to have positive bands for antibodies to core proteins such as p24 without being infected. In a study of 318 low risk individuals in West Virginia, 29 individuals were found to have positive bands for p24 antibody. In the Multicenter AIDS Cohort Study, 8.0 per cent of 1595 men who remained seronegative for at least 4 years had Western blots at entry which showed reactions to core proteins. Moreover, in Africa and other developing countries, the specificity may not be adequate, since some individuals living there may be infected with other viruses that react to the Western blot. Even if more definitive tests, such as PCR, are used to detect infection, the problem of misclassification may not be completely solved. HIV mutates readily, and PCR may not detect new mutants.

For trials involving subunit vaccines, participants could be screened at each study visit with an assay that detects antibodies to non-vaccine HIV antigens. Those individuals who are found to be 'seropositive' would then have their infection confirmed using a virus detection method such as PCR. This approach should have sufficient specificity. Also, performing PCR on all participants at the end of the study might detect infections for which there is no antibody response to natural infection. If PCR is used, one way to reduce the number of tests is to perform PCR on pooled blood samples. Blood from several individuals might be pooled and PCR applied to the
combined sample. If the test result is negative, all individuals would be considered uninfected. If the result is positive (or indeterminate), individual blood samples would be tested. This approach would add another layer of complexity to the misclassification problem, but it could result in substantial savings in the number of tests required. Alternatively, PCR might be applied to blood samples of the highest risk participants. If few additional infections are found, it may not be cost effective to test the remaining participants. See Hoff et al. for further discussion of monitoring HIV infection in trials.

3.3. Randomized controlled trials

It is expected that initial vaccine efficacy trials will be randomized double-blind with a placebo control. Randomization is essential to maximize the likelihood that the vaccine and control groups are comparable with respect to factors which influence HIV exposure and transmission, thus permitting any observed difference in HIV infection rates to be attributed to the effects of the vaccine. Blinding is important to ensure that risk behaviour, and therefore exposure to HIV, is comparable between vaccinees and controls. If unblinded, vaccinees might increase risk behaviour in the belief that they are protected, while control subjects remain or become cautious. This could lead to a paradox of vaccinees experiencing more infections than controls. Blinding also guards against differential rates of loss to follow-up between vaccinees and controls. It further provides an unbiased assessment of study outcomes. Last, the need for concurrent controls is twofold. First, historical controls may differ from trial participants with respect to factors which influence HIV exposure and transmission. For example, treatment of other sexually transmitted diseases may have improved, thereby decreasing the risk of HIV acquisition. Second, if all trial participants are immunized, they may increase their risk behaviour, making it difficult to detect a vaccine effect even if suitable historical controls can be found.

Blinding, however, may be difficult to maintain in some study populations since study participants who receive vaccine are likely to test positive for HIV by standard antibody assays. Some participants may thus unblind themselves by seeking HIV testing outside of the trial such as at anonymous testing sites. Others may inadvertently unblind themselves when they denote blood or plasma, but the extent to which participants are likely to unblind themselves is unknown. In the early trials of zidovudine to treat HIV-infected patients, there was a similar concern that patients might unblind themselves as to their treatment status. The degree of unblinding appeared to have been relatively small in these trials and may be small in vaccine trials as well. Nevertheless, it will be incumbent on those conducting trials to stress to participants that the vaccine's efficacy is unknown and that participants should continue to use safer sex and other risk reduction practices. It also will be essential to collect data on risk behaviour, including whether participants unblind themselves, to assess whether vaccinees and controls behave differently.

3.4. Sample size considerations

The number of trial participants required to assess the efficacy of a candidate vaccine will depend on many factors, the first of which is the choice for the primary endpoint. At least for trials conducted in the United States, HIV infection is expected to be the primary endpoint. In developing countries, the development of AIDS might be considered as the primary endpoint, as mentioned previously.

One of the most important factors affecting sample size is the incidence rate of HIV infection in the study population. The smaller the rate is, the larger the required sample size will be. Thus, it is most efficient to enroll individuals with the highest risk of HIV infection. There may be difficulties
SOME STATISTICAL ISSUES IN HIV VACCINE TRIALS

with recruiting or retaining these individuals, however. For example, active injecting drug users (IDUs) have a greater infection rate than users in drug treatment, but they may be more difficult to recruit or retain. Also, some IDUs may become incarcerated during the trial, making it more difficult to follow these individuals.

By ethical necessity, those conducting trials must provide counselling to trial participants on ways to avoid exposure to HIV. However, counselling and other non-vaccine interventions, such as providing clean needles to IDUs, could have a significant impact on the expected incidence of HIV infection. To the extent that these prevention efforts are successful, the overall risk of infection within the trial will drop, and the power to detect a benefit from immunization will fall correspondingly. Therefore, the required sample size will increase according to the expected benefit of counselling or other interventions on the HIV infection rate. Thus, to estimate sample size accurately, seroconversion studies, conducted in preparation for efficacy trials, should determine incidence rates under planned trial conditions including counselling and other interventions.

Another factor affecting sample size is the minimum level of efficacy considered worthwhile for a vaccine. The smaller the level of efficacy to be tested against the null hypothesis that the vaccine has no effect, the larger the sample size required, but while it may take fewer participants to test that a vaccine is 90 per cent rather than 50 per cent with adequate power, there is less precision in the estimate of efficacy. For example, 166 total subjects enrolled in a 3-year study were sufficient to have 90 per cent power to detect 90 per cent efficacy in a study population with a 7 per cent annual rate of infection. Even if one observes 90 per cent fewer infections among vaccinees, the 95 per cent confidence interval for the true vaccine efficacy could have a lower bound of 24 per cent. This lower bound may be an unacceptable level of efficacy to make a national commitment to distribute a vaccine widely. Thus, sample size calculations also need to consider the precision of the estimate of vaccine efficacy.

The length of follow-up of trial participants also affects sample size. With longer follow-up, fewer subjects are required, since longer follow-up allows time for additional endpoints. Thus, one way to reduce sample size is to increase the length of follow-up on each participant. However, the trial would take longer to reach a conclusion about vaccine efficacy. Moreover, loss to follow-up is likely to increase with longer follow-up which could affect the generalizability of the trial results. Long-term follow-up will be required, none the less, to assess the long-term effects of immunization. Identifying data such as a participant's social security number might be collected to utilize existing surveillance systems such as the U.S. National Death Index. Obviously, the confidentiality of such data would need to be assured.

In many clinical trial settings, one can assume that the effects of treatment are immediate. With HIV vaccines, this is not likely to be the case. Immunization is expected to involve a series of injections over a 6-month period. Protection may build up gradually over the course of these injections, with maximum protection achieved shortly after the last inoculation. In a more traditional vaccine trial, one might be able to begin the formal observation period after the last injection so that those in the vaccine group would have achieved the maximum possible protection. For example, in the 1954 Salk polio vaccine trial, efforts were made (unsuccessfully) to complete immunization of all children before the start of the polio season. In the AIDS setting, however, the study populations are likely to be at risk, before, during, and after the immunization period. It would be inappropriate to ignore infections that occur during that period, because the immunized subjects should have obtained at least partial benefit. (Similarly, failure to complete the full series of injections must not cause exclusion of a subject from the main analysis.) Thus, the method of sample size calculation must take account of the possibility that benefit would accumulate gradually.
Lakatos\textsuperscript{34, 35} derived formulae to calculate sample size based on the logrank statistic for clinical trials in which a lag in treatment effect is expected. Using his methods and assuming the full effect of immunization is achieved gradually over the course of immunization, one sees an appreciable increase in sample size, especially if the immunization period is long relative to the follow-up period. This is to be expected since the lag in vaccine effect dampens the difference in the cumulative infection rates between vaccinees and controls. Dixon et al.\textsuperscript{40} illustrate the increase in sample size when one accounts for a linear lag of 6 months. Here, sample sizes are based on an unweighted logrank statistic to test for vaccine efficacy.

A weighted logrank statistic, however, which places maximum weight where the full effect of immunization is expected to occur and less weight where the effect is not fully achieved should be more efficient. Zucker and Lakatos,\textsuperscript{41} motivated by such trials as the Lipid Research Clinics Coronary Primary Prevention Trial, in which a treatment lag was observed, proposed a weighted logrank statistic with weight at time \( t \) equal to the proportion of the treatment effect achieved by time \( t \). If vaccine efficacy is truly achieved gradually over the 6 month immunization period, then this weighted logrank statistic should enjoy certain optimality properties.

The approach of Zucker and Lakatos assumes that once the full effect of immunization is attained, that benefit persists, at least for the duration of the trial, but protection may wane over time as evidenced, for example, by declining antibody titres in vaccine recipients. In this situation, one might define weights \textit{a priori} based on antibody titres or other measures of immune system response observed in phase II immunogenicity studies. Less weight would be assigned where antibody levels are expected to be in decline. However, if antibody levels do not correlate with protection, basing the weights on this measure could have dire consequences on the power of the test. The demonstration of declining antibody levels in phase II trials may also indicate the need for a booster injection, say at 12 months.

Another approach is to use an adaptive weighted logrank statistic as proposed by Self.\textsuperscript{42} The general shape of the weight function is specified in advance of the trial, but the actual weights are determined by the observed rates of infection among vaccinees and controls. Such an approach has been proposed for cancer prevention trials, where, for example, participants' adherence to dietary intervention may decrease with time on study.

An open question, however, is how does the interim monitoring of trial results affect this approach. Each trial will have the oversight of an independent Data and Safety Monitoring Board which will monitor the progress of the trial on a regular basis and make recommendations concerning its continuation, termination, or modification. Early termination may occur because of unacceptable adverse effects of immunization, early demonstration of vaccine efficacy, lack of power to detect efficacy levels of interest, or external events such as the availability of scientific results from other similar vaccine trials.

Table II depicts sample size requirements for a two-arm trial under varying trial conditions. A linear lag of 6 months in attainment of the full effect of immunization is assumed. The weighted logrank statistic uses weights which increase from 0.0 to 1.0 linearly over the 6 month immunization period and equal 1.0 thereafter. Sample size is calculated to provide 90 per cent power to detect an ultimate 50 per cent reduction in the hazard of infection. Subject accrual is assumed to occur uniformly over the first year of the trial. Subjects are further assumed to become lost to follow-up at a rate of 10 per cent per year. Also, 5 per cent of vaccinees are assumed not to complete the full course of immunization and to experience little or no benefit from the injections received.

From Table II, one sees that there is a sample size savings of 12 to 18 per cent by use of the weighted logrank statistic. However, the unweighted logrank statistic is preferred for the primary data analysis. The vaccine might induce enhancing antibodies which increase the risk of infection.
among vaccinees. Trial participants may unblind themselves and change risk behaviour. In either case, there could be more infections among vaccinees than controls. By down-weighting infections which occur early, one could potentially make a harmful vaccine seem acceptable. None the less, a weighted logrank statistic may be useful as part of a secondary analysis.

3.5. Potential trial cohorts

Sample size will be determined in large part by the cohorts selected for study. For gay and bisexual men followed in the Multicenter AIDS Cohort Study, the HIV seroincidence rate is about 1 per cent per year. For younger gay men, the rate is expected to be somewhat higher. Preliminary estimates from the San Francisco Young Men's Health Study indicate a seroincidence rate of 2 to 3 per cent per year among homosexual men between the ages of 18 and 29 years. In a cohort of injecting drug users the rate was between 2 and 5 per cent per year.

In developing countries, incidence rates are expected to be much higher. In a rural Uganda cohort, the incidence rate of heterosexually transmitted HIV varied from 2 to 7 per cent per year depending on the age of the individual. Among conscripts of the Royal Thai Army, an incidence rate of 3-9 infections per 100 person years of follow-up has been observed. In Haiti, a study of discordant couples showed an incidence rate of 7-2 infections per 100 person years of follow-up, despite counselling to avoid high risk behaviour. Follow-up has also been exceptional with 95 per cent of the cohort still under observation at 4 years. Moreover, medical treatment of the seropositive partner could provide a means for maintaining contact with the seronegative partner enrolled in a vaccine trial.

Among commercial sex workers in Mombasa, Kenya, an incidence rate of 15-3 infections per 100 person years of follow-up was seen. However, compliance with follow-up in this cohort was only 59 per cent at 3 months. Furthermore, inducements to avoid sexual contact during the immunization period must be found to permit these women to experience the full benefit of immunization.

Another population to consider for efficacy trials is infants born to HIV-infected mothers. Fourteen to 39 per cent of these infants are expected to become HIV infected. Furthermore, the incubation period for AIDS is much shorter in children than adults, with the median time to AIDS estimated to be on the order of 4 years. Thus, trials in this population might determine the ability of a candidate vaccine to prevent infection or disease in a relatively short period of time. The demonstration of efficacy in paediatric populations would increase enthusiasm for similar trials in adult populations.
While maternal antibodies to measles virus appear to render live-attenuated measles vaccines ineffective in infants less than 9 months of age, maternal or other passively acquired antibodies do not appear to interfere with vaccines based on protein antigens. Infants, for example, who are born to women who are chronic carriers of hepatitis B virus are effectively immunized at birth by the joint administration of hepatitis B vaccine containing surface antigen and antibodies derived from the sera of infected persons. Furthermore, this active plus passive immunization is more effective than either strategy alone. Thus, a similar approach with HIV vaccines may prove to be effective in paediatric populations as well.

It is not known, however, what fraction of infected infants acquire their infection in utero and what fraction become infected during labour or delivery. If the majority of infants are infected in utero and are not likely to benefit from immunization, then only modest reductions in the overall HIV transmission rate may be possible. As methods for determining the time of infection improve, it will be informative to assess vaccine efficacy in infants exposed at birth separately from infants infected in utero. Trials must also take into account that infants may become infected after birth through breast feeding, particularly in developing countries where safe alternatives to breast milk are not available.

Effective recruitment and retention strategies will be critical to enroll and retain the necessary number of trial participants. Those conducting trials may need to allay fears about government involvement, especially with the continuing legacy of the Tuskegee study of untreated syphilis in African-American males. Trials must also ensure that vulnerable populations and developing countries do not serve as 'guinea pigs'. Therefore, negotiations with target communities and host countries must address in advance of trials plans for the provision of vaccines found to be safe and effective. Furthermore, confidentiality of trial participants will need to be ensured, since a person’s very participation identifies him as a high-risk individual which could jeopardize his employment or his ability to obtain insurance. Access to health insurance, consequently, may need to be secured by trial sponsors, and because immunized individuals may test HIV positive by standard assays, free access to highly specific testing may need to be provided to these individuals after the trial is completed. Other issues in trial implementation are discussed in Rida et al.34

3.6. Estimation of vaccine efficacy

Vaccine efficacy is estimated generally by \( VE = 1 - RR \) where \( RR \) is some measure of the relative risk of infection between vaccinees and controls. If \( (C_v, Y_v, N_v) \) represents the number of cases (that is, infections), person years at risk, and number of individuals enrolled in the vaccine arm and \( (C_u, Y_u, N_u) \) represents those numbers for the ‘unvaccinated’ control arm, then vaccine efficacy can be estimated by either

\[
VE_1 = 1 - \frac{C_v}{C_u/N_u}
\]

or

\[
VE_2 = 1 - \frac{C_v}{Y_u}
\]

The first estimator uses cumulative attack rates to measure the relative risk of infection. The second estimator uses estimates of the instantaneous hazard of infection. These estimators, however, behave differently under varying assumptions about trial length and the mode of action of the vaccine. For a vaccine which reduces the per contact probability of infection by a fraction \( p \) in a ll vaccinees, \( p \) is estimated by \( VE_2 \) for any length of study. \( VE_1 \), on the other hand, will tend
to underestimate \( p \) and will decrease toward zero as trial length is increased. For a vaccine which provides complete protection to a fraction \( p \) of vaccinees while providing no direct benefit to the remaining vaccinees, \( p \) is estimated by \( VE_1 \) for any trial length. \( VE_2 \), in this situation, will tend to overestimate \( p \) and will increase toward unity as the length of follow-up is increased. Thus, the appropriate selection of an estimator for \( p \), an important measure of a vaccine's direct effect on a vaccinee's susceptibility, will depend on what is known about a vaccine's mode of action.

These estimators, furthermore, assume that the ratio of the instantaneous hazards of infection between vaccinees and controls is constant over time, but because of the likely lag in vaccine effect and the possible waning of protection over time, the proportional hazards assumption may not be reasonable. If one is willing to assume a linear lag in vaccine effect of 6 months, say, then one can derive estimates for the maximum effect of immunization, but one may want to estimate other components of the vaccinee hazard function such as the time to maximum protection. Maximum protection may or may not be achieved in 6 months as assumed in sample size calculations.

Independent outcomes are also assumed. However, in contrast to non-communicable diseases such as cardiovascular disease and most cancers, independent outcomes cannot be assumed with communicable diseases such as AIDS. The infection status of one individual depends on the infection status of others with whom he has contact. Thus, a vaccine can have an indirect effect by reducing the number of infections individuals to which a susceptible person is exposed. This effect is referred to as herd immunity. Although prevalence of HIV may change over time in the study population due to the indirect effects of immunization, randomization at the individual level should assure comparable exposure to HIV between vaccinees and controls. In this case, \( VE_1 \) and \( VE_2 \) still measure the direct effect of a vaccine, but if randomization is by some other sample unit, such as clinic or village, these estimators may no longer estimate a vaccine's direct effect alone. Halloran et al.\(^5\) discuss the assessment of a vaccine's direct and indirect effects.

These estimators also assume that the risk of infection is identical for each unvaccinated individual. Individual risk, however, is likely to vary since individuals may have different numbers of sexual or needle-sharing partners. Other behaviour affecting exposure to HIV, such as the use of condoms, may vary as well, but while it may be desirable to collect behavioural data longitudinally to assess whether vaccinees and controls have comparable exposure to HIV, detailed data may not be required to measure vaccine efficacy reliably. Incorporating behaviour data into the estimate of efficacy may increase the precision of the estimate. However, such data may not be necessary to obtain an unbiased estimate, provided participants do not unblind themselves and differentially change risk behaviour. Furthermore, if behavioural data are unreliable, they may introduce considerable bias to the estimate of efficacy. For example, participants may have difficulty recalling risk behaviour or they may not disclose behaviour which they have been counselled to avoid.

Trials conducted in populations likely to unblind themselves and differentially change risk behaviour will not provide direct measures of a vaccine's biological efficacy. Attempts to estimate biological efficacy indirectly by statistical adjustment for change in risk behaviour at the individual level may not succeed. Even if data on risk behaviour are reliable, a positive result from an adjusted analysis could be due to residual confounding. There could be differences in risk behaviour not measured by a risk assessment questionnaire that account for this result. In this setting, one might have to be satisfied with estimating a vaccine's 'effectiveness' rather than its biological efficacy. None the less, efforts to estimate biological efficacy by statistical adjustment may be useful as part of a secondary analysis of the data. Halloran et al.\(^9\) illustrate the impact of behavioural change on estimates of vaccine efficacy. Bias introduced by changes in risk behaviour is most evident for vaccines with modest levels of biological efficacy.
Finally, a vaccine might prevent or delay disease rather than prevent infection. In this case, $C_v$ and $C_w$ represent the number of infected individuals who develop disease in the vaccine and control group, respectively. A vaccine could also reduce the contagiousness of vaccinees who become infected. This could impact the epidemic significantly. Koopman and colleagues\textsuperscript{58} propose an ‘exposure risk’ approach to assess the effect of a vaccine on contagiousness. However, this approach which entails collecting data on a sample of individuals at risk rather than the partners of infected individuals would be very difficult to implement in efficacy trials. Such an approach might be carried out during a phase IV post-marketing study. Further thought on defining and estimating vaccine efficacy in phase III trials is needed.

4. SECONDARY OBJECTIVES

4.1. Correlates of protective immunity

In addition to assessing vaccine efficacy, trials may attempt to determine what immune responses induced by the vaccine correlate with protection in vaccine recipients. There are a number of reasons for determining the correlates of protective immunity. First, if the vaccine is highly effective then correlates of immunity could serve as surrogate endpoints in future trials. Future trials might involve the same vaccine, but a different study population. Also, new but similar acting vaccines might be developed which have fewer side effects or provide more durable protection. Second, even if the vaccine fails or is only moderately effective, immunologic assessment of vaccine recipients could provide valuable clues to guide future vaccine development.

For many infectious diseases, one can search for the correlates of protective immunity by examining the immune responses of individuals who recover from natural infection. Individuals who recover from hepatitis B, for example, develop antibodies to the surface antigen of the virus. No one, however, has been known to recover from HIV infection. Another way to identify the correlates of immunity is to study the immune responses of individuals who have had known repeated exposure to HIV, but who have not become infected. Nairobi prostitutes, for example, who have remained seronegative for several years despite continued exposure to HIV, are being carefully studied.\textsuperscript{59} A third approach uses animal models for HIV infection and disease. Chimpanzees, as mentioned previously, can be infected with HIV, but they do not develop disease. Their immune response to infection is being carefully studied. Rhesus macaques, immunized with live-attenuated SIV and then challenged with pathogenic virus, may also provide valuable information about what immune responses are necessary for protection.\textsuperscript{10}

In clinical trials, one might assess the correlates of protective immunity by comparing the immune responses of vaccinees who become infected (cases) with the responses of vaccinees who do not (controls). To judge the importance of a humoral response to immunization, for example, one might calculate the mean antibody titre to a specific vaccine antigen one month after the last inoculation for the cases and compare that to the mean for the controls. (Obviously, one would need to control for factors which affect risk of infection such as the number of sexual or needle-sharing partners and the length of follow-up since infected vaccinees may have had greater exposure to HIV than their uninfected counterparts.) This approach was taken by the Ad Hoc Group for the Study of Pertussis Vaccines\textsuperscript{60} in a Swedish study of two acellular pertussis vaccines to prevent whooping cough in children. As in that trial, there may be no difference in mean antibody titres between cases and controls. In this situation, it will be important to determine if antibodies decay more quickly among cases than controls. Cases may have a similar initial rise in antibody levels as controls, but their antibody levels decline faster. Granström and Granström\textsuperscript{61} postulate that this is what might have occurred in the Swedish pertussis trials.
If rates of decay are similar between cases and controls, however, this may indicate that antibodies are not protective and that some other immune response such as the development of cytotoxic T lymphocytes (CTLS) is important for protection. HIV is transmitted by both cell-free and cell-associated virus. While neutralizing antibodies may be sufficient to prevent infection by cell-free virus, their ability to protect against cell-associated virus is questionable. CTLS, on the other hand, recognize and kill HIV-infected cells that display fragments of viral proteins on their surface. In this regard, CTLS may play an important role in preventing or clearing infection.

Given the interplay between the humoral and cellular arms of the immune system, the assessment of the correlates of protective immunity may well be a multivariate problem. Furthermore, Granström and Granström stress that one must realize the technical difficulties in designing studies of correlates. Consequently, correlate data must be interpreted with caution and with the understanding that failure to show a correlation does not necessarily mean that a correlate does not exist.

### 4.2. HIV variation

HIV exhibits considerable variation in its genetic sequence which could have considerable impact on vaccine development. There are at least five genetic ‘subtypes’ of HIV-1 with each subtype containing many different strains. Vaccines furthest along in development are based on a single virus strain rather than a ‘cocktail’ of several different strains. Furthermore, the strains are of subtype B, the predominant subtype of North America and Europe. It is not known, however, if a vaccine based on one strain will protect against exposure to other strains, even of the same subtype. It may be important then to sequence virus from participants who become infected. The pattern of virus strains in infected controls is likely to reflect the pattern in the study population. On the other hand, the pattern in the infected vaccines may show fewer infections from strains that match or are very similar to the strain on which the vaccine is based. This finding would suggest a strain or subtype-specific protection and may indicate the need for a multivalent approach to immunization.

For populations in which two or more virus subtypes are found, immunization might reduce the number of infections of the subtype(s) on which the vaccine is based, but appear to cause more infections of a subtype on which the vaccine is not. This ‘pseudo-enhancement’ may occur because an individual who is immunized against one subtype may have increased exposure to virus of another subtype due to a lack of natural competition between viruses. This phenomenon was observed to some degree in hepatitis B vaccine trials. Individuals immunized against hepatitis B appeared to experience more cases of non-A, non-B hepatitis than similar controls. This phenomenon could occur in a trial conducted in a country such as Thailand where two subtypes of HIV-1 predominate. Subtype A is found primarily in sexually infected individuals, while subtype B is found mostly in injecting drug users. Even if a trial of a subtype B vaccine were tested only in IDUs, one would still need to be concerned with vaccinees being infected with viruses of subtypes A. Careful analysis of virus sequence may be required.

### 4.3. Other virologic and immunologic assessments

To gain insight on a candidate vaccine’s ability to prevent or delay disease, investigators must follow trial participants who become infected for an extended period after infection is discovered. Virologic and immunologic measures such as viral load and CD4+ cell count, measured in the months following infection, may provide short-term clues to a candidate vaccine’s clinical benefit.
Initial follow-up may also reveal the occurrence of transient infections in vaccinees, indicating that immunization might permit some individuals to clear their infection. However, it is not known how quickly infection might be cleared, if it is cleared at all. Furthermore, virus may be cleared from the peripheral circulation, but remain in lymph nodes and other tissues, making it difficult to judge if an infection has truly been eliminated.

Viral load, as measured by quantitative PCR, is of interest for a number of reasons. First, viral load may be lower in infected vaccinees, suggesting a possible clinical benefit of immunization. Alternatively, a higher viral load in vaccine recipients may indicate disease enhancement due to immunization, a serious safety concern. In either case, the use of viral load as a surrogate endpoint in future trials must be validated by long-term follow-up to assess its correlation with clinical outcome. Finally, viral load may be related to the infectiousness of infected individuals with lower loads associated with reduced transmission of HIV. However, one would need to measure infectious virus in seminal fluids and vaginal secretions, since the amount of virus in peripheral blood may not correlate with the amount in these fluids and secretions. Furthermore, one would need to correlate viral load with transmissibility of the virus. Also, characterization of HIV cell-tropism may reveal differences between infected vaccinees and controls which affect HIV transmission.

CD4 count is of interest for many of the same reasons as viral load. Natural history studies such as the San Francisco Men’s Health Study have noted an appreciable drop in CD4 count shortly after seroconversion. If the CD4 drop is observed to be less pronounced in infected vaccinees, this might indicate a possible clinical benefit of immunization, but because CD4 count is costly to measure and must be measured in real time with current technology, it will not be feasible to obtain baseline CD4 count against which to measure a drop. To do so would require knowing who will become infected or measuring CD4 on all participants at baseline. However, CD4 count is known to further decline in the first 12 months after seroconversion. Comparing CD4 change over this period between vaccinees and controls may still be meaningful. None the less, the use of CD4 count as a surrogate endpoint must be validated by long-term follow-up to determine what fraction if any of a vaccine’s benefit is mediated through CD4. Monitoring CD4 count after infection is also important for safety considerations. If vaccine recipients exhibit a more rapid decline in CD4 count, this may indicate disease enhancement. It also must be pointed out that at least in the U.S., the potential use of licensed or experimental antiretroviral drugs or therapeutic vaccines to treat infected trial participants may further complicate the interpretation of CD4 count and other biological markers.

It is critical to validate the use of biological markers such as viral load and CD4 count as surrogate endpoints in future trials for which the use of clinical endpoints would require lengthy follow-up or very large numbers of participants, perhaps more than can be realistically enrolled. If unvalidated, the marker could lead to erroneous conclusions about vaccine efficacy. Furthermore, a marker which is validated for a particular candidate vaccine must be validated for other candidates, especially if the new candidate has a different mechanism of action.

There has been considerable debate over what constitutes a good surrogate marker. It is not sufficient for natural history studies to demonstrate that the marker is a good predictor of clinical outcome. Immunization must also mediate its effect through the marker to a large degree, but because immunization may affect many underlying processes which influence clinical outcome, reliance on a biological marker to assess vaccine efficacy is problematic. For example, CD4 count is a strong predictor of disease progression. In a study of macaques immunized with a whole, inactivated vaccine, all immunized monkey became infected after challenge with pathogenic SIV and showed the same drop in CD4 count as the unvaccinated controls. However, the immunized monkeys showed evidence of disease amelioration and prolonged
Table III. Participant enrolment into trial which adds a second vaccine

<table>
<thead>
<tr>
<th>Enrollment period (months)</th>
<th>Control</th>
<th>Trial arm A</th>
<th>Trial arm B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–6</td>
<td>1250</td>
<td>1250</td>
<td>0</td>
</tr>
<tr>
<td>6–12</td>
<td>1250</td>
<td>1250</td>
<td>2500</td>
</tr>
<tr>
<td>12–18</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–6</td>
<td>1250</td>
<td>1250</td>
<td>0</td>
</tr>
<tr>
<td>6–12</td>
<td>1250</td>
<td>1250</td>
<td>1250</td>
</tr>
<tr>
<td>12–18</td>
<td>1250</td>
<td>1250</td>
<td></td>
</tr>
</tbody>
</table>

Survival. Thus, CD4 would lead to a false negative conclusion about clinical efficacy. Interestingly, CD8+ lymphocytes declined precipitously in the unvaccinated monkeys, but remained stable in the vaccinated.) Even if CD4 had remained stable in the immunized monkeys, one could not conclude a clinical benefit of immunization on CD4 alone. Immunization might have a deleterious effect on some other underlying process that more than offsets the benefit of stabilized CD4 count.

5. OTHER TRIAL DESIGNS

5.1. Adaptive designs

Given that many candidate vaccines are in development, new candidates could become available for efficacy testing while phase III trials of existing candidates are still enrolling participants. In this situation, one might consider adding new arms to the existing study rather than starting new trials. This approach might provide some economy in sample size over that required to conduct multiple separate trials. It may also afford a more direct (albeit limited) comparison of the candidate vaccines.

Suppose a two-arm trial of candidate A is designed to enroll 5000 total participants with equal allocation between the vaccine and control arms. After half the participants are enrolled, a second candidate B becomes available. Instead of initiating a second trial with another 5000 subjects, one might consider adding another 2500 participants to the first trial. The remaining participants might be randomized to the control, candidate A, and candidate B arms in a 1:1:2 ratio, respectively, see Table III (case 1). However, a comparison of the HIV infection rates between the control and candidate B arms may not be valid. Individuals assigned to candidate B will have been enrolled later on average than individuals randomized to the control group. Individuals entering the study later, after candidate B is offered, may be different from individuals who entered the study when only candidate A was available. This could introduce a selection bias. Also, the background rate of HIV infection may have changed over time, making the comparison questionable.

A more appropriate approach after candidate B is introduced is to continue randomization in a 1:1:1 ratio until 2500 participants have been assigned to candidate A, see Table III (case 2). Subsequent participants are randomized in a 1 to 1 ratio between the control and candidate B arms until 2500 individuals have been assigned to candidate B. Under this scheme, 3750
individuals are randomized to the control arm whereas 2500 individuals are randomized to each of the candidate arms. Total sample size is 8750 which is still less than the 10,000 required to conduct two separate trials. Individuals randomized to candidate A are compared with the first 2500 individuals randomized to the control group while individuals randomized to candidate B are compared with the last 2500 individuals randomized to the control group. Notice that individuals randomized to the control group in the middle of the enrollment period are used twice, so to speak.

There may be logistical problems, however, with adding arms to existing trials. The new candidates may also cause different reactions in vaccinees, making it difficult to fully blind the trial. Furthermore, if the new candidates contain different antigens, then the method of HIV detection may need to be different. For example, an assay to detect antibody to the transmembrane protein gp41 can be used to detect infection in individuals immunized with a gp120 subunit vaccine. It cannot be used in individuals immunized with a gp160 subunit vaccine since these vaccinees may develop antibody to the gp41 portion of gp160. Careful consideration, therefore, needs to be given to such issues when deciding to add arms or not.

5.2. Factorial designs

In a more comprehensive effort to prevent the spread of HIV, investigators may study the benefits of combining other intervention strategies such as the use of microbicides with the administration of vaccines. Nonoxynol-9, for example, is an effective viricide that may enhance the effect of a vaccine by lowering the viral inoculum to which an individual's immune system is challenged. With a smaller viral challenge, the immune system may be more capable of preventing infection or disease. A factorial design in this situation might be an efficient design for studying the individual and combined effects of vaccine and microbicide. Other situations in which factorial designs may be useful include the treatment of non-HIV sexually transmitted diseases in combination with vaccine, the active and passive immunization of infants born to HIV-infected women and the study of combination vaccines. A factorial design may be particularly efficient for two vaccines with different mechanisms of action. One vaccine might induce a strong antibody response while the other induces a strong cell-mediated immune response.

The classic factorial design is often advertised as 'two studies for the price of one'. This claim may be reasonable if the two treatments under study affect different endpoints such as in the Physicians' Health Study. In this study, physicians were randomized to receive aspirin (to prevent cardiovascular disease), beta-carotene (to prevent cancer), both, or neither. Since aspirin is not thought to prevent cancer and beta-carotene is not thought to prevent cardiovascular disease, sample size is determined by calculating the sample size for each treatment question as if two separate studies were being planned and then taking the larger of the two numbers.

However, if the two treatments (prevention strategies) affect the same endpoint, then one does not usually get 'two for one'. If both treatments are effective to some degree, then the expected number of endpoints may be substantially smaller than if only one or none were effective. The smaller number of endpoints will drive up the required sample size. If one wishes to have reasonable power to detect interaction effects, the sample size will be greater still. None the less, there are situations in which the sample size using a factorial design will be substantially less than the sample size required to conduct two separate studies.

In considering a factorial design, one has to weight the difficulties that may arise with such a design. There may be logistical complications with the administration of two or more prevention efforts. There may be problems with participant compliance. If the two treatments cannot be given on the same schedule, for example, participants may find it difficult to adhere to the study
protocol. Non-compliance could seriously decrease the statistical power of the study. There also may be monitoring issues. An unexpectedly high adverse event rate in an arm with combination treatment could result from the toxic effect of a single component or from a toxic interaction. These issues need careful consideration when deciding to use a factorial design.

6. CONCLUSIONS

Efficacy trials of prophylactic HIV vaccines will be among the most difficult clinical trials ever attempted. Not only will there be challenges with the recruitment and retention of high-risk uninfected individuals, there will be many statistical challenges to the design, conduct, analysis and interpretation of these trials.

Traditionally, most vaccines have prevented disease rather than infection, but because of the long incubation period of AIDS, prevention of infection may be the only practical goal to study in efficacy trials, especially in the U.S. and other developed countries. Suitable populations, none the less, may be found in which a vaccine's ability to prevent or delay disease can be studied. In these populations, biological markers such as viral load and CD4 cell count as well as potential correlates of immunity might be measured and their use as surrogate endpoints in future trials validated. Research, however, remains to be conducted on what makes a marker or constellation of markers a reasonable surrogate for disease outcome.

A candidate vaccine might also reduce the infectiousness of a vaccinee who does become infected. This effect of immunization could have a positive as well as a negative impact on the epidemic. If infectiousness is reduced without greatly extending the infectious period, the basic reproduction number of the epidemic might be reduced with a subsequent slowing of the epidemic. On the other hand, if infectiousness is only moderately reduced and the infectious period is extended, the basic reproduction number could increase with negative consequences on the epidemic. Assessing a vaccine's ability to reduce infectiousness, however, may be very difficult. While reduced viral load due to immunization may suggest reduced infectiousness, this must be verified somehow. Efficacy trials, by themselves, are unlikely to determine if a vaccine reduces infectiousness.

If a candidate vaccine were highly effective, then biases that might affect the analysis of vaccine efficacy might not be so great as to completely obscure the effects of immunization. However, given the urgency of the pandemic, less than ideal vaccines may enter into efficacy testing. If these vaccines are only moderately effective, then biases could significantly affect the ability to detect small, yet meaningful benefits of immunization. Great care in analysing and interpreting such data will be required.

ACKNOWLEDGEMENTS

The authors are gratefull to Bill Blackwelder, Dennis Dixon, Patricia Fast, Bob Fischer, Mary Foulkes, Rod Hoff, Nzeera Ketter, Jim McNamara, Nancy Miller and Amy Sheon for helpful discussions.

REFERENCES


