Statistical Data Analysis Plan

A Randomized, Worldwide, Placebo-Controlled, Double-Blind Study to Investigate the Safety, Immunogenicity, and Efficacy on the Incidence of HPV 16/18-Related CIN 2/3 or Worse of the Quadrivalent HPV (Types 6, 11, 16, 18) L1 Virus-Like Particle (VLP) Vaccine in 16- to 23-Year-Old Women—The FUTURE II Study (Females United to Unilaterally Reduce Endo/Ectocervical Disease)

(Study 015)

Prepared by

Lisa Lupinacci

Data Analysis Plan

A Randomized, Worldwide, Placebo-Controlled, Double-Blind Study to Investigate the Safety, Immunogenicity, and Efficacy on the Incidence of HPV 16/18-Related CIN 2/3 or Worse of the Quadrivalent HPV (Types 6, 11, 16, 18) L1 Virus-Like Particle (VLP) Vaccine in 16- to 23-Year-Old Women—The FUTURE II Study (Females United to Unilaterally Reduce Endo/Ectocervical Disease)

(Study 015)

TABLE OF CONTENTS

PAGE

Data Analysis Plan

(Study 015)

TABLE OF CONTENTS (CONT.)

PAGE

Data Analysis Plan

(Study 015)

TABLE OF CONTENTS (CONT.)

Data Analysis Plan

A Randomized, Worldwide, Placebo-Controlled, Double-Blind Study to Investigate the Safety, Immunogenicity, and Efficacy on the Incidence of HPV 16/18-Related CIN 2/3 or Worse of the Quadrivalent HPV (Types 6, 11, 16, 18) L1 Virus-Like Particle (VLP) Vaccine in 16- to 23-Year-Old Women—The FUTURE II Study (Females United to Unilaterally Reduce Endo/Ectocervical Disease)

(Study 015)

EXECUTIVE SUMMARY

Background

Over 50% of sexually active adults will become infected with human papillomavirus (HPV) during their lifetime. HPV infection can result in genital warts and anogenital dysplasia that may result in cancer. These diseases are associated with substantial morbidity and mortality. Every year, 471,000 cases of cervical cancer are diagnosed worldwide. The 5-year survival for this disease is 70%. The impact of cervical cancer is accentuated by the fact that this disease generally affects women in their 30's to 50's, a time of peak productivity. In the developed world, routine Papanicolaou (Pap) screening has reduced the incidence of cervical cancer by 75%. However, sporadic Pap screening in the developing world and among the disadvantaged in the United States has failed to reduce the incidence of cervical cancer worldwide.

Over 90 HPV types have been identified. HPV Types 16 and 18 cause ~70% of high-grade cervical dysplasia (cervical intraepithelial neoplasia 2/3 or CIN 2/3) cases and cervical and anal cancers, while HPV Types 6 and 11 cause >90% of genital warts. Therefore, it is expected that a prophylactic quadrivalent HPV vaccine that reduces infection with these 4 HPV types will reduce the incidence of CIN 2/3 and cervical cancer.

CIN 2/3 or Cervical Intraepithelial Neoplasia Grades 2 and 3 represents replacement of normal cervical epithelium with dysplastic cells. CIN 2 is defined as moderate to high-grade dysplasia encompassing between 1/3 and 2/3 of the thickness of the cervical epithelium. CIN 3 represents 2/3 to full thickness high-grade dysplasia or carcinoma in situ. Both CIN 2 and CIN 3 are considered high-grade cervical lesions and are indications for wide excision worldwide. In women with CIN 2/3, excision prevents the development of cervical cancer.

CIN 2/3 lesions are the immediate and obligate clinical and pathologic precursors to cervical cancer. An assessment of efficacy in reducing HPV 16- and 18-related CIN 2/3 lesions is the most clinically relevant and accurate way of determining whether or not administration of a prophylactic HPV vaccine containing types 16 and 18 will reduce the incidence of cervical cancer related to HPV Types 16 and 18. Thus the primary objective of Protocol 015 is to demonstrate that the quadrivalent HPV vaccine (Types 6, 11, 16, 18) reduces the incidence of HPV 16/18-related CIN 2/3.

Embedded in the HPV 16/18-related CIN 2/3 efficacy study is a substudy to demonstrate similarity in immune responses to the 4 HPV types in the vaccine among subjects who received vaccine from 3 consistency lots.

Study Objectives and Hypotheses

Safety: The primary safety objective of Protocol 015 is to demonstrate that a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 virus-like particle (VLP) vaccine is generally well tolerated. The corresponding primary safety hypothesis states that the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine is generally well tolerated in 16- to 23-year-old females.

Efficacy: The primary efficacy objective of Protocol 015 is to demonstrate that intramuscular administration of a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine reduces the incidence of the composite endpoint of HPV 16- and 18-related high-grade cervical abnormalities (CIN 2/3) or HPV 16- and 18-related invasive cervical carcinoma in subjects who are polymerase chain reaction (PCR) negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for the relevant HPV type. The corresponding primary hypothesis states that administration of a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine reduces the incidence of the composite endpoint of HPV 16- or HPV 18-related CIN 2/3 or invasive cervical carcinoma compared with placebo in subjects who are PCR negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for the relevant HPV type. *(The statistical criterion for success requires that the lower bound of the confidence interval for the vaccine efficacy exclude 0%.)*

Consistency Lot Substudy: The primary immunogenicity objective of the consistency lot substudy in Protocol 015 is to demonstrate that the Final Manufacturing Process (FMP) results in quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine that, when given in a 3-dose regimen, induces consistent serum anti-HPV 6, anti-HPV 11, anti-HPV 16, and anti-HPV 18 responses 4 weeks Postdose 3. The corresponding co-primary hypotheses are:

(1) Three separate lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine induce similar immune responses, as measured by the percentage of subjects who achieve serum anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL, at Week 4 Postdose 3. *(Each*

vaccine component will be analyzed separately. The statistical criterion for similarity requires that the upper bound of the confidence interval for the maximum absolute difference in proportions between any 2 of the 3 lots exclude 10 percentage points or more for each HPV type.)

(2) Three separate lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine induce similar immune responses, as measured by the serum geometric mean titers (GMTs) to HPV 6, 11, 16, and 18, at Week 4 Postdose 3. (*Each vaccine component will be analyzed separately. The statistical criterion for consistency requires that the upper bound of the confidence interval for the fold-difference in GMTs between any 2 lots exclude a fold-difference of 2 or greater for each HPV type.)*

Study Design

Protocol 015 is a prospective randomized, double-blind, placebo-controlled, parallel, multicenter study operating under in-house blinding procedures. Embedded within Protocol 015 is a consistency lot substudy. Approximately 11,500 subjects will be randomized in a 1:1 ratio to receive either 3 injections of quadrivalent HPV (Type 6, 11, 16, 18) L1 VLP vaccine or 3 injections of placebo. Within the quadrivalent HPV vaccine group, subjects will be further randomized to receive 3 different lots of the vaccine in a 1:1:1 ratio. The actual lots received will differ between subjects enrolled in the consistency lot substudy and the remaining subjects in Protocol 015. Each dose of quadrivalent vaccine will contain 20 μg HPV 6 L1 VLP, 40 μg HPV 11 L1 VLP, 40 μg HPV 16 L1 VLP, and 20 μg HPV 18 L1 VLP. Vaccine or placebo is administered at 0, 2, and 6 months. Approximately 1 month following the third vaccination visit (Month 7), 6 months following the third vaccination visit (Month 12) and every year thereafter until Month 48, subjects will return to the study center for collection of specimens, gynecologic examinations and Pap tests. Subjects with Pap test abnormalities at scheduled visits will be referred for colposcopy based on a mandatory triage strategy. Lesions observed during the colposcopy will be biopsied. Sections of each biopsy will be sent to members of the Pathology Panel for pathological diagnosis. Additional sections will be sent to Merck Research Laboratories (MRL) for PCR analysis. Should a subject be referred for definitive therapy based on a diagnosis of CIN 2/3 or worse, sections of the tissue excised during the definitive therapy procedure will be sent to the Pathology Panel for diagnosis. Additional sections will be sent to MRL for PCR analysis.

While most subjects in Protocol 015 will be followed only for the occurrence of serious adverse experiences (SAEs) following vaccination, the subjects who enroll in the United States (~1150 of the subjects [10%]) will participate in a nonserious adverse experience (NSAE) substudy. These subjects will undergo a full assessment of nonserious adverse experiences that occur within 14 days following each vaccination visit in addition to the follow-up for serious adverse experiences. Subjects enrolled in the United Kingdom will also undergo active follow-up for nonserious adverse experiences, as this is a regulatory requirement in the United Kingdom. Subjects at all other study sites may report nonserious AEs spontaneously, but there will be no solicitation for such events at these study sites.

Approximately 3000 of the subjects in Protocol 015 will be enrolled into the consistency lot substudy. The subjects in the consistency lot substudy will be randomized in a 1:1:1:3 ratio to receive Consistency Lot 1 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 2 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 3 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, or placebo. Therefore, among the 3000 subjects enrolled in the substudy, 1500 will be randomized to receive the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, and 1500 will be randomized to receive placebo. The placebo subjects will be enrolled in the consistency lot substudy only to maintain blinding and to maintain the 1:1 ratio of subjects in the vaccine and placebo treatment groups for the larger efficacy study. These subjects will not be used to address the immunogenicity objectives of the substudy. The 1500 subjects randomized to receive the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine will be randomized in a 1:1:1 ratio to receive 1 of 3 consistency lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine.

The efficacy study employs a fixed-number-of-events design. An interim analysis addressing the primary efficacy hypothesis of the study is planned to be conducted at the time that at least 19 cases of CIN 2/3 or invasive cervical cancer related to HPV 16 or 18 have been observed. The final efficacy analysis will be conducted when at least 29 cases have been observed.

At the time of the interim analysis of Protocol 015, an additional "interim" analysis will be conducted using combined CIN 2/3 or worse data from 4 protocols within the HPV vaccine program: Protocols 005, 007, 013, and 015. The combined data set will include the complete data from Protocols 005, 007, and 013 and the interim data from Protocol 015. Like the interim analysis of Protocol 015, this analysis will assess the efficacy of the vaccine in reducing HPV 16- or 18-related CIN 2/3 or invasive cervical cancer. The analysis of the combined data sets will have sufficient power to show a vaccine efficacy >25%, however.

If both the interim analysis of Protocol 015 and the interim analysis of the combined data sets meet the statistical criteria for success prespecified in the data analysis plans/protocol, the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine may be submitted for regulatory review prior to the conclusion of this study (Protocol 015). If this occurs, the remainder of the study will be considered an extension. The final efficacy analysis will be conducted regardless of the results of the interim analysis. If the vaccine is submitted for regulatory review based on the interim results, the data collected following the interim analysis will be used to obtain more precise estimates of the efficacy parameters of primary and secondary interest and will be combined with data from Protocols 005, 007, and 013 for an analysis of the efficacy of the vaccine in reducing CIN 2/3 or worse due to any HPV type.

If either of the interim analyses does not meet the statistical criteria for success prespecified in the data analysis plans/protocol, the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine will be submitted for regulatory review based on the final analysis of Protocol 015 and the combined CIN 2/3 data set. The final efficacy analysis of

Protocol 015 will be conducted when at least 29 cases of the primary endpoint have been observed. The target number of cases should be reached at approximately the same time the study is complete.

Endpoints

Safety: The primary variables of interest for safety/tolerability are serious vaccine-related adverse experiences, if any occur, and severe, injection-site adverse experiences. Unless they are considered serious adverse events, reports of injection-site adverse experiences will only be solicited from subjects participating in the nonserious adverse experience substudy and from subjects enrolled in the United Kingdom. Subjects from any study site may spontaneously report such events, however.

Efficacy: The primary variable of interest for efficacy is the combined incidence of HPV 16-related CIN 2/3 or worse and HPV 18-related CIN 2/3 or worse. This endpoint will occur **if on any single biopsy, endocervical curettage (ECC) or Loop Electrosurgical Excision Procedure (LEEP)/conization tissue block,** the following occur:

- Pathology panel consensus diagnosis of: CIN 2, CIN 3 (including squamous carcinoma in situ), adenocarcinoma in situ, invasive squamous cervical carcinoma, or invasive adenocarcinoma of the cervix,

  AND

- Detection of HPV 16 and/or HPV 18 by biopsy Thinsection PCR in an adjacent section from the same tissue block.

Of secondary interest are the incidences of colposcopic biopsy and definitive excisional cervical procedures (LEEP, laser conization, cold-knife conization) performed due to HPV 16- and 18-related cervical disease and the overall incidence of ALL CIN 2/3 and invasive cervical cancer (caused by any vaccine or non-vaccine HPV type).

Immunogenicity (Consistency Lot Substudy): For immunogenicity, the following endpoints are of primary interest: (1) the percentage of subjects who achieve anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL at Week 4 Postdose 3 and (2) the geometric mean titers (GMTs) to HPV 6, 11, 16, and 18 at Week 4 Postdose 3. Of secondary interest are the GMTs to HPV 6, 11, 16, and 18 at Months 24 and 48. The anti-HPV 6, 11, 16, and 18 responses at Months 7, 24, and 48 will only be measured in the subjects participating in the consistency lot substudy.

Additional exploratory efficacy and immunogenicity endpoints will be summarized. These endpoints are described in the main body of the Data Analysis Plan in Section III.A. for efficacy and Section IV.A. for immunogenicity.

Data Analysis and Decision Criterion for Success of the Study

Success or failure in the consistency lot substudy and in the main CIN 2/3 efficacy study are independent. That is, if the consistency lot substudy fails to meet its primary endpoint, the main CIN 2/3 efficacy study may still be declared a success. It is expected that both the main efficacy study and the consistency lot substudy will be successful. In fact, there is a notable advantage to demonstrating both efficacy and consistency within the same study.

Efficacy: The primary efficacy analysis will be per-protocol. To address the primary efficacy hypothesis of the study, a one-sided test of the null hypothesis that the vaccine efficacy (defined as 100[1 - Relative Risk]) is ≤0% will be conducted. The alternative hypothesis states that the vaccine efficacy is >0%. A point estimate of the vaccine efficacy and the corresponding multiplicity-adjusted two-sided confidence interval will be provided as well as a p-value for the test. Rejection of the null hypothesis in the hypothesis test (i.e., success of the hypothesis test) will correspond to the lower bound of the confidence interval for the vaccine efficacy exceeding 0%. An exact analysis will be used. Such an analysis will not adjust for demographic factors such as age and number of lifetime sexual partners but will account for the person-years at risk.

A multiplicity adjustment will be made to account for the 2 separate efficacy analyses of the data (the interim and the final). Table 1 gives the effective two-sided alpha level and power for each analysis, assuming various values for the true vaccine efficacy. The overall two-sided alpha level for the study is controlled at the 0.05 level.

Table 1

Power Sensitivity Analysis

Analysis	α-Level	Number of Cases	True Vaccine Efficacy	Power
Interim	0.0204	19	80%	80%
			85%	90%
			90%	97%
Final	0.0411	29	80%	95%
			85%	99%
			90%	>99%

Consistency Lot Substudy: The primary immunogenicity analysis will also be per-protocol. The first primary immunogenicity hypothesis, regarding consistency of the 3 lots of quadrivalent HPV vaccine with respect to the percentages of subjects who achieve anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL at Week 4 Postdose 3, will be addressed by 3 pairwise comparisons for each HPV type (12 comparisons in total). For each HPV type, each pairwise comparison will test the equivalence of 2 of the 3 lots

(within 10 percentage points) using 2 one-sided tests at the 0.05 level. This criterion requires that the two-sided 90% confidence interval for the difference in rates be entirely contained within the interval (-10%, 10%). This criterion will ensure that the upper bound of the confidence interval for the maximum absolute difference in rates between any 2 of the 3 lots is <10 percentage points. The assumed response rate to each HPV type in the vaccine is 90% for each lot.

The second primary immunogenicity hypothesis, regarding consistency of the 3 lots of quadrivalent HPV vaccine with respect to the GMTs to HPV 6, 11, 16, and 18 at Week 4 Postdose 3, will also be addressed by 3 pairwise comparisons for each HPV type (12 comparisons total). Each pairwise comparison will test the equivalence of 2 of the 3 lots (within 2-fold) using 2 one-sided tests at the 0.05 level. This criterion requires that the two-sided 90% confidence interval for the ratio of GMTs be entirely contained within the interval (0.5, 2.0). An analysis of variance (ANOVA) model will be used with a response variable of the natural log of the individual titers and fixed effects for study center and quadrivalent HPV lot. The assumed standard deviation of the natural log titers is 1.2.

If equivalence can be established in all 3 pairwise comparisons for a given HPV type and endpoint, then the 3 lots will be considered consistent for that HPV type and endpoint. Equivalence must be established for both endpoints (rates and GMTs) and for all 4 vaccine HPV types for the 3 quadrivalent HPV lots to be considered consistent. The overall type I error rate for this consistency testing will be controlled at the 0.05 level. The overall power for the consistency lot substudy is 93%.

All subjects who received at least 1 injection and have follow-up data will be included in the summary of serious adverse experiences. All subjects in the NSAE substudy and in the United Kingdom who received at least 1 injection and have follow-up data will be included in the safety summaries of nonserious adverse experiences. Non-serious adverse experiences among subjects who received at least 1 injection and who are not in the NSAE substudy nor enrolled in the United Kingdom will be listed separately. An overall adverse experience profile will be provided as well as summaries of specific adverse experiences reported during the study.

I. INTRODUCTION

A. Objective of the Data Analysis Plan

This data analysis plan (DAP) is intended to be a comprehensive and detailed description of the strategy, rationale, and statistical techniques that will be used to address the clinical safety, tolerability, and immunogenicity of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine and the efficacy of the vaccine in the prevention of HPV 16 and 18-related CIN 2/3 or worse in 16- to 23-year-old women. Protocol 015 and Amendment 015-01 are covered by this DAP. The protocol is entitled, "A Randomized, Worldwide, Placebo-Controlled, Double-Blind Study to Investigate the Safety, Immunogenicity, and Efficacy on the Incidence of HPV 16/18-Related CIN 2/3 or Worse of the Quadrivalent HPV (Types 6, 11, 16, 18) L1 Virus-Like Particle (VLP) Vaccine in 16- to 23-Year-Old Women—The FUTURE II Study (Females United to Unilaterally Reduce Endo/Ectocervical Disease)."

B. Description of the Study and Objectives/Hypotheses

1. Study Design

Protocol 015 is a prospective randomized, double-blind, placebo-controlled, parallel, multicenter study operating under in-house blinding procedures. Embedded within Protocol 015 is a conistency lot substudy. Approximately 11,500 subjects will be randomized in a 1:1 ratio to receive either 3 injections of quadrivalent HPV (Type 6, 11, 16, 18) L1 VLP vaccine or 3 injections of placebo. Within the quadrivalent HPV vaccine group, subjects will be further randomized to receive 3 different lots of the vaccine in a 1:1:1 ratio. The actual lots received will differ between subjects enrolled in the consistency lot substudy and the remaining subjects in Protocol 015. Each dose of quadrivalent vaccine will contain 20 µg HPV 6 L1 VLP, 40 µg HPV 11 L1 VLP, 40 µg HPV 16 L1 VLP, and 20 µg HPV 80 L1 VLP. Vaccine or placebo is administered at 0, 2, and 6 months. Approximately 1 month following the third vaccination visit (Month 7), 6 months following the third vaccination visit (Month 12) and every year thereafter until Month 48, subjects will return to the study center for collection of specimens, gynecologic examinations, and Pap tests.

While most subjects in Protocol 015 will be followed only for the occurrence of serious adverse experiences (SAEs) following vaccination, the subjects who enroll in the United States (~1150 of the subjects [10%]) will participate in a nonserious adverse experience (NSAE) substudy. These subjects will undergo a full assessment of nonserious adverse

B. Description of the Study and Objectives/Hypotheses (Cont.)

experiences that occur within 14 days following each vaccination visit in addition to the follow-up for serious adverse experiences. In addition, due to regulatory requirements in the United Kingdom, all subjects enrolled in the United Kingdom will be solicited for serious and nonserious adverse experiences that were observed during the 14 days following each vaccination as well. The method for collecting nonserious adverse experience data will be different in the NSAE substudy and in the United Kingdom. In the NSAE substudy, each subject will record any nonserious adverse experiences that she observes during the 14-day follow-up period on a vaccination report card that is sent home with her at the vaccination visit. She will return the report card to the study site at the next scheduled visit, and the data will be entered from the report card into the study database. (The data will have been captured in writing by the subject during the appropriate period, however.) In the United Kingdom, subjects will not receive a vaccination report card. However, at each visit after a vaccination visit, each subject will be asked by study site personnel to recall any nonserious adverse experiences that occurred during the 14-day period following the prior vaccination. Subjects enrolled in study sites outside of the United States and the United Kingdom may report nonserious adverse experiences spontaneously, but solicitation for such events will not be conducted at these study sites.

Approximately 3000 of the subjects in Protocol 015 will be enrolled into the consistency lot substudy. The subjects in the consistency lot substudy will be randomized in a 1:1:1:3 ratio to receive Consistency Lot 1 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 2 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 3 of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, or placebo. Therefore, among the 3000 subjects enrolled in the consistency lot substudy, 1500 will be randomized to receive the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, and 1500 will be randomized to receive placebo. The placebo subjects will be enrolled in the consistency lot substudy only to maintain blinding and to maintain the 1:1 ratio of subjects in the vaccine and placebo treatment groups for the larger efficacy study. These subjects will not be used to address the immunogenicity objectives of the substudy. The 1500 subjects randomized to receive the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine will be randomized in a 1:1:1 ratio to receive 1 of 3 consistency lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine.

With respect to the efficacy evaluation in Protocol 015, procedures performed at both scheduled and unscheduled visits will provide efficacy

B. Description of the Study and Objectives/Hypotheses (Cont.)

data. The procedures performed at scheduled visits for the purpose of efficacy data collection include serum sample collection at enrollment, collection of cervicovaginal specimens at enrollment and Month 7, and Pap testing at Month 7 and all subsequent visits. The procedures that are typically performed at unscheduled visits for the purpose of efficacy data collection are repeat Pap tests, colposcopy, and biopsy following Month 7.

For the primary immunogenicity evaluation in the consistency lot substudy, subjects in the substudy will have a serum sample collected at Month 7 in addition to the sample collected at enrollment. Subjects in the substudy will also have serum samples collected at Months 24 and 48 to address a secondary immunogenicity objective.

The serum samples collected at enrollment and Months 7, 24, and 48 will be tested for antibody to HPV vaccine types using a radioimmunoassay (RIA) for each vaccine HPV type. The cervicovaginal specimens collected at enrollment and Month 7 will be tested for HPV vaccine-type deoxyribonucleic acid (DNA) using type-specific PCR assays. The RIA results at enrollment and the PCR results from the cervicovaginal specimens collected at enrollment and Month 7 will be used to exclude subjects who are positive for vaccine HPV types at baseline or who develop infection before 1 month following the administration of the third dose of the vaccination regimen from the evaluation of vaccine efficacy. The postvaccination RIA results will be used to assess the immunogenicity of the vaccine.

The results of the Pap tests performed at Month 7 and subsequent visits will be used to identify subjects with HPV disease. Mandatory protocol-specified guidelines will be used to triage subjects with Pap abnormalities to colposcopy. Colposcopy is performed by an experienced colposcopist who has been trained in protocol-specific colposcopy procedures. If a lesion is seen during the colposcopy, it is biopsied.

All biopsies will be read by a pathologist at a central laboratory (b)(4) for the purpose of patient management. However, the biopsies will also be read by an independent panel of expert pathologists who will provide the final pathologic diagnosis for study purposes. To determine the causal HPV type within a cervical biopsy for efficacy assessments, Merck's type-specific HPV localizing assay (HPV Type-Specific Thinsection PCR Assay) will be performed on paraffin-embedded tissue samples.

B. Description of the Study and Objectives/Hypotheses (Cont.)

If the central laboratory diagnosis of a biopsy from a subject is CIN 2/3 or worse, the subject will be referred for definitive therapy. Sections of the tissue excised during the definitive therapy procedure will be treated in the same fashion as the biopsy specimens.

Though not a protocol-approved practice, it is possible that some subjects will have colposcopies performed by private physicians outside the context of the study rather than by the study sites. If cervical biopsies or tissue specimens are collected during such colposcopies, these tissue specimens will not be collected according to protocol-specified procedures and will be read by the local laboratory used by the private physician rather than the central laboratory for patient management. Nevertheless, study sites actively solicit information regarding whether or not such procedures have been conducted outside the study, and if these procedures have been conducted, every effort is made by the study site to obtain permission from the subject to access the local laboratory diagnosis of the tissue specimen, the slides prepared from the specimen and the tissue block itself. If the slides are obtained, they will be sent to the pathology panel for diagnosis. If the tissue block is obtained, sections will be sent for PCR analysis. Thus, when a cervical biopsy or tissue specimen is collected outside of the study, MRL will have access to: (1) the local laboratory diagnosis for the specimen only; (2) the local laboratory diagnosis and a pathology panel diagnosis for the specimen; (3) the local laboratory diagnosis, a pathology panel diagnosis, and PCR results for the specimen; or (4) none of these. Such specimens will not be used in the primary efficacy analysis but will be used in sensitivity analyses.

The efficacy study employs a fixed-number-of-events design. The final efficacy analysis will be conducted when at least 29 cases of CIN 2/3 or invasive cervical cancer related to HPV 16 or 18 have been observed.

An interim analysis addressing the primary efficacy hypothesis of the study is planned to be conducted at the time that at least 19 cases have been observed. The interim analysis of Protocol 015 will be conducted in conjunction with an interim analysis of the combined CIN 2/3 or worse data from Protocols 005, 007, 013, and 015. If both interim analyses meet the statistical criteria for success prespecified in the data analysis plans/protocol, the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine may be submitted for regulatory review prior to the conclusion of this study (Protocol 015) (See Section VI.A for details regarding the success criteria). If this occurs, the remainder of the study will be considered an extension. The final efficacy analysis will be conducted

B. Description of the Study and Objectives/Hypotheses (Cont.)

regardless of the results of the interim analysis. If the vaccine is submitted for regulatory review prior to the end of the study, the data collected following the interim analysis will be used to obtain more precise estimates of the efficacy parameters of primary and secondary interest and will be combined with the data from 3 other HPV vaccine studies for an analysis of the efficacy of the vaccine in reducing CIN 2/3 or worse due to any cause.

2. Study Reporting

The data from Protocol 015 will be analyzed and reported in 3 distinct stages.

a. Since Month 7 is the primary time point for the immunogenicity and safety analyses, these analyses will be conducted when all subjects in the consistency lot substudy have completed the Month 7 visit. (Since the consistency lot substudy will include the last 3000 subjects enrolled, all subjects in the study will have finished Month 7 by this time.)

b. The efficacy portion of the protocol employs a fixed event design, whereby interim and final primary analyses of efficacy will be conducted at the time that specific target numbers of cases of the primary endpoints are observed. Given the event rates assumed in the protocol, the target number of cases for the interim analysis is expected to be reached around the time that all subjects have completed their Month 30 visits (i.e., 2.5 years of follow-up). The target number of cases for the final analysis is expected to be reached around the time that all subjects have completed their Month 48 visits (i.e., 4 years of follow-up).

c. The efficacy data collected from Protocol 015 will be combined in a prespecified analysis with efficacy data from Protocols 005, 007, and 013 to evaluate the vaccine efficacy with respect to: (1) vaccine-type HPV-related CIN 2/3 and cervical carcinoma and (2) CIN 2/3 and cervical cancer related to all HPV types. The interim and final analyses of the combined studies with respect to these efficacy endpoints is also endpoint driven. To accrue an adequate number of cases for the combined efficacy evaluation of CIN 2/3 or cancer related to all HPV types, the subjects in Protocol 015 will need to be followed for ~4 years.

B. Description of the Study and Objectives/Hypotheses (Cont.)

Success on the interim analyses of Protocol 015 and the combined efficacy data set could trigger a regulatory submission (see Section VI.A.). If a regulatory submission is prepared at this time, the time period between the interim and final analyses of Protocol 015 will be considered an extension phase. With respect to the objectives of this protocol, at the end of the extension phase, the vaccine efficacy for the primary and secondary endpoints will be re-estimated in order to refine the precision of the estimates. If a submission is prepared using the interim data from Protocol 015, then the efficacy results from the interim and final analyses will appear in separate reports. The plan for the CIN 2/3 efficacy analysis in the combined data sets from Protocols 005, 007, 013, and 015 appears in a separate DAP, and the results of this analysis will also appear in a separate report.

The details of maintaining the study blinding for the various stages of reporting are discussed in Section VII.C.

3. Primary Objectives

Safety (Overall Study)

To demonstrate that a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine is generally well tolerated.

Efficacy (Overall Study)

To demonstrate that intramuscular administration of a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine reduces the incidence of the composite endpoint of HPV 16- and 18-related high-grade cervical abnormalities (CIN 2/3) or HPV 16- and 18-related invasive cervical carcinoma in subjects who are PCR negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for the relevant HPV type.

Immunogenicity (Consistency Lot Substudy)

To demonstrate that the Final Manufacturing Process (FMP) results in quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine that, when given in a 3-dose regimen, induces consistent serum anti-HPV 6, anti-HPV 11, anti-HPV 16, and anti-HPV 18 responses 4 weeks Postdose 3.

B. Description of the Study and Objectives/Hypotheses (Cont.)

4. Secondary Objectives

Efficacy (Overall Study)

To estimate the impact of the administration of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine on the rates of colposcopic biopsy and definitive excisional cervical procedures (LEEP, laser conization, cold-knife conization) performed due to HPV 16- and HPV 18-related disease.

Immunogenicity (Consistency Lot Substudy)

To evaluate the persistence of vaccine-induced serum anti-HPV 6, anti-HPV 11, anti-HPV 16, and anti-HPV 18 responses in subjects who are PCR negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for the relevant HPV type.

5. Exploratory Objectives

Efficacy (Overall Study)

To estimate the impact of the administration of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine on the incidence of the composite endpoint of ALL CIN 2/3 or invasive cervical carcinoma (caused by any vaccine or non-vaccine HPV type) in subjects who are PCR negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for high risk HPV types.

6. Primary Hypotheses

Safety (Overall Study)

The quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine is generally well tolerated in 16- to 23-year-old females.

Efficacy (Overall Study)

Administration of a 3-dose regimen of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine reduces the incidence of the composite endpoint of HPV 16- or HPV 18-related CIN 2/3 or invasive cervical carcinoma compared with placebo in subjects who are PCR negative and seronegative at baseline and PCR negative 1 month after completion of the vaccination series for the relevant HPV type. *(The statistical criterion for success requires that the lower bound of the confidence interval for the vaccine efficacy exclude 0%.)*

B. Description of the Study and Objectives/Hypotheses (Cont.)

Immunogenicity (Consistency Lot Substudy)

a. Three separate lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine induce similar immune responses, as measured by the percentage of subjects who achieve serum anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL, at Week 4 Postdose 3. *(Each vaccine component will be analyzed separately. The statistical criterion for similarity requires that the upper bound of the confidence interval for the maximum absolute difference in proportions between any 2 of the 3 lots exclude 10 percentage points or more for each HPV type.)*

b. Three separate lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine induce similar immune responses, as measured by the serum geometric mean titers (GMTs) to HPV 6, 11, 16, and 18, at Week 4 Postdose 3. *(Each vaccine component will be analyzed separately. The statistical criterion for consistency requires that the upper bound of the confidence interval for the fold-difference in GMTs between any 2 lots exclude a fold-difference of 2 or greater for each HPV type.)*

II. STUDY PARTICIPANTS CHARACTERISTICS

Subject characteristics of age, gender, race/ethnicity, number of lifetime male and female sexual partners, tobacco use, pregnancy history, age at first sexual intercourse, lifetime gynecologic medical history, history of sexually transmitted diseases (STDs), and contraceptive use will be summarized by treatment group and overall for all subjects who entered the study. In addition, baseline (Day 1) serostatus for HPV 6, 11, 16, and 18, PCR status for HPV 6, 11, 16, and 18 at Day 1 and Month 7, baseline (Day 1) Pap test results, and baseline (Day 1) presence of other STDs will be presented by treatment group and overall. Co-infections with multiple vaccine HPV types will be also summarized by treatment group and overall. Balance between treatment groups with respect to subject characteristics will be determined by observation.

The number and percentage of subjects with specific prior medications within 3 days prior to the first vaccination will be summarized by treatment group for all specific prior medications reported by ≥1% of the subjects in either treatment group. Similarly, the number and percentage of subjects with specific concomitant medications within 14 days following any vaccination will be summarized by treatment group for all specific concomitant medications reported by ≥1% of the subjects in either treatment group. Again, balance between treatment groups will be assessed by observation only.

III. EFFICACY ANALYSES

A. Efficacy Endpoints

The primary variable of interest for efficacy is the combined incidence of HPV 16-related CIN 2/3 or worse and HPV 18-related CIN 2/3 or worse. This endpoint will occur **if on any single biopsy, ECC, or LEEP/conization tissue block,** the following occur:

- Pathology panel consensus diagnosis of: CIN 2, CIN 3 (including squamous carcinoma in situ), adenocarcinoma in situ, invasive squamous cervical carcinoma, or invasive adenocarcinoma of the cervix,

AND

- Detection of HPV 16 and/or HPV 18 by biopsy Thinsection PCR in an adjacent section from the same tissue block.

Of secondary interest are the incidences of colposcopic biopsy and definitive excisional cervical procedures (LEEP, laser conization, cold-knife conization) performed due to HPV 16- and 18-related cervical disease.

Of primary exploratory interest is the overall incidence of ALL CIN 2/3 and invasive cervical cancer (caused by any vaccine or non-vaccine HPV type). Several other exploratory endpoints will be evaluated including: (1) CIN 2/3 or worse related to HPV 16 and 18 separately; (2) vaccine-type-HPV-related CIN 1 or worse; (3) CIN 1 or worse due to any cause; (4) HPV 6/11-related genital warts; (5) any HPV 6/11/16/18-related lesion; (6) any HPV-related lesion; (7) colposcopy; (8) colposcopic biopsy due to any CIN; (9) definitive therapy due to any CIN; and (10) Pap abnormalities. In addition, the antibody responses in vaccine recipients who have breakthrough cases of HPV 16/18-related CIN 2/3 or worse will be evaluated. The potential therapeutic effects of administering the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine to subjects who are already infected at baseline with a vaccine HPV type, but who have not yet mounted an immune response to the HPV infection, will be assessed. The time-to-event (or disease-free time) distributions in the vaccine and placebo groups will be compared. Finally, an observational assessment of a potential waning in vaccine efficacy will be performed.

B. Study Participant Populations for Efficacy Analyses

Four patient populations will be considered for the efficacy analyses: 1 per-protocol population and 3 modified intention-to-treat (MITT) populations.

B. Study Participant Populations for Efficacy Analyses (Cont.)

The efficacy analysis using the per-protocol population will be the primary analysis. Supplemental analyses of the primary and secondary efficacy endpoints will be provided using the 3 MITT populations. The populations that will be considered for the exploratory efficacy endpoints are described in their respective sections.

The per-protocol and MITT populations differ with regard to the inclusion/exclusion of protocol violators who are described below.

Per-Protocol Population

To be included in the per-protocol analyses, subjects must:

1. receive all 3 injections with the correct dose of the correct clinical material;

2. be seronegative by RIA to the appropriate HPV types before the first injection and PCR-negative to the appropriate HPV types through Month 7;

3. not receive any nonstudy inactivated vaccine within 14 days of a study vaccine or any nonstudy live virus vaccine within 21 days before or 14 days after a study vaccine;

4. not receive immune globulin (including RhoGAM™ [Ortho-Clinical Diagnostics Inc.]) or blood-derived products at any time through Month 7 of the study;

5. not receive immunosuppressives or have an immune disorder considered by the Clinical Monitor to potentially interfere with the subject's response to the vaccine;

6. not concurrently participate in any other clinical studies of investigational agents or clinical studies involving collection of cervical specimens which, in the opinion of the Clinical Monitor, may potentially interfere with the subject's response to the vaccine;

7. have a Month 7 visit within a day range considered acceptable for defining the subject's Month 7 PCR status.

Only subjects who are HPV 16 seronegative by RIA at enrollment and HPV 16 PCR-negative from enrollment through Month 7 will be eligible to be

B. Study Participant Populations for Efficacy Analyses (Cont.)

counted as cases of HPV 16-related CIN 2/3 or cervical cancer. Only subjects who are HPV 18 seronegative by RIA at enrollment and HPV 18 PCR-negative from enrollment through Month 7 will be eligible to be counted as cases of HPV 18-related CIN 2/3 or cervical cancer. Therefore, to be included in the per-protocol population for the primary efficacy analysis regarding CIN 2/3 and cervical cancer related to HPV 16 and 18, subjects are required to be seronegative by RIA at enrollment and PCR-negative from enrollment through Month 7 to HPV 16, to HPV 18, or to both. The same population will be considered for the secondary analysis of biopsy and definitive therapy procedures due to HPV 16- and 18-related disease.

For the exploratory efficacy estimates regarding CIN 1, CIN 2/3 and cervical cancer due to any cause, again only subjects who are HPV 16 seronegative by RIA at enrollment and HPV 16 PCR-negative from enrollment through Month 7 will be eligible to be counted as cases of HPV 16-related CIN 2/3 or cervical cancer. The same rules apply for HPV 18. For HPV types 31, 33, 35, 45, 51, 52, 55, 56, 58, 59, and 68, only subjects who are PCR-negative from enrollment through Month 7 for a given HPV type will be eligible to be counted as cases of CIN 1, CIN 2/3, or cervical cancer related to that type. No baseline serology testing will be performed for these 11 high-risk HPV types. Therefore, all subjects who are PCR-negative from enrollment through Month 7 for 1 of the types, even though they may be baseline seropositive for that type, will be eligible to be counted as cases of CIN 1, CIN 2/3, or cervical cancer related to that type. (This approach is conservative, since subjects who have baseline positive serology for a given HPV type have already mounted an immune response to that type.) With respect to disease endpoints related to high-risk HPV types that are not among the 13 listed above, subjects who meet criteria (1) and (3) through (7) above, have normal Pap results from Day 1 through Month 7, and have a history of normal Pap test results will be eligible to meet the case criteria. A sensitivity analysis will be performed in which all subjects who meet criteria (1) and (3) through (7) above (regardless of Pap test results from Day 1 through Month 7 and Pap history) will be eligible to meet the case criteria.

Modified Intention-to-Treat Populations

The first MITT analysis will include all subjects who are seronegative at enrollment and PCR-negative from enrollment through Month 7 to the appropriate HPV types, who receive all 3 vaccinations, and who have any follow-up visit following Month 7. The primary difference between this population and the per-protocol population is the inclusion of general protocol violators.

B. Study Participant Populations for Efficacy Analyses (Cont.)

The second MITT analysis will include all subjects who are seronegative and PCR-negative at enrollment to the appropriate HPV types, who receive at least 1 dose of quadrivalent HPV (Types 6, 11, 16, and 18) L1 VLP vaccine or placebo, and who have any follow-up visit after 1 month following the first injection.

The third MITT analysis will include all subjects who receive at least 1 dose of quadrivalent HPV (Types 6, 11, 16, and 18) L1 VLP vaccine or placebo and who have any follow-up visit after 1 month following the first injection, regardless of initial serology and PCR status.

The MITT populations consider incorrectly randomized subjects and subjects who received the incorrect clinical material in the analysis according to the treatment group to which they were randomized by the study allocation schedule.

C. Approaches to Efficacy Analysis

1. Primary Efficacy Hypothesis

The interim and final primary efficacy analyses will be conducted when fixed numbers of cases of the primary endpoint have been observed. Specifically, the interim analysis will be conducted when at least 19 cases of the primary endpoint have been observed, and the final analysis will be conducted when at least 29 cases of the primary endpoint have been observed.

To address the primary hypothesis of the study, a one-sided test of the null hypothesis that the vaccine efficacy (defined as 100[1 - Relative Risk]) is ≤0% will be conducted. The alternative hypothesis is that the vaccine efficacy is >0%. A point estimate of the vaccine efficacy and the corresponding multiplicity-adjusted two-sided confidence interval will also be provided. Rejection of the null hypothesis (i.e., success in the study) corresponds to the lower bound of the confidence interval exceeding 0%. An exact analysis will be used. Such an analysis will not adjust for extraneous factors such as age and number of lifetime sexual partners.

The estimate of the vaccine efficacy will account for the follow-up (i.e., person-time at risk) in the vaccine and placebo groups. Since the endpoint for the primary analysis is a composite endpoint, any subject who is eligible to be considered as an endpoint according to the HPV 16-related

C. **Approaches to Efficacy Analysis (Cont.)**

disease definition, the HPV 18-related disease definition, or both will be included in the population at risk for the primary analysis. The method that will be used to compute the follow-up time for each subject in the population at risk is described in Section VII.D.1.c.

In addition, a sensitivity analysis will be performed in which cervical biopsies and tissue specimens collected outside the study (i.e., not according to protocol-specified procedures) are included in the primary efficacy analysis if they meet the case criteria. A cervical biopsy or tissue specimen collected outside the study will be eligible to be counted as an endpoint for this analysis *only* if HPV 16 and 18 PCR results *and* a pathology panel diagnosis are available for the specimen.

2. Secondary Efficacy Analysis

The efficacy of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine in reducing the incidence of: (1) colposcopic biopsy performed due to HPV 16- and 18-related cervical disease; (2) definitive excisional cervical procedures (LEEP, laser conization, cold-knife conization) performed due to HPV 16- and 18-related cervical disease; and (3) either colposcopic biopsy or definitive excisional cervical procedures performed due to HPV 16- and 18-related cervical disease will be estimated using the same methodology that will be used for the primary analysis.

3. Exploratory Efficacy Analyses

a. CIN 2/3 Due to Any HPV Type

The efficacy of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine in reducing the overall incidence of ALL CIN 2/3 and invasive cervical cancer (caused by any vaccine or non-vaccine HPV type) will be estimated using the same methodology that will be used for the primary analysis. Only the point estimate of the vaccine efficacy and a 95% confidence interval for the efficacy with respect to this endpoint will be provided. No statistical testing will be conducted as there will be insufficient power to draw conclusions regarding this endpoint from the cases that are likely to be observed.

It is expected that all subjects will be included in the population at risk for the "all CIN 2/3" endpoint, since almost all subjects should be eligible to be an endpoint with respect to at least 1 of the 13 high-risk HPV types for which subjects are being tested.

C. Approaches to Efficacy Analysis (Cont.)

As a supplement to the "all CIN 2/3" evaluation, the (b)(4) swabs collected from subjects at Months 24, 36, and 48 may be tested (if resources permit) for non-vaccine HPV types by PCR in order to assess the potential "replacement" of HPV types in CIN 2/3 lesions as a result of vaccination. The swabs would likely only be analyzed in subjects who developed CIN 2/3.

A sensitivity analysis will be performed in which cervical biopsies and tissue specimens collected outside the study (i.e., not according to protocol-specified procedures) are included in the "all CIN 2/3" efficacy analysis if they meet the case criteria. A cervical biopsy or tissue specimen that is collected outside the study will be eligible to be counted as an endpoint for this analysis *only* if a pathology panel diagnosis is available for the specimen.

b. Supplementary Summaries of the Primary Endpoints

The vaccine efficacy with respect to the primary endpoint will be summarized by HPV type. A point estimate of the vaccine efficacy will be provided along with a 95% confidence interval using the methodology described above for the primary analysis. These estimates will be provided for descriptive purposes only in the per-protocol and all of the MITT populations. No statistical testing will be conducted.

In addition, to assess the impact of the vaccine on HPV 16/18-related CIN 2/3 as diagnosed by a laboratory whose purpose is to provide pathologic diagnoses for patient management, a sensitivity analysis will be conducted in which the central laboratory (b)(4) diagnosis rather than the pathology panel diagnosis of each biopsy is used for case identification. Cervical biopsies and tissue specimens collected outside of the study will be included in this analysis if a local laboratory diagnosis and PCR results are available. Since these specimens were collected outside of the study, the local laboratory diagnosis will be used to identify cases rather than a central laboratory (b)(4) diagnosis. This analysis will be performed in the per-protocol and the first MITT population.

c. Time-to-Event Distributions

For the primary endpoint, survival analysis techniques will be used to compare the time-to-event (or disease-free time) distributions between

C. **Approaches to Efficacy Analysis (Cont.)**

the vaccine and placebo groups. Kaplan-Meier (product limit) estimates of the time-to-event curves and the corresponding 95% confidence intervals will be displayed by treatment group, and the estimated time-to-event curves will be plotted. The time-to-event distributions between the 2 treatment groups will be compared using the log-rank test. The log-rank statistic and the corresponding p-value will be noted on the plot of the curves. For each subject who meets the endpoint definition, the time-to-event will be the time between the subject's Month 7 visit date and the earliest date a tissue sample is collected from the subject which meets the endpoint criterion. For all other subjects, the time between the subject's Month 7 visit date and the date the subject (1) drops out of the study, (2) has definitive therapy, or (3) completes the study will be used in the analysis. The time-to-event analysis will be conducted for the per-protocol and the 3 MITT populations.

d. Correlates of Protection

An exploratory analysis of any breakthrough HPV 16/18-related CIN 2/3 or worse cases (cases of HPV 16/18-related CIN 2/3 or worse in vaccine recipients) will be performed to try to establish the relationship between immune markers and protection from disease endpoints. Specifically, the immune responses in the non-breakthrough vaccinees will be compared with the immune responses in the breakthrough cases at all available time points (i.e., at Months 7, 24, and 48). However, if there are few cases among the vaccine recipients, there may be insufficient information to establish an immune correlate of protection in this study.

Immune markers will be explored using the per-protocol population. However, if the vaccine is highly efficacious, very few primary endpoints will be observed in the vaccine group in the per-protocol analysis. If this is the case, it may be more informative to explore correlates of protection in the second MITT population. This population still includes only initially HPV naïve subjects but is likely to yield the most endpoints in the vaccine group.

e. CIN 1 or Worse

The vaccine efficacy with respect to the incidence of HPV 16- or 18-related CIN 1 or worse and HPV 6-, 11-, 16- or 18-related CIN 1 or worse will be estimated. In addition, the vaccine efficacy with respect

C. **Approaches to Efficacy Analysis (Cont.)**

to all CIN (any grade due to any cause) will be estimated. Point estimates and confidence intervals for the vaccine efficacy will be provided using the methodology described above for the primary analyses. The HPV 16- or 18-related CIN 1 or worse endpoint and the HPV 6/11/16/18-related CIN 1 or worse endpoint will be summarized in the per-protocol and the first MITT population. To be included in the analysis of CIN due to any cause, subjects will be required to be seronegative and PCR negative from Day 1 through Month 7 for relevant HPV types or have normal Pap test results from Day 1 through Month 7 and a history of normal Pap test results (see Section III.B.). Only point estimates of the vaccine efficacy and 95% confidence intervals for the efficacy with respect to these endpoints will be provided. No statistical testing will be conducted.

f. Other HPV-Related Lesions

The impact of the vaccine on other HPV-related lesions and other combinations of HPV-related lesions will also be evaluated. The person-time based incidences of: (1) HPV 6- or 11-related external genital warts; (2) any HPV 6, 11, 16, or 18-related lesion; and (3) any HPV-related lesion will be estimated by treatment group. Corresponding 95% confidence intervals for the incidence rates will be provided. To be eligible to be counted as an HPV 6/11-related endpoint, subjects will be required to be seronegative at Day 1 and PCR negative from Day 1 through Month 7 for HPV 6 and HPV 11. To be eligible to be counted as an HPV 16- or 18-related endpoint, subjects will be required to be seronegative at Day 1 and PCR negative from Day 1 through Month 7 for the relevant type. To be eligible to be counted as an endpoint related to other HPV types, subjects will be required to be PCR negative from Day 1 through Month 7 for the relevant HPV types or have normal Pap test results from Day 1 through Month 7 and have a history of normal Pap test results (see Section III.B.). Only point estimates of the incidence rates and 95% confidence intervals for the rates with respect to these endpoints will be provided. No statistical testing will be conducted.

g. Gynecologic Procedures

The efficacy of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine in reducing the incidence of: (1) colposcopy, (2) colposcopic biopsy due to any CIN, and (3) definitive therapy due to any CIN will be estimated using the same methodology that will be used for the

C. **Approaches to Efficacy Analysis (Cont.)**

primary analysis. Only point estimates of the vaccine efficacy and 95% confidence intervals for the efficacy with respect to this endpoint will be provided. Estimates will be provided using the per-protocol and all of the MITT populations. No statistical testing will be conducted.

h. Pap Abnormalities

Person-time based incidences of the following Pap abnormalities will be summarized by treatment group: (1) high-grade squamous intraepithelial lesion (HSIL) or worse, (2) low-grade squamous intraepithelial lesion (LSIL) or worse, (3) atypical squamous cells of undetermined significance (ASC-US) with positive HPV probe or worse (atypical squamous cells suggestive of a high-grade intraepithelial lesion [ASC-H], HSIL, LSIL, Atypical Glandular Cells), and (4) ASC-US or worse. To be included in these summaries, subjects will be required to be seronegative and PCR negative from Day 1 through Month 7 to all HPV types for which subjects are tested, have normal Pap test results from Day 1 through Month 7, and have a history of normal Pap test results. The incidences of the above Pap abnormalities will also be summarized in subjects who are seronegative and PCR negative at Day 1 to all HPV types for which subjects are tested, have normal Pap test results at Day 1, and have a history of normal Pap test results. Only point estimates of the incidence rates and 95% confidence intervals for the rates with respect to these endpoints will be provided. These evaluations will be conducted on the subsets of subjects above both including and excluding general protocol violators. No statistical testing will be conducted.

i. Potential Therapeutic Effect of the Vaccine

Data permitting, an exploratory analysis to assess the potential therapeutic effects of administering the quadrivalent vaccine to subjects who are already infected with a vaccine HPV type, but who have not yet mounted an immune response to the HPV infection at baseline, will be performed. Specifically, the vaccine's impact in reducing the incidence of (1) HPV 16- and 18-related CIN 2/3 or worse and (2) HPV 6-, 11-, 16- and 18-related CIN 1 or worse will be evaluated in these subjects.

C. **Approaches to Efficacy Analysis (Cont.)**

For example, among subjects HPV 16 seronegative and PCR positive at Day 1 or HPV 18 seronegative and PCR positive at Day 1, the incidence of CIN 2/3 related to the relevant HPV type will be estimated in each treatment group and the 95% confidence intervals for these rates will be provided. Only cases of CIN that are detected following the Month 7 visit will be included in this analysis.

j. **Assessment of Potential Waning of Vaccine Efficacy**

Due to the timing between the development of HPV infection and the development of clinical sequelae, it is expected that the target number of cases required to estimate the vaccine efficacy with respect to the primary endpoint of the study at the interim analysis (19 cases) will be observed ~2.5 years into the study. By Year 4, it is anticipated that ~10 to 12 additional cases will be observed. Thus, there is limited power in this study to draw any definitive conclusions regarding changes in vaccine efficacy over time. Nevertheless, an exploratory analysis will be conducted to assess a possible waning of the vaccine efficacy. Specifically, at the conclusion of the study, the vaccine efficacy by year will be estimated, and any observed trends in the efficacy over time will be reported. The first year estimate will include all follow-up through Month 12 for each subject; the second year estimate will include all follow-up through Month 24 for each subject; the third year estimate will include all follow-up through Month 36 for each subject; and the final estimate will include all follow-up through Month 48 for each subject. A potential waning of vaccine efficacy will be explore using the per-protocol and first MITT populations.

D. **Definition of Compliance Measure**

Treatment compliance is defined in this study as receipt of all scheduled study vaccinations. To summarize treatment compliance, the numbers of subjects who receive each vaccination will be tabulated by treatment group. As stated in Section III.B., subjects who do not complete the full vaccination regimen will be excluded from the per-protocol primary analyses of efficacy but will be included in 2 MITT analyses.

Another important compliance measure in this study is the degree of subject compliance with follow-up visits following the completion of the vaccination series. Since colposcopy is a required procedure for the identification of potential study endpoints, subjects who miss study follow-up visits or who

D. Definition of Compliance Measure (Cont.)

visit their private gynecologists for examinations and/or treatment rather than the study investigators represent missed opportunities to observe endpoints. If these subjects develop HPV 16- or 18-related CIN 2/3 or worse during the time that they missed study visits or if they had HPV 16- or 18-related CIN 2/3 or worse identified and treated outside of the study, then these events would not be captured in the primary estimate of the vaccine efficacy in this study.

To summarize compliance with respect to follow-up study visits, the numbers of subjects who complete each follow-up visit will be tabulated by treatment group. To provide measures of compliance with respect to follow-up study visits, the average interval between scheduled study visits will be computed for each subject and across subjects within a treatment group. The number of subjects with intervals longer than 9 months will be summarized by treatment group. In addition, the percentages of subjects in each group who had biopsies or excision procedures performed outside of the study will be summarized, and the percentages of subjects in each group who had biopsies or excision procedures performed outside of the study for whom the tissue samples were unavailable to Merck will be summarized. Differences in these measures between the treatment groups will be assessed observationally and the potential impact on the efficacy analyses will be noted. In addition, a sensitivity analyses for assessing the potential impact of missing data on the primary efficacy analysis is proposed in Section VII.D.1.f.

IV. IMMUNOGENICITY ANALYSES

A. Immunogenicity Endpoints

For immunogenicity, the endpoints are of primary interest are:

1. the percentage of subjects who achieve anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL at Week 4 Postdose 3 and

2. the GMTs to HPV 6, 11, 16 and 18 at Week 4 Postdose 3.

Of secondary interest are the GMTs to HPV 6, 11, 16 and 18 at Months 24 and 48. These variables will be measured only in the subjects participating in the consistency lot substudy.

Exploratory endpoints of interest are: (1) the immune responses to each vaccine HPV type in subjects who are baseline positive for that type and (2) the associations among vaccine HPV types.

B. Study Participant Populations for Immunogenicity Analyses

Two subject populations will be considered for the immunogenicity analyses: the per-protocol population and the "all type-specific HPV naïve subjects with serology data" population. The immunogenicity analyses using the per-protocol population will be the primary analyses. Supportive analyses will be performed using the "all type-specific HPV naïve subjects with serology data" population. The 2 populations differ with regard to the inclusion/exclusion of protocol violators as described below.

Per-Protocol Immunogenicity Population

For immunogenicity, subjects in the per-protocol population must meet all of the same inclusion criteria as described in Section III.B. for the per-protocol efficacy population. In addition, they must receive all 3 vaccinations within acceptable day ranges (outlined in Section VIII.A.), and they must have the postvaccination serum samples collected within acceptable day ranges (outlined in Section VIII.A.).

"All Type-Specific HPV Naïve Subjects With Serology Data" Population

The "all type-specific HPV naïve subjects with serology data" population will include all subjects who receive all 3 vaccinations, who are seronegative by RIA to the appropriate HPV vaccine component(s) at enrollment, who are PCR-negative to the appropriate HPV vaccine component(s) from enrollment through Month 7, and who have a valid serology result after the 3rd injection. This population includes general protocol violators and considers incorrectly randomized subjects and subjects who received the incorrect clinical material in the analysis according to the treatment group to which they were randomized by the study allocation schedule. This population will be considered for the primary and secondary analyses and is considered a secondary approach.

C. Approaches to Immunogenicity Analysis

1. Primary Immunogenicity Hypotheses

Success must be achieved on both hypotheses for the study to be a success.

The first primary immunogenicity hypothesis, regarding consistency of the immune responses to quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine as measured by the percentages of subjects with anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL and anti-HPV 18 ≥200 mMU/mL, will be addressed by

C. **Approaches to Immunogenicity Analysis (Cont.)**

3 pairwise comparisons (1 for each possible pairing of the 3 consistency lots) for each vaccine HPV type individually (12 comparisons in total). For each vaccine HPV type, each pairwise comparison is intended to demonstrate equivalence of the 2 lots (within an equivalence margin of 10 percentage points) using 2 one-sided tests at the $\alpha = 0.05$ level. This criterion is equivalent to requiring that the two-sided 90% confidence interval for the difference in percentages between lots be entirely contained within (-10%, 10%). These comparisons will be tested using the method of Miettinen and Nurminen [1] stratified by study center.

The second primary immunogenicity hypothesis, regarding consistency of the immune responses to quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine as measured by the GMTs to HPV 6, 11, 16 and 18, will also be addressed by 3 pairwise comparisons (1 for each possible pairing of the 3 consistency lots) for each vaccine HPV type individually (12 comparisons in total). For each vaccine HPV type, each pairwise comparison is intended to demonstrate equivalence of the 2 lots (within an equivalence margin of 2-fold) using 2 one-sided tests at the $\alpha = 0.05$ level. This criterion is equivalent to requiring that the two-sided 90% confidence interval for the ratio of the GMTs for the lots be entirely contained within (0.5, 2). Using the data from all 3 lots, an ANOVA model will be constructed for each vaccine HPV type with a response variable of natural log titer and fixed effects for study center, lot, and lot-by-study center interaction. Pairwise comparisons between lots will be made using the mean squared error (MSE) from the final ANOVA model as an estimate of variance.

If equivalence can be established in all 3 pairwise comparisons for a given HPV type and endpoint, the 3 lots will be considered consistent for that HPV type and endpoint. Equivalence must be established for both endpoints (rates and GMTs) and for all 4 HPV types for the 3 quadrivalent HPV lots to be considered consistent.

2. **Secondary Immunogenicity Analysis**

To evaluate the persistence of vaccine-induced serum anti-HPV 6, anti-HPV 11, anti-HPV 16 and anti-HPV 18 responses, GMTs and the proportions of subjects with anti-HPV ≥200 mMU/mL for each vaccine HPV type will be summarized at Month 7, Month 24, and Month 48. Graphical displays of the distribution of titers among subjects administered each lot of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine will be provided by HPV type.

C. Approaches to Immunogenicity Analysis (Cont.)

3. Exploratory Immunogenicity Analyses

a. Immune Responses to Each Vaccine HPV Type in Subjects Who Are Baseline Seropositive for the Given Type

Assessments in this section will be performed only when data permits; that is, when the number of subjects in the subgroup being analyzed is sufficient.

To explore a potential association between baseline seropositivity for a given vaccine HPV type and improved Postdose 3 immunogenicity for the same HPV type (immune memory), the proportions of subjects with anti-HPV ≥200 mMU/mL and the GMTs at Month 7 for each vaccine HPV type will be summarized in subjects who are seropositive at Day 1 for that vaccine HPV type (6, 11, 16, or 18), regardless of PCR status. The responses will be compared observationally with the proportions of subjects with anti-HPV ≥200 mMU/mL and the GMTs at Month 7 in (1) subjects who are seronegative at Day 1 and PCR-negative from Day 1 through Month 7 for the vaccine HPV type of interest and (2) subjects who are seronegative at Day 1 and PCR-negative from Day 1 through Month 7 for all vaccine types. RCDF (Reverse Cumulative Distribution Function) plots will also be used to observationally compare these groups. These summaries will evaluate the anamnestic responses to the individual components of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine. Table 2 outlines the groups that will be compared for this analysis.

Table 2

Comparisons to Assess a Potential Association Between Baseline Seropositivity and Improved Immunogenicity at Month 7

Comparison Group 1	Comparison Group 2
HPV type 6 sero+ at Day 1[†]	HPV type 6 naïve[‡]
HPV type 6 sero+ at Day 1[†]	Naïve for all vaccine HPV types[‡]
HPV type 11 sero+ at Day 1[†]	HPV type 11 naïve[‡]
HPV type 11 sero+ at Day 1[†]	Naïve for all vaccine HPV types[‡]
HPV type 16 sero+ at Day 1[†]	HPV type 16 naïve[‡]
HPV type 16 sero+ at Day 1[†]	Naïve for all vaccine HPV types[‡]
HPV type 18 sero+ at Day 1[†]	HPV type 18 naïve[‡]
HPV type 18 sero+ at Day 1[†]	Naïve for all vaccine HPV types[‡]

Comparisons will be made between Group 1 and Group 2.
Sero+ = Seropositive.
[†] Regardless of PCR status.
[‡] Naïve = Seronegative at Day 1 and PCR negative from Day 1 through Month 7.

C. Approaches to Immunogenicity Analysis (Cont.)

To investigate whether or not the baseline titer of a subject who is seropositive at Day 1 for a given vaccine HPV type has an impact on the subject's response to the vaccine at Month 7, the natural log of the Month 7 titers for a given HPV type will be modeled as a function of treatment group and the natural log of the baseline HPV titer for that same HPV type for subjects who are seropositive for that HPV type at baseline. Three regression models will be constructed. The first will be a linear regression model with the natural log titers at Month 7 as the response variable. The second will be a logistic regression model with an indicator of whether or not the subject's Month 7 titer is ≥200 mMU/mL as the response variable. The third will be a linear regression model with the difference in natural log titer between Month 7 and Day 1 as the response variable (this model is equivalent to modeling the fold-rise in titer from Day 1 to Month 7).

The study will also explore whether subjects who are HPV infected but who do not have detectable anti-HPV at baseline have an active immune response to the infection. In particular, peak (Month 7) and persistence anti-HPV responses for a given vaccine HPV type will be compared between subjects who are infected at Day 1 (PCR positive) with that HPV type, but who have not developed a detectable anti-HPV response (seronegative) to that type, and those who are naïve for that HPV type. It is assumed that newly infected subjects are PCR positive but still have negative RIA titers. Therefore, the antibody responses in subjects who are initially PCR positive but seronegative for a given vaccine HPV type will be summarized and compared observationally with subjects who are HPV-naïve for that HPV type and with subjects who are HPV-naïve for all vaccine HPV types (naïve implies seronegative at Day 1 and PCR-negative from Day 1 through Month 7). The GMTs and the proportions of subjects with anti-HPV ≥200 mMU/mL at Month 7 and through the persistence phase will be summarized. Table 3 outlines the groups that will be compared for this analysis.

C. Approaches to Immunogenicity Analysis (Cont.)

Table 3

Comparisons to Explore Implied Immunity
in Subjects With New HPV Infections

Comparison Group 1	Comparison Group 2
HPV type 6 PCR+ and sero- at Day 1	HPV type 6 naïve†
HPV type 6 PCR+ and sero- at Day 1	Naïve for all vaccine HPV types†
HPV type 11 PCR+ and sero- at Day 1	HPV type 11 naïve†
HPV type 11 PCR+ and sero- at Day 1	Naïve for all vaccine HPV types†
HPV type 16 PCR+ and sero- at Day 1	HPV type 16 naïve†
HPV type 16 PCR+ and sero- at Day 1	Naïve for all vaccine HPV types†
HPV type 18 PCR+ and sero- at Day 1	HPV type 18 naïve†
HPV type 18 PCR+ and sero- at Day 1	Naïve for all vaccine HPV types†
Comparisons will be made between Group 1 and Group 2. PCR+ = PCR positive. Sero- = Seronegative. † Naïve = Seronegative at Day 1 and PCR negative from Day 1 through Month 7.	

To assess the difference between subjects who may have natural immunity (subjects who are initially seropositive but PCR negative) and subjects with persisting immune responses after natural infection, GMTs will be summarized and compared at Month 7 between subjects who are initially seropositive and PCR positive for a given vaccine HPV type and subjects who are initially seropositive but PCR negative for that type. These summaries will allow an observational assessment of the differences in RIA titers between the 2 groups. Table 4 lists the groups that will be compared for this analysis.

Table 4

Comparisons to Assess the Difference Between Subjects With
Probable Natural Immunity and Subjects With Persisting Immune
Responses After Natural Infection

Comparison Group 1	Comparison Group 2
HPV type 6 sero+ and PCR+ at Day 1†	HPV type 6 sero+ but PCR- at Day 1†
HPV type 11 sero+ and PCR+ at Day 1†	HPV type 11 sero+ but PCR- at Day 1†
HPV type 16 sero+ and PCR+ at Day 1†	HPV type 16 sero+ but PCR- at Day 1†
HPV type 18 sero+ and PCR+ at Day 1†	HPV type 18 sero+ but PCR- at Day 1†
Comparisons will be made between Group 1 and Group 2. Sero+ = Seropositive. PCR+ = PCR positive. PCR- = PCR negative. † HPV seronegative at Day 1 and PCR negative from Day 1 through Month 7 for all other vaccine HPV types.	

C. Approaches to Immunogenicity Analysis (Cont.)

b. Association Among Vaccine HPV Types

The exploratory analysis of the association among vaccine HPV types has 2 goals: (1) to ascertain if seropositivity for one HPV type impacts the immune response to another vaccine HPV type and (2) to assess whether or not the magnitude of the association appears to increase by the number of other vaccine HPV types for which the subject is positive.

To assess whether or not baseline seropositivity for 1 vaccine HPV type impacts the immune response to another vaccine HPV type, the GMTs and the proportions of subjects with anti-HPV ≥200 mMU/mL at Month 7 for a given vaccine HPV type will be observationally compared between subjects who are HPV naïve for all vaccine HPV types from Day 1 to Month 7 and subjects who are HPV naïve for the given vaccine HPV type but seropositive for each 1 of the other 3 vaccine HPV types individually. For example, the HPV 6 immune responses will be summarized and compared between subjects who are HPV naïve for all vaccine HPV types from Day 1 to Month 7 and: (1) subjects who are naïve for HPV type 6 from Day 1 to Month 7 but seropositive to HPV type 11 at Day 1; (2) subjects who are naïve for HPV type 6 from Day 1 to Month 7 but seropositive to HPV type 16 at Day 1; and (3) subjects who are naïve for HPV type 6 from Day 1 to Month 7 but seropositive to HPV type 18 at Day 1. RCDF plots will also be used to compare these groups. Table 5 outlines the groups that will be compared for this analysis.

C. Approaches to Immunogenicity Analysis (Cont.)

Table 5

Comparisons to Assess the Impact of Baseline Seropositivity for One Vaccine HPV Type on the Immune Responses to Another Vaccine HPV Type

HPV Type to be Summarized	Comparison Group 1	Comparison Group 2
6	Baseline sero+ for HPV type 11[†] Baseline sero+ for HPV type 16[†] Baseline sero+ for HPV type 18[†]	Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡]
11	Baseline sero+ for HPV type 6[†] Baseline sero+ for HPV type 16[†] Baseline sero+ for HPV type 18[†]	Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡]
16	Baseline sero+ for HPV type 6[†] Baseline sero+ for HPV type 11[†] Baseline sero+ for HPV type 18[†]	Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡]
18	Baseline sero+ for HPV type 6[†] Baseline sero+ for HPV type 11[†] Baseline sero+ for HPV type 16[†]	Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡] Vaccine HPV type naïve[‡]

Comparisons will be made between Group 1 and Group 2.
Sero+ = Seropositive.
[†] HPV seronegative at Day 1 and PCR negative from Day 1 through Month 7 for all other vaccine HPV types.
[‡] Seronegative at Day 1 and PCR negative from Day 1 through Month 7 for all vaccine HPV types.

Finally, to explore the impact of multiple positive vaccine HPV types on immune responses to the other vaccine HPV types, the proportions of subjects with anti-HPV ≥200 mMU/mL and the GMTs at Month 7 for a given vaccine HPV type will be summarized and compared between subjects who are HPV naïve for all vaccine HPV types from Day 1 to Month 7 and: (1) subjects who are seronegative at Day 1 for the HPV type being summarized but who are seropositive at Day 1 for all possible combinations of 2 of the other 3 vaccine HPV types and (2) subjects who are seronegative at Day 1 for the HPV type being summarized but who are seropositive at Day 1 for all 3 of the other vaccine HPV types. Table 6 provides all summaries that will be performed for this purpose.

Data permitting, the question of associations among the vaccine HPV types will be explored with regression models. The first set of models will explore the individual associations between HPV vaccine types. The first model will model the Month 7 responses for 1 vaccine HPV

C. Approaches to Immunogenicity Analysis (Cont.)

type (e.g., HPV type 6) as a function of the baseline titers for the other 3 vaccine HPV types (e.g., HPV type 11, 16, and 18). The second model will model the Month 7 responses for 1 vaccine HPV type (e.g., HPV type 6) as a function of the Month 7 responses for the other 3 vaccine HPV types (HPV types 11, 16, or 18). These models will be constructed separately for subjects who are: (1) seronegative for the HPV type of interest and (2) seropositive for the HPV type of interest at Day 1.

An additional model will assess the impact of the number of vaccine HPV types for which the subject is positive at baseline on the response to a different given HPV type. This model will model the Month 7 responses for one vaccine HPV type (e.g., HPV type 6) as a function of the number of other vaccine HPV types that are seropositive at Day 1 (0, 1, 2, or 3 seropositive titers among the set of HPV types 11, 16, and 18).

Linear regression models will be used with the natural log of the Month 7 titers as the response variable. The exploratory immunogenicity analyses will be performed on the per-protocol population. Due to possible sample size limitations, all comparisons between groups will be observational, and some assessments may not be viable due to small cell sizes.

C. Approaches to Immunogenicity Analysis (Cont.)

Table 6

Comparisons to Explore the Impact of Multiple Baseline Positive Vaccine HPV Types on the Immune Responses to the Other Vaccine Types

HPV Type to be Summarized	Comparison Group 1	Comparison Group 2
6	Sero- for HPV 6 but sero+ for 2 of vaccine HPV types 11, 16, and/or 18	Vaccine HPV type naïve†
	Sero- for HPV 6 but sero+ for HPV types 11, 16, and 18	Vaccine HPV type naïve†
11	Sero- for HPV 11 but sero+ for 2 of vaccine HPV types 6, 16, and/or 18	Vaccine HPV type naïve†
	Sero- for HPV 11 but sero+ for HPV types 6, 16, and 18	Vaccine HPV type naïve†
16	Sero- for HPV 16 but sero+ for 2 of vaccine HPV types 6, 11, and/or 18	Vaccine HPV type naïve†
	Sero- for HPV 16 but sero+ for HPV types 6, 11, and 18	Vaccine HPV type naïve†
18	Sero- for HPV 18 but sero+ for 2 of vaccine HPV types 6, 11, and/or 16	Vaccine HPV type naïve†
	Sero- for HPV 18 but sero+ for HPV types 6, 11, and 16	Vaccine HPV type naïve†
Comparisons will be made between Group 1 and Group 2. Sero- = Seronegative. Sero+ = Seropositive. † Seronegative at Day 1 and PCR negative from Day 1 through Month 7 for all vaccine HPV types.		

D. Definition of Compliance Measure

In the consistency lot substudy, the numbers and percentages of subjects who complete each vaccination visit and who complete the follow-up blood sampling at Month 7 will be summarized by treatment group (consistency lot) and in total.

V. SAFETY ANALYSES

All subjects who received at least 1 injection and have follow-up data will be included in the summary of serious adverse experiences. All subjects in the NSAE substudy and in the United Kingdom who received at least 1 injection and have follow-up data will be included in the safety summaries of nonserious

V.SAFETY ANALYSES (CONT.)

adverse experiences. In all summaries, subjects who are incorrectly randomized will be summarized for safety according to the clinical material they received. If a subject received vaccine at any of the visits, then the subject will be summarized in the appropriate vaccine group. Subjects will be summarized in the placebo group only if they received placebo at all visits.

For the nonserious adverse experience summaries, the data from the NSAE substudy and the data from the subjects in the United Kingdom will be summarized separately and combined. Separate summaries will be provided because the differences in the nonserious adverse experience reporting mechanisms in the 2 sets of subjects (vaccination report card in the NSAE substudy versus solicitation at the next visit in the United Kingdom) may lead to differences in the adverse experience rates. In particular, without the written reminders from the vaccination report cards, the subjects in the United Kingdom may be susceptible to recalling only the more significant adverse experiences when they are asked for adverse experience data at the next visit (which will likely be 2 to 4 months following the previous vaccination).

Subjects enrolled at study sites outside of the United States and the United Kingdom may report nonserious adverse experiences spontaneously even though solicitation of such events is not a part of the adverse experience follow-up at these study sites. Any nonserious adverse experiences reported by subjects outside of the United States and the United Kingdom will be listed. Incidence rates among these subjects will not be estimated due to the difficulties in associating a correct denominator with data of this type (spontaneously reported).

The summaries of serious adverse experiences will be provided both for the vaccine groups pooled versus placebo and for the 3 consistency lots separately. However, since only ~650 subjects are expected to be enrolled into both the consistency lot substudy and NSAE substudy, the primary focus of the nonserious adverse experience safety summaries will be on comparisons between the placebo recipients and the pooled vaccine recipients. Nonserious adverse experience data will be tabulated by consistency lot to allow an observational assessment of the comparability of the safety profiles across lots.

A. Special Safety Analysis

The primary endpoint for safety is the incidence of vaccine-related serious adverse experiences. Point estimates of the incidence of serious adverse experiences and vaccine-related serious adverse experiences and the corresponding 95% confidence intervals will be provided for each treatment group for all subjects enrolled in the study.

A. Special Safety Analysis (Cont.)

Severe injection site reactions are also of special interest. Therefore, for all subjects enrolled in the NSAE substudy and in the United Kingdom, point estimates of the incidences of severe injection site reactions will be provided for each treatment group along with the corresponding 95% confidence intervals.

For all subjects enrolled, pregnancies that occur during the study and their outcomes will be tabulated by treatment group and overall. Pregnancies that were detected anytime between Day 1 and 1 month following the completion of the vaccination regimen (Month 7) will be summarized separately from those detected following Month 7. In addition, the outcomes of all breast-feeding experiences in subjects who were breast-feeding during the vaccination regimen will be tabulated by treatment group and overall.

B. Adverse Experiences

To reduce the number of statistical tests performed for adverse experiences and the accompanying multiplicity issues associated with them, the following approach will be used for the analyses of adverse experiences. These analyses will be performed using the subset of subjects participating in the NSAE substudy only.

To provide an overall assessment of safety, measures such as the incidence of: (1) any adverse experience, (2) any injection-site adverse experience, (3) any systemic adverse experience, and (4) any vaccine-related adverse experience that occurred throughout the study will be summarized.

To address specific adverse experiences, the incidences of injection-site adverse experiences Days 1 to 5 and specific systemic adverse experiences within 14 days postvaccination occurring in at least 1% of the subjects will be tabulated across injections. Risk differences will be estimated, and the associated 95% two-sided confidence intervals will be provided. The method of Miettinen and Nurminen [1] will be used for all comparisons.

Statistical testing will be conducted for the specific adverse experiences for which subjects were prompted on the vaccination report card. Therefore, the proportions of subjects who experience redness, swelling, and pain/tenderness at the injection site and the proportions of subjects with elevated temperatures (≥100°F, oral equivalent) will be formally analyzed. In addition to the count and the percentage, the risk difference, the associated 95% confidence interval, and the two-sided p-value testing for a difference in incidence rates will be presented.

B. Adverse Experiences (Cont.)

Separate summaries will be performed for subjects who are: (1) seronegative at Day 1 and PCR-negative Day 1 through Month 7 to all 4 vaccine components, and (2) seropositive at Day 1 or PCR-positive at any time between Day 1 and Month 7 to at least 1 vaccine component. Tabulations of adverse experience data by injection will also be provided.

The distribution of adverse experiences across severity/intensity ratings will also be provided for each treatment group across all injections. Each episode of each reported adverse experience will be counted separately for this summary. The summary will be provided for any adverse experiences, any systemic adverse experiences, and any injection-site adverse experiences. For the measured adverse experiences of redness and swelling, 0 to 1 inch will be categorized as mild, >1 inch but ≤2 inches will be categorized as moderate and >2 inches will be categorized as severe. In addition, the distribution of injection-site adverse experiences will be summarized separately by adverse experience for the "prompted" adverse experiences of pain/tenderness/ soreness, redness, and swelling.

To complement the analyses above, the number of subjects reporting an adverse experience within each severity/intensity category will be summarized across all adverse experiences reported. Each subject will be counted only once in these summaries. The worst severity/intensity rating will be utilized for each subject. These summaries will be performed for any adverse experiences, any injection-site adverse experiences, and any systemic adverse experiences. The summaries will be provided for each treatment group across all injections.

For specific injection-site adverse experiences Day 1 to 5 postvaccination and specific systemic adverse experiences Day 1 to 14 postvaccination with an incidence of ≥ 1% in either treatment group, the distribution of maximum severity per subject will also be tabulated by treatment group in Section II.11 of the clinical study report (CSR). For the measured injection-site adverse experiences that are prompted for on the VRC Days 1 to 5 postvaccination, erythema (redness) and swelling, the distribution of maximum reported size will also be tabulated by treatment group.

VI. INTERIM ANALYSES/DATA AND SAFETY MONITORING BOARDS

A. Efficacy

An interim efficacy analysis of Protocol 015 is planned to be conducted at the time that at least 19 cases of the primary endpoint have been observed. The interim analysis will involve a test of the primary hypothesis in the per-protocol and the first 2 MITT populations. An estimate of the incidence rate of all CIN 2/3 will also be provided. The critical database fields for identifying protocol violators, identifying cases, and conducting the primary analysis will be identified, and all critical data that are in-house will be screened prior to the interim analysis. A separate document outlining the critical data fields and detailed screening plan will be written.

The interim analysis of Protocol 015 will be performed in conjunction with an interim analysis of the combined data from Protocols 005, 007, 013, and 015, and, therefore, will only be conducted when at least 19 cases of CIN 2/3 or cervical cancer related to HPV 16 or 18 have been observed in Protocol 015 *and* at least 33 cases have been observed across all 4 studies.

The interim analysis of Protocol 015 will be performed by a designated unblinded statistician unrelated to the study. The unblinded statistician will provide the results of the analysis to a data and safety monitoring board (DSMB) along with the results of the interim analysis of the combined data set (Protocols 005, 007, 013, and 015). If the interim analysis of Protocol 015 meets the primary statistical criterion for success (lower bound of the confidence interval for the vaccine efficacy >0%) *and* the interim analysis of the combined data sets meets the primary criterion for success prespecified in the DAP for that analysis (lower bound of the confidence interval for the vaccine efficacy >25%) in the per-protocol and the 2 MITT populations, then the DSMB will communicate this information to the HPV vaccine project team at the SPONSOR, and the project team may proceed with submission of the quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine for regulatory review. At that time, any remaining unscreened data in the official clinical database (the non-critical data) for Protocol 015 that were collected prior to the interim analysis will be screened and cleaned by a designated clinical, statistical, and data management team. The database will be audited and a copy will be unblinded and frozen. The designated statistical team will analyze the trial data and prepare the submission for regulatory review. The results will then be disclosed to the regulatory agency as part of the submission.

A. Efficacy (Cont.)

If either the interim analysis of Protocol 015 or the interim analysis of the combined data set fails to meet the statistical criterion for success at the interim time point in either the per-protocol or 1 of the 2 MITT populations, the DSMB will communicate this information to the HPV vaccine project team at the SPONSOR, and the project team will not proceed with a submission to regulatory authorities at that time. In this case, it may be necessary for the interim analysis results to be provided to a senior management committee by the designated unblinded statistician to enable decisions to be made regarding the overall HPV program. The senior management committee would consist of 4 to 5 members representing Research, Clinical, Regulatory, and Biostatistics. This committee would not make decisions that would impact the conduct of the individual studies.

The final analysis of Protocol 015 will be performed when at least 29 cases of the primary endpoint have been observed regardless of the outcome of the interim analysis. This will allow for an assessment of longer-term vaccine efficacy and will allow a sufficient number of cases of CIN 2/3 or worse due to any cause to accrue for a secondary analysis of this endpoint using the combined data sets from Protocols 005, 007, 013, and 015 (this analysis is outlined in the combined analysis DAP). A multiplicity adjustment will be made to account for the 2 separate analyses of Protocol 015. A two-sided alpha of 0.0204 will be spent at the interim analysis, and a two-sided alpha of 0.0411 will be spent at the final analysis.

B. Safety

In order to ensure that no alarming, unusual, or unexpected safety problems are occurring with the vaccine, safety will be monitored during the study by the DSMB who will determine if any actions should be taken based on the data. Periodically during the study (approximately every 6 months during the vaccination period and approximately every year thereafter), all available safety data from the study will be summarized by the designated unblinded Merck statistician and sent to the DSMB. Summaries will also be provided in the event that there is a specific safety concern during the study.

In general, adverse experiences will be summarized descriptively as frequencies and percentages by treatment group and type of adverse experience (this will be done by vaccination visit and across all vaccination visits), and temperatures will be summarized. The DSMB manual specifies the details of the operating procedure for this process.

In particular, the DSMB will monitor the following parameters during the course of Protocol 015:

B. Safety (Cont.)

- the incidence and characteristics of serious adverse experiences;
- the incidence and characteristics of nonserious adverse experiences – a general summary will be provided as well as summaries of specific injection-site and systemic adverse experiences;
- pregnancies and their outcomes;
- the incidence of breast-feeding during the vaccination period and the associated outcomes; and
- new medical conditions that arise during the study.

Serious adverse experiences will be reported to the DSMB at the same time Merck reports them to the FDA.

All summaries will be provided by treatment group and in total. However, in the by-treatment-group summaries, the treatment groups will be labeled as Group A and Group B unless the DSMB has a specific concern and requests that the groups be unblinded.

The interim summaries of safety will be performed on partially unscreened and unaudited data to ensure that the DSMB has the quickest access to the largest possible data set in any given summary. It is recognized that there may be minor changes in incidences and severity gradings of adverse experiences based on the screening and auditing process. However, since DSMB decisions with regards to safety are likely to be based on large, unexpected differences in safety parameters between treatment groups, such inaccuracies are not likely to result in incorrect decisions.

VII. STATISTICAL TECHNICAL ISSUES

A. Planned Statistical Power and Sample Size

1. Efficacy (Overall Study)

In order to avoid problems with imprecise incidence and efficacy estimates, the efficacy study is powered based on a fixed event design with an interim analysis. To ensure adequate power for the interim analysis (80 to 90% power) and the final analysis (≥90%) for varying true vaccine efficacies after the multiplicity adjustment, at least 19 cases of the primary endpoint are required for the interim analysis, and at least

A. Planned Statistical Power and Sample Size (Cont.)

29 cases are required for the final analysis. Table 7 gives the power for the primary efficacy analysis (based on the methodology described in Section VII.D.1.a.) for varying true vaccine efficacy values with 19 cases at the interim analysis and 29 cases at the final analysis.

Table 7

Power Sensitivity Analysis

Analysis	α-Level	Number of Cases	True Vaccine Efficacy	Power
Interim	0.0204	19	80%	80%
			85%	90%
			90%	97%
Final	0.0411	29	80%	95%
			85%	99%
			90%	>99%

An overall sample size of 11,500 subjects should yield at least 29 cases of the primary endpoint by the time all subjects have completed Month 48 of the study based on the following assumptions:

a. ~18% of the subjects enrolled will be HPV 16 seropositive at Day 1 or PCR positive at Day 1 or Month 7;

b. ~18% of the subjects enrolled will be HPV 18 seropositive at Day 1 or PCR positive at Day 1 or Month 7;

c. the attrition rate will be no more than 15% through Month 7 of the study and no more than 5% per year thereafter;

d. all subjects followed beyond Month 7 are eligible to be endpoints and any subject who discontinues the study is at risk for HPV 16- or 18-related CIN 2/3 or worse for half of the study;

e. the incidence of CIN 2/3 or worse related to HPV Type 16 will be ~0.19% in the placebo group over the 41 months after Postdose 3, while the incidence of CIN 2/3 or worse related to HPV Type 18 will be ~0.038% in the placebo group over the 41 months after Postdose 3; and

f. if the vaccine efficacy is extremely high, all cases will have to accrue in the placebo group before the efficacy analysis may be performed.

A. Planned Statistical Power and Sample Size (Cont.)

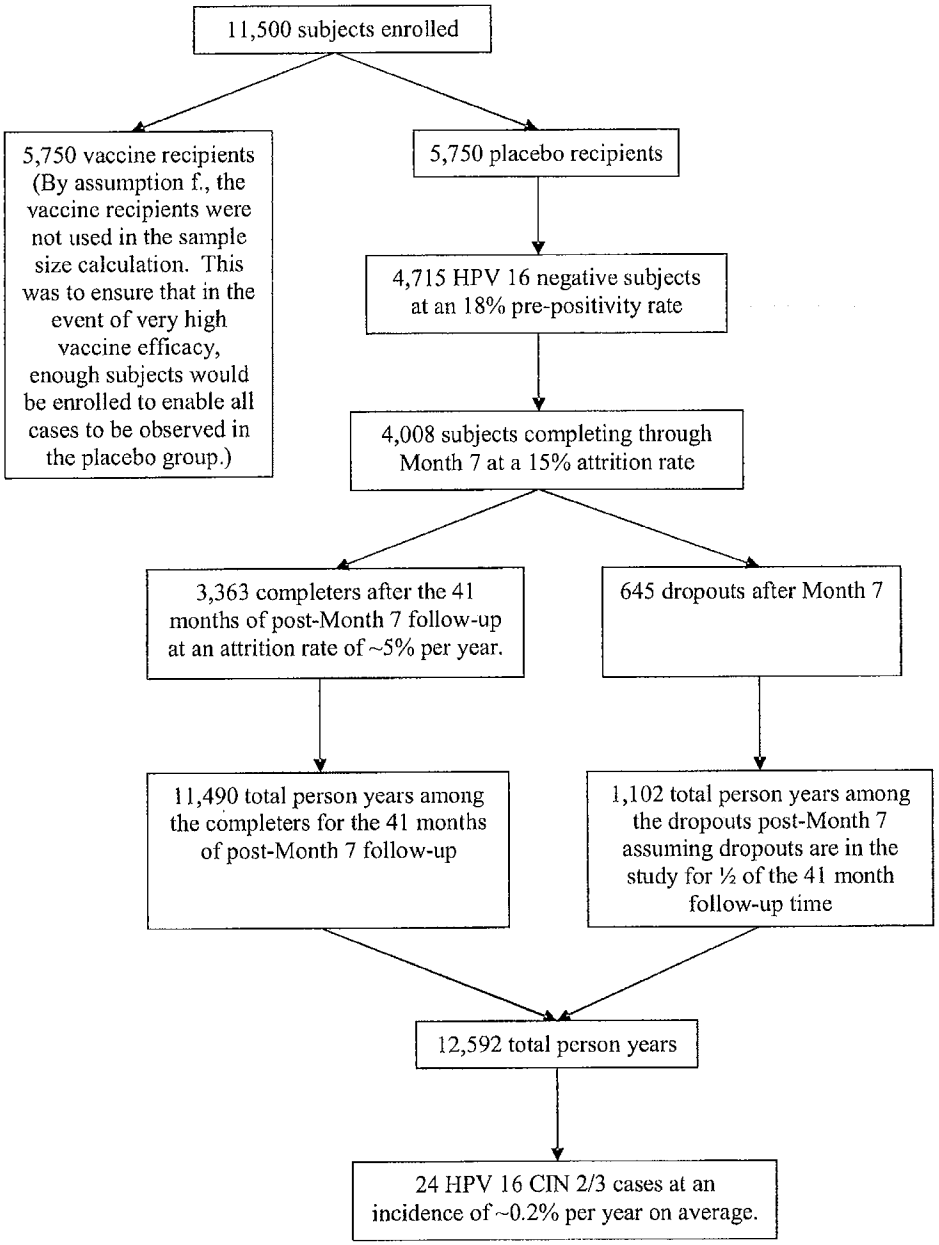
Assuming that all subjects followed beyond Month 7 are eligible to be considered as endpoints, a sample size of 11,500 will provide ~8000 subjects who are eligible to be considered as cases according to the HPV 16-related disease definition and ~8000 subjects who are eligible to be cases according to the HPV 18-related disease definition. Since the primary endpoint is a composite endpoint, any subject who is eligible to be considered as an endpoint according to the HPV 16-related disease definition, the HPV 18-related disease definition or both will be included in the population at risk for the primary analysis. Therefore, assuming an ~8% overlap in the number of subjects who are HPV 16 or 18 seropositive at Day 1 or PCR positive Day 1 through Month 7, the population at risk for the primary analysis will include ~9000 subjects. Using a person time approach under assumptions (3) and (4) and given the event rates in assumptions (5) and (6), ~9000 subjects in the population at risk will provide the required number of endpoints by Year 4 of the study.

Figures 1 and 2 illustrate the derivation of the HPV 16-related CIN 2/3 cases and HPV 18-related CIN 2/3 cases, respectively, from the 11,500 subjects enrolled.

A. Planned Statistical Power and Sample Size (Cont.)

Figure 1

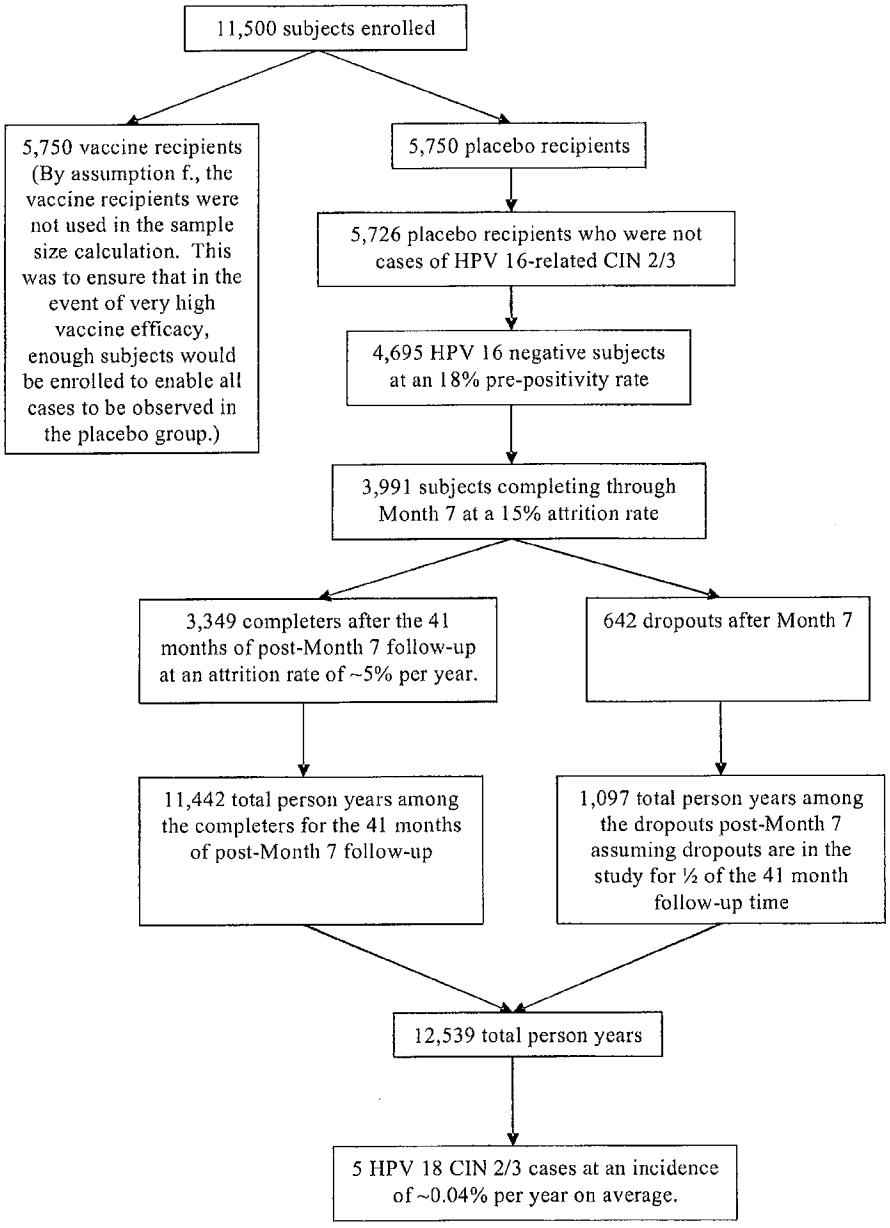
Derivation of HPV 16-Related CIN 2/3 Cases



A. Planned Statistical Power and Sample Size (Cont.)

Figure 2

Derivation of HPV 18-Related CIN 2/3 Cases



A. **Planned Statistical Power and Sample Size (Cont.)**

2. **Immunogenicity (Consistency Lot Substudy)**

The primary time point for the immunogenicity analysis is Month 7. It is expected that the attrition rate through Month 7 of the study will be no more than 15%. In addition, it is assumed that 15%, 18%, and 18% of the subjects enrolled in the consistency lot substudy will be seropositive at Day 1 or PCR positive at Day 1 or Month 7 for HPV Types 6/11, 16, and 18, respectively. Therefore, it is expected that 361 (500 x 0.85 x 0.85) of the 500 subjects randomized to receive each of the consistency lots will be evaluable for the HPV 6 and 11 endpoints and 348 (500 x 0.82 x 0.85) of the 500 subjects will be evaluable for the HPV 16 and 18 endpoints.

For the first primary immunogenicity hypothesis involving the consistency of the 3 lots with respect to the percentage of subjects who achieve anti-HPV 6 ≥200 mMU/mL, anti-HPV 11 ≥200 mMU/mL, anti-HPV 16 ≥200 mMU/mL, and anti-HPV 18 ≥200 mMU/mL at Week 4 Postdose 3, the assumed response rate for each HPV type in each lot is 90%. (The assumed response rate is based on Protocols 001, 002, 004, and 006). With 361 evaluable subjects for the HPV 6 and 11 endpoints and 348 evaluable subjects for the HPV 16 and 18 endpoints, this study has 98.3% power to rule out a ≥10-percentage-point difference in rates among the 3 lots ($\alpha = 0.05$) for each HPV type. Assuming independence of the 4 HPV types, the overall power for the first primary immunogenicity hypothesis is 93.4%.

For the second primary immunogenicity hypothesis involving the consistency of the 3 lots of quadrivalent HPV vaccine with respect to GMTs for HPV 6, 11, 16, and 18, the standard deviations of the natural logarithm of the Month 7 titers are assumed to be no more than 1.2 for each HPV type. (The assumed standard deviation is based on Protocols 001, 002, and 004.) With 361 evaluable subjects for the HPV 6 and 11 endpoints and 348 evaluable subjects for the HPV 16 and 18 endpoints, this study has 99.9% power to rule out a ≥2-fold difference in the ratio of GMTs ($\alpha = 0.05$) for each HPV type. Assuming independence of the 4 HPV types, the overall power for the second primary immunogenicity hypothesis is 99.6%.

If the 2 primary immunogenicity hypotheses are assumed to be independent, the overall power for the consistency lot substudy is 93%.

B. Method of Assigning Study Participants to Treatment Groups

All subjects in the study will be randomized in a 1:1 ratio to receive either quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine or placebo within investigative sites. Within the quadrivalent HPV vaccine group, subjects will be further randomized to receive 3 different lots of the vaccine in a 1:1:1 ratio. The actual lots received will differ between subjects enrolled in the consistency lot substudy and the remaining subjects in Protocol 015.

Subjects participating in the consistency lot substudy will be randomized in a 1:1:1:3 ratio to receive Consistency Lot 1 of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 2 of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, Consistency Lot 3 of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine, or placebo within investigative sites. This randomization will ensure a 1:1:1 allocation of subjects to the 3 quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine consistency lots. It will also maintain a 1:1 ratio of subjects receiving quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine (any lot) and placebo for the overall efficacy study.

For the non-consistency lot part of the study, the Interactive Voice Response System (IVRS) will assign a block of 18 allocation numbers to a study site upon allocation of the first subject at that site. Within the block of 18 numbers, 9 will be allocated to vaccine (3 to Lot No. VAI018I001, 3 to Lot No. VAI025T001, and 3 to Lot No. VAI025T002) and 9 to placebo in a 3:3 ratio in blocks of 6. The system will assign an additional block of allocation numbers to a study site once it has depleted its supply of numbers.

For the consistency lot substudy, the IVRS will assign a block of 18 allocation numbers to a study site upon allocation of the first subject at that site. Within the block of 18 numbers, 3 will be allocated to Consistency Lot 1, 3 to Consistency Lot 2, 3 to Consistency Lot 3, and 9 to placebo in a 1:1:1:3 ratio in blocks of 6. The system will assign an additional block of allocation numbers to a study site once it has depleted its supply of numbers.

For both parts of the study, case numbers will be assigned to the subjects sequentially at each study site beginning with the lowest number available at the site as subjects are enrolled in the study. This will retain the randomization and the treatment group ratio within study sites. Case numbers should not be reassigned for any reason.

C. Blinding/Unblinding

This is a double-blinded study (operating under in-house blinding procedures). Therefore, the subjects enrolled in the study, the investigators and study

C. Blinding/Unblinding (Cont.)

personnel, the laboratory personnel conducting the PCR and RIA assays on the clinical samples, and the Pathology Panel will be blinded to the treatment group assignments of the subjects for the duration of the study (~4 years). The clinical, data management, and statistical personnel involved with the protocol at the SPONSOR will be blinded to the treatment group assignments of the subjects through the primary efficacy evaluations (~2.5 to 3 years).

To ensure blinding of the subjects, investigators and study personnel, the vaccine and placebo are supplied in identical vials labeled with a double-panel, blinded label. The vaccine and placebo are visually indistinguishable.

The duration of the primary efficacy phase of Protocol 015 will be ~4 years. However, the primary time point for the immunogenicity analysis for the consistency lot substudy is Month 7. The immunogenicity analysis for the consistency lot substudy and the safety analysis for the NSAE substudy may be conducted when all subjects in the respective substudies have completed the Month 7 follow-up. In order to keep the data for the efficacy evaluations blinded while preparing the study reports for the substudies, restricted clinical, statistical, and data management teams will be formed to analyze the data from the immunogenicity and safety substudies. All clinical, statistical, and data management personnel can perform the screening and cleaning of data through Month 7, a process that includes data review and correction and identification of protocol violators, in a blinded fashion. Once the data through Month 7 are cleaned and audited, a copy of the clean database will be made, and this copy will be unblinded for the restricted clinical, statistical, and data management teams that are responsible for the immunogenicity and safety analyses. Only the teams responsible for the immunogenicity and safety analyses will have access to the allocation schedules and the unblinded database. The primary database will remain blinded for the efficacy analysis. No member of the previously unblinded team (who participated in the immunogenicity and safety analyses) will be responsible for screening and cleaning the data for the efficacy analysis. Therefore, the data for the efficacy analysis will be screened and cleaned in a blinded manner. The efficacy analysis will then be performed on the clean data set.

Protocol 015 will also have an interim analysis performed when at least 19 cases of the primary endpoint have been observed. The interim analysis will be performed on screened and cleaned data by a designated unblinded statistician. If the interim analysis of Protocol 015 meets the primary statistical criterion for success *and* the interim analysis of the combined data sets from Protocols 005, 007, 013, and 015 meets the primary criterion for success prespecified in the DAP for that analysis, the quadrivalent HPV

C. Blinding/Unblinding (Cont.)

(Types 6, 11, 16, and 18) L1 VLP vaccine may be submitted to regulatory authorities for review as soon as possible. The interim results from both the analysis of Protocol 015 and the analysis of the combined studies will then be disclosed to all personnel at the SPONSOR responsible for preparing the regulatory submission. Following the unblinding of the data for the preparation of the submission, the clinical, statistical, and data management personnel assigned to the HPV vaccine project at the SPONSOR will most likely have access to the individual treatment assignments of the subjects in Protocol 015, even though the study will continue. However, Protocol 015 will be complete for the purposes of addressing the primary objectives.

If the early submission occurs, the remainder of Protocol 015 will be considered an extension phase. The purpose of the extension phase will be to collect data on the longer-term efficacy of the vaccine with respect to the primary endpoint and to allow a sufficient number of cases of CIN 2/3 and cervical cancer due to any cause to accrue for a secondary analysis of this endpoint using the combined data sets from Protocols 005, 007, 013, and 015.

It is important to note that throughout the efficacy phase *and* the extension phase of Protocol 015, the laboratory personnel, the pathology panel, the investigators, study site personnel, and subjects will remain blinded to the individual treatment allocations of the subjects. The data that impact ascertainment of the efficacy endpoints are: (1) the Pap tests, which are read by the central laboratory; (2) the biopsy samples, which are read by the pathology panel; and (3) the PCR test results, which are provided by MRL's blinded laboratory. Thus, all investigators and technicians who have the entire responsibility for the ascertainment and confirmation of efficacy endpoints will be blinded for the duration of the study.

If either the interim analysis of Protocol 015 or the interim analysis of the combined data sets fails to meet the statistical criterion for success at the interim time point, the quadrivalent HPV (Types 6, 11, 16, and 18) L1 VLP vaccine will not be submitted for regulatory review at that time. The results from the interim analyses will remain blinded to everyone except the designated unblinded statistician, the DSMB, and possibly a small senior management committee until the end of the studies (i.e., until the complete data have been screened, cleaned, and audited for analysis). It may be necessary for the interim analysis results to be provided to a senior management committee by the designated unblinded statistician to enable decisions to be made regarding the overall HPV program. The senior management committee would consist of 4 to 5 members representing Research, Clinical, Regulatory, and Biostatistics.

D. Statistical Methods

1. Methods for Efficacy Analyses

a. Vaccine Efficacy

A one-sided test of H_0: $\pi = 0$ versus H_1: $\pi > 0$ will be conducted, where $\pi = 1 - RR$ is the vaccine efficacy, and RR is the relative risk of the vaccine compared with placebo. In addition, a multiplicity-adjusted two-sided exact confidence interval for the vaccine efficacy will be provided. An exact conditional procedure will be used under the assumption that the number of cases in the vaccine group in each study, C_v, and the number of cases in the placebo group, C_p, are independent Poisson random variables with means λ_v and λ_p. (The relative risk of the vaccine is λ_v/λ_p.) Assuming the follow-up times in the vaccine group and placebo group are k_v and k_p, respectively, the number of cases in the vaccine group C_v, given the total number of cases observed, $T = C_v + C_p$, is binomially distributed with parameters, $C_v + C_p$, and $p = k_v\lambda_v/(k_v\lambda_v + k_p\lambda_p)$, where p is the probability that a subject in the vaccine group is a case.

In the case of equal follow-up in the vaccine and placebo groups, the point estimate for p is $C_v/(C_v + C_p)$. The lower bound of the $100(1-\alpha)\%$ exact confidence interval for p can be calculated by searching for the p_L such that the probability of observing C_v or more vaccine cases out of C_v+C_p total cases is $\alpha/2$. The upper bound of the exact confidence interval for p can be calculated by searching for p_U such that the probability of observing C_v or fewer vaccine cases out of C_v+C_p total cases is $\alpha/2$. It follows that the exact confidence interval for vaccine efficacy will then have lower bound $VE_L = (1-2*p_U)/(1-p_U)$ and upper bound $VE_U = (1-2*p_L)/(1-p_L)$.

In the event that there is unequal follow-up between groups, the estimate of the vaccine efficacy will rely not solely on the ratio of the number of cases in the vaccine group to the number of cases in the placebo group, but must also account for the follow-up in each group. Rather than $VE = 1 - C_v/C_p$, the point estimate for the vaccine efficacy will be $VE = 1 - (C_v/k_v)/(C_p/k_p)$, where k_v and k_p are the person years of follow-up in the vaccine and placebo groups, respectively. The following formulas will be used for the endpoints of the confidence interval for the vaccine efficacy: lower bound $VE_L = (1-p_U/k)/(1-p_U)$ and upper bound $VE_U = (1-p_L/k)/(1-p_L)$, where $k= k_v/(k_v + k_p)$, Follow-Up in Vaccine Group/Total Follow-Up.

D. Statistical Methods (Cont.)

b. Counting of Efficacy Endpoints

Subjects will be counted only once for the primary efficacy analysis. If a subject develops CIN 2/3 related to HPV 16 at Year 2 and develops CIN 2/3 related to HPV 18 at Year 3, the subject will be counted as one CIN 2/3 endpoint in the primary efficacy analysis. For the efficacy summaries by HPV type, the subject will be counted once in the HPV 16-related CIN 2/3 analysis and once in the HPV 18-related CIN 2/3 analysis.

c. Computation of Follow-Up Time

In the primary per-protocol analysis, follow-up for CIN 2/3 and cervical cancer begins following the Month 7 visit. Therefore, each subject's follow-up time will be computed by calculating the number of days between her Month 7 visit date and her final visit date. For cases, the "final visit" will be the first visit at which CIN 2/3 or cervical cancer was detected for the purpose of person-time computation. This value will be converted to person-years by dividing by 365.25. To obtain the total follow-up time in each treatment group, k_v and k_p, the person-years will be summed over all subjects in the given treatment group.

Although the primary analysis is per-protocol and HPV type-specific, all subjects who are eligible for the per-protocol analysis based on at least 1 of the 2 HPV types being analyzed (HPV 16 or 18) will contribute follow-up time to the totals. For the exploratory analysis of all CIN 2/3, all subjects who are eligible for the per-protocol analysis based on at least 1 high-risk HPV type will contribute follow-up time to the totals.

d. Correlates of Protection

In addition to the methods described in Section III.C.3.c. for assessing the anti-HPV serum RIA responses as potential correlates of protection, the odds of developing breakthrough HPV 16/18-related CIN 2/3 or worse may be modeled as a function of the corresponding Postdose 3 RIA titers using a logistic regression model. Length of postvaccination follow-up and study site will be included as additional model terms. Statistical significance of the coefficient for the RIA titer will be an indication that RIA titer is a predictor of breakthrough

D. Statistical Methods (Cont.)

infection. The estimate of the coefficient will indicate the nature of the relationship. If there are few cases among the vaccine recipients, there may be insufficient information to establish an immune correlate of protection in this study.

e. Potential Bias

1) Ascertainment Bias

Protocol 015 requires pelvic examinations which include an examination for genital warts. A genital wart examination introduces the potential for ascertainment bias in this study. Subjects are triaged to colposcopy and biopsy based on the results of the Pap tests performed approximately every year. However, in clinical practice, colposcopy is frequently performed based solely on a discovery of genital warts as well. Since a quadrivalent HPV (Types 6, 11, 16, and 18) L1 VLP vaccine is being administered, this could introduce bias if the vaccine is efficacious against genital warts. If the vaccine is efficacious against genital warts, warts will be found more often in the placebo recipients than in the vaccine recipients. Thus, the placebo recipients will be referred for colposcopy more often than the vaccine recipients, and, consequently, more CIN lesions will be detected in these subjects. This may cause the vaccine to appear more efficacious against CIN than it really is. To avoid this type of ascertainment bias, cases of CIN 2/3 or worse that are detected during a colposcopy that was performed *solely* based on a finding of genital warts (i.e., without an accompanying abnormal Pap test result) will not be included in the primary or secondary analyses outlined in this DAP, nor will these CIN cases count toward the total number of cases needed to trigger the analyses. Sensitivity analyses will be performed in which all cases of CIN 2/3 or worse are included regardless of the reason the colposcopy was performed. The Pap test result that will be considered in determining whether or not the colposcopy was performed based on a genital wart finding only will be the Pap test result from the most recent *scheduled* visit prior to the colposcopy date or on the colposcopy date. Pap test results from unscheduled visits at which colposcopies were performed will not be considered because they also represent data that could be influenced by the vaccine's impact on the genital wart endpoint.

D. Statistical Methods (Cont.)

2) Censoring Due to Definitive Therapy

When a subject has a biopsy performed, the subject is referred for definitive therapy based on the central laboratory's diagnosis of the biopsy (not the consensus diagnosis of the pathology panel). A subject is typically referred for a central laboratory diagnosis of CIN 2 or worse. Once a subject is referred for definitive therapy, the subject is censored for the primary and secondary evaluations of vaccine efficacy. If the subject meets the case criteria for an endpoint based on any tissue samples collected up to the time of definitive therapy, or if the tissue sample collected at the time of definitive therapy qualifies her as an endpoint, she will be counted as a case.

Since the pathology panel diagnosis is the diagnosis that determines the endpoint status of a subject and the central laboratory diagnosis determines whether or not the subject is referred for treatment, there is the potential for a subject to be censored for the analysis based on a central laboratory diagnosis of CIN 2 or worse, while the pathology panel diagnosis of the same lesion is less severe disease than CIN 2 or no disease at all. Though this situation is expected to occur infrequently, it could cause some of the subjects with the most potential for being developing an endpoint to be censored before developing endpoints. Consequently, it could cause a slight bias in the estimate of vaccine efficacy. To assess the impact of this situation on the efficacy estimate, a sensitivity analysis will be performed in which the endpoint definition is based on the more severe of the central laboratory or pathology panel diagnosis of each biopsy. (Since subjects are referred for definitive therapy based on a diagnosis of CIN 2 or worse, this type of analysis will essentially include all subjects referred for definitive therapy as cases of CIN 2/3 or worse.)

f. Missing Efficacy Data

Occasionally in the context of clinical studies, critical data are unavailable for analysis. This usually occurs because the data are not collected (e.g., the subject missed a visit or refused an examination) or because the samples from which the data are generated are mishandled (e.g., samples are lost or broken before reaching the laboratory) or are unsatisfactory (e.g., the slide cannot be read by the pathologist or the swab does not amplify in the PCR assay).

D. Statistical Methods (Cont.)

With respect to counting cases of CIN 2/3 and cervical cancer for the efficacy analyses, if a subject has a biopsy, ECC, or LEEP/conization specimen collected during the efficacy evaluation phase (post-Month 7) and the PCR result or pathology panel diagnosis is missing for the specimen, then the subject cannot be classified as a case based on that specimen.

When the missing data are needed to establish a subject's eligibility for analysis, the following rules will apply:

- With respect to missing serology results, subjects who are missing a baseline RIA result for a particular vaccine HPV type *will not* be eligible to be classified as cases of CIN 2/3 or cervical cancer related to that HPV type in the primary analysis.

- With respect to missing PCR results for cervicovaginal specimens, a subject's eligibility for analysis depends on the number of missing results. In Protocol 015, the PCR results for 2 cervicovaginal specimens collected at enrollment and 2 cervicovaginal specimens collected at Month 7 are used to determine each subject's eligibility for analysis. Subjects who are missing one or both of the PCR results for a given vaccine HPV type at enrollment or Month 7 will not be eligible to be cases of CIN 2/3 or cervical cancer related to that HPV type in the primary analysis.

- If the PCR result from a biopsy sample taken between enrollment and Month 7 (inclusive) is missing for a given vaccine HPV type, and the biopsy is diagnosed as abnormal, the subject *will not* be eligible to be classified as a case of CIN 2/3 or cervical cancer related to that type. If the PCR result is missing and the diagnosis is normal, the subject *will* be eligible. This rule was established because abnormal tissue is likely to be HPV PCR positive.

Similar rules will be applied for the secondary analysis.

Missing data which results from a subject dropping out of the study will be treated as missing (non-existent) in the primary analysis. However, a sensitivity analysis will be conducted in which the number of cases expected among subjects who are lost to follow-up will be projected assuming that these subjects would have been classified as cases with the same probability as those subjects who were not lost to

D. Statistical Methods (Cont.)

follow-up. In this analysis, it will be assumed that subjects who are lost to follow-up prior to Month 7 (i.e., prior to completion of the vaccination series) would have been classified as cases at the same rate as the placebo recipients in the primary analysis regardless of treatment group. That is, an assumption of "no vaccine efficacy" will be made in subjects who do not complete the vaccination regimen. The incidence rate that will be assumed for subjects who are lost to follow-up at Month 7 or following Month 7 (i.e., after completion of the vaccination series) will depend on each subject's treatment group. It will be assumed that vaccine recipients who complete the vaccination series and then drop out subsequently would have developed the endpoint of interest at the same rate as the vaccine recipients in the primary analysis (who will all have also completed the vaccination series). It will be assumed that placebo recipients who complete the vaccination series and then drop out subsequently would have developed the endpoint at the same rate as the placebo recipients in the primary analysis.

g. Interaction

Prior to conducting the formal statistical analyses described above, tests for study site by treatment group interaction will be performed. This section describes the general approach that will be taken for assessing the presence and nature of the interactions.

All interaction tests will be conducted at the $\alpha = 0.10$ level. If no statistically significant interaction is found, then the study sites will be pooled for the analysis. If a significant treatment-by-study center interaction is found, the nature of the interaction will be examined and the relevance of the interaction in the context of the planned analysis will be evaluated. If the interaction cannot be disregarded, certain study sites may be analyzed separately.

To assess the impact of treatment group-by-study center interaction on the incidence rates of the various disease endpoints, an exact test for homogeneity in relative risks among the study sites will be conducted using the method of Martin and Austin [2]. Due to the small numbers of cases of certain disease endpoints expected in this study, there will be low power to detect treatment group-by-study center interaction. Therefore, treatment group-by-study center interaction may need to be assessed observationally. If this is the case, any observed interaction will be described, but it may not be possible to analyze study sites separately and still maintain adequate power for the primary and secondary analyses.

D. Statistical Methods (Cont.)

2. Methods for Immunogenicity Analyses

a. Consistency of Immune Responses as Measured by Proportions

The methodology of Wiens, Heyse, and Matthews [3] will be used to demonstrate consistency among the 3 separate lots of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine. According to this methodology, the evaluation of similarity in response rates (the proportions of subjects with anti-HPV ≥200 mMU/mL at 4 weeks Postdose 3) between lots will be based on 3 pairwise comparisons for each vaccine HPV type. For each HPV type, each pairwise comparison is intended to demonstrate the equivalence of 2 of the 3 lots within 10 percentage points using 2 one-sided tests at the $\alpha = 0.05$ level for an overall type I error rate of 0.05 as outlined in Table 8. The tests will be stratified by study center as described in Section VII.D.2.b. Since there are 3 consistency lots, there will be 6 one-sided comparisons for each vaccine HPV type.

Table 8

Comparisons for Establishing Consistency

Pairwise Comparison	Group 1	Group 2
1	Lot 1	Lot 2
2	Lot 1	Lot 3
3	Lot 2	Lot 3

Two lots will be considered similar if both one-sided p-values associated with the 2 comparisons within the pair are <0.05 (corresponding to the rejection of the associated null hypothesis). This criterion is equivalent to requiring the two-sided 90% confidence interval for the difference in response rates between the pair of lots be entirely within (-10%, 10%). All 3 lots will be considered consistent with respect to response rates for a given vaccine HPV type if all 6 one-sided p-values for that vaccine HPV type are <0.05, or, equivalently, if all 3 pairwise 90% two-sided confidence intervals for the vaccine HPV type are entirely within (-10%, 10%).

D. Statistical Methods (Cont.)

b. Equivalence With Respect to Proportions

As described above in Section VII.D.2.a., the first primary immunogenicity hypothesis will be addressed by 3 pairs of one-sided tests for 2 binomial proportions conducted at the $\alpha = 0.05$ level for each vaccine HPV type. The tests will be stratified by study center.

The statistical hypotheses that will be tested are:

$$H_0: \; p_1 - p_2 \leq -S_0 \text{ or } p_1 - p_2 \geq S_0$$

$$H_a: \; -S_0 < p_1 - p_2 < S_0$$

where H_0 and H_a are the null and alternative hypotheses, respectively, p_1 and p_2 are the proportions of subjects with anti-HPV ≥200 mMU/mL at 4 weeks Postdose 3 for the HPV type of interest in the 2 quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine lots being compared, and S_0 is the prespecified equivalence margin for the 2 proportions.

Statistical significance for the 2 one-sided equivalence tests for each pair of lots is established if the p-values for the hypothesis tests are each <0.05. This corresponds to the 90% confidence interval for $p_1 - p_2$ being entirely contained within the interval $(-S_0, S_0)$ where S_0 is 10 percentage points for this study. Since both one-sided tests for each pair of lots are required to reject the null hypothesis at the $\alpha = 0.05$ level, the overall α-level will be controlled at $\alpha = 0.05$ for each pair of lots.

The proposed method for testing equivalence is based on a test for a non-zero difference in binomial proportions by Miettinen and Nurminen [1] with stratification by study center. The test statistic (Z-statistic) is given by:

$$Z_{diff} = \frac{\hat{p}_1^* - \hat{p}_2^* - S_0}{\sqrt{\sum_{i=1}^{I} (W_i / \sum_{k=1}^{K} W_k)^2 \, \tilde{V}_i}} \quad \text{for the first one-sided test, and}$$

$$Z_{diff} = \frac{\hat{p}_1^* - \hat{p}_2^* + S_0}{\sqrt{\sum_{i=1}^{I} (W_i / \sum_{k=1}^{K} W_k)^2 \, \tilde{V}_i}} \quad \text{for the second one-sided test, (1)}$$

where

D. Statistical Methods (Cont.)

$$\hat{p}_s^* = \sum_{i=1}^{I} (W_i / \sum_{k=1}^{K} W_k) \ \hat{p}_{si} \quad \text{for } s = 1, 2;$$

$$\tilde{V}_i = \left[\frac{\tilde{p}_{1i}(1-\tilde{p}_{1i})}{n_{1i}} + \frac{\tilde{p}_{2i}(1-\tilde{p}_{2i})}{n_{2i}} \right] \frac{(n_{1i}+n_{2i})}{(n_{1i}+n_{2i}-1)} \qquad (2);$$

and

$\hat{p}_{1i}$ = observed proportion with the response of interest in Group 1 at study center *i*;

$\hat{p}_{2i}$ = observed proportion with the response of interest in Group 2 at study center *i*;

n_{1i} = number of subjects in Group 1 at study center *i*;

n_{2i} = number of subjects in Group 2 at study center *i*;

$W_i = n_{2i} + n_{1i}$ (i.e., the weight for study center *i*, the total sample size in the 2 groups at study center *i*);

$I = K$ = total number of study centers;

$i = k$ = study center;

$\tilde{p}_{1i}$ and $\tilde{p}_{2i}$ = the maximum likelihood estimates for the proportion of responders in Group 1 and Group 2, respectively, computed under the constraint $\tilde{p}_{2i} = \tilde{p}_{1i} + S_0$; and

S_0 = prespecified difference in proportions under the null hypothesis.

The test statistic, Z_{diff}, has an asymptotic standard normal distribution under the null hypothesis, H_0. Large values of Z_{diff} favor the alternative hypothesis, H_a.

Although the study conclusions will be based on the p-values for the hypothesis tests, two-sided 90% confidence intervals for the difference in proportions between each pair of lots will also be provided. Based on Miettinen and Nurminen [1], the two-sided 90% confidence interval is given by the roots for $PD = p_1 - p_2$ in the following equation:

D. Statistical Methods (Cont.)

$$\chi^2_\alpha = \frac{(\hat{p}_1^* - \hat{p}_2^* - PD)^2}{\sum_{i=1}^{I} (W_i / \sum_{k=1}^{K} W_k)^2 \ \tilde{V}_i} \qquad (3)$$

where

χ^2_α is the 100(1-2α)th percentile from the central chi-square distribution with 1 degree of freedom ($\chi^2_\alpha = 2.71$ for two-sided 90% confidence interval);

$\hat{p}_1^*$, $\hat{p}_2^*$, $\tilde{V}_i$ and W_i, are the same as defined earlier in the Z-statistic;

PD is the difference between 2 proportions ($PD = p_1 - p_2$);

I =K= total number of study centers;

i =k = study center;

$\tilde{p}_{1i} = \tilde{p}_{2i} + PD$ for i=1,...,I;

$\tilde{p}_{2i}$ is the maximum likelihood estimate for p_{2i} under the constraint $\tilde{p}_{1i} = \tilde{p}_{2i} + PD$.

Equation (3) does not allow for explicit solutions for PD; therefore, a numerical algorithm will be used to obtain the 2 roots for PD (the upper and lower bounds of the confidence interval). Programming details for the numerical algorithm and the maximum likelihood estimates can be found in Miettinen and Nurminen [1].

c. Consistency in Immune Responses as Measured by GMTs

Using the methodology of Wiens, Heyse and Matthews [3], the evaluation of similarity in GMTs between lots will be based on 3 pairwise comparisons for each vaccine HPV type. For each HPV type, an ANOVA model will be constructed with natural log titer as the dependent variable and treatment group, study center, and treatment-by-study center interaction as fixed effects. Sample size weights will be used in estimating an overall treatment effect. Using the mean squared error (MSE) from the final ANOVA model as an estimate of variance, each pair of lots will be compared using 2 one-sided tests at the $\alpha = 0.05$ level, for an overall type I error rate of 0.05 as outlined in Table 8. Since there are 3 consistency lots, there will be 6 one-sided comparisons for each vaccine HPV type.

D. Statistical Methods (Cont.)

Two lots will be considered similar if both one-sided p-values associated with the 2 comparisons within the pair are <0.05 (corresponding to the rejection of the associated null hypothesis). This criterion is equivalent to requiring the two-sided 90% confidence interval for the difference in response rates between the pair of lots be entirely within (0.5, 2). All 3 lots will be considered consistent with respect to response rates for a given vaccine HPV type if all 6 one-sided p-values for that vaccine HPV type are <0.05, or, equivalently, if all 3 pairwise 90% two-sided confidence intervals for the vaccine HPV type are entirely within (0.5, 2).

d. Equivalence With Respect to GMTs

As described above in Section VII.D.2.c., the second primary immunogenicity hypothesis will be addressed using 3 pairs of one-sided tests conducted at the $\alpha = 0.05$ level for each vaccine HPV type. Fold differences in GMTs, the associated two-sided 90% confidence intervals and p-values will be provided using an ANOVA model. The ANOVA model will model the natural log of the individual titers of the subjects as a function of treatment group, study center and treatment-by-study center interaction, all of which will be fixed effects. In estimating the differences between treatment groups, a weight proportional to the sample size at a given study center divided by the total evaluable sample size will be used for each study center. Using the Mean Squared Error (MSE) from the final ANOVA model as an estimate of variance, one-sided tests for the similarity of 2 means will be performed at the $\alpha = 0.05$ level.

The 2 one-sided tests for each pair of lots will be used to test that the fold difference in GMTs between the lots is no greater than 2-fold in either direction. The statistical hypotheses that will be tested are as follows:

$$H_0: \ GMT_1/GMT_2 \leq 0.5 \text{ or } GMT_1/GMT_2 \geq 2$$

$$H_a: \ 0.5 < GMT_1/GMT_2 < 2$$

where GMT_1 and GMT_2 represent the GMTs in Group 1 and Group 2, respectively. Since the analysis will be conducted on the natural log scale, the corresponding hypotheses on this scale are:

$$H_0: \ \mu_1 - \mu_2 \leq \ln(0.5) \text{ or } \mu_1 - \mu_2 \geq \ln(2)$$

$$H_a: \ \ln(0.5) < \mu_1 - \mu_2 < \ln(2)$$

D. Statistical Methods (Cont.)

where μ_1 and μ_2 are the true means of the natural log titers in Group 1 and Group 2, respectively.

Statistical significance for the 2 one-sided equivalence tests for each pair of lots is established if the p-values for the hypothesis tests are each less than 0.05. This corresponds to the 90% confidence interval for the fold difference in the 2 lots being contained entirely within (0.5, 2). Assuming the normality of the natural log titers, the 90% confidence interval for the geometric mean fold difference in antibody titers will be provided using a t-distribution. The MSE from the final ANOVA model will be used as an estimate of variance.

Since both one-sided tests for each pair of lots are required to reject the null hypothesis at the $\alpha = 0.05$ level, the overall α-level will be controlled at $\alpha = 0.05$ for each pair of lots.

e. Interaction

Prior to conducting the formal statistical analyses of response rates and GMTs, tests for study center-by-treatment group interaction will be performed. This section describes the general approach that will be taken for assessing the presence and nature of the interactions. Study center-by-treatment group interaction will be assessed using 2 approaches. One will address interactions when the variables of interest are proportions, and the other will address interactions when the variables of interest are the GMTs.

1) Interaction Assessment for Difference in Proportions

A test for treatment-by-study center interaction will be conducted to evaluate whether or not the difference in proportions between 2 lots is consistent across study centers. Treatment-by-study center interaction will be tested separately for each pair of the 3 consistency lots and for each vaccine HPV type. An improved version (Mehrotra [4], Mehrotra & Chan [5]) of the Brown and Forsythe [6] testing procedure will be used.

The test statistic to be used is as follows:

$$F^* = \frac{\sum_{k=1}^{K} n_k \left(d_k - \bar{d}\right)^2}{\sum_{k=1}^{K} \left(1 - \frac{n_k}{N}\right) n_k V_k}$$

D. Statistical Methods (Cont.)

where,

$$d_k = \hat{p}_{1k} - \hat{p}_{2k}$$

$$\bar{d} = \frac{\sum_{k=1}^{K} n_k d_k}{N}$$

$$n_k = \frac{2n_{1k}n_{2k}}{n_{1k} + n_{2k}}$$

$$V_k = \left(\frac{\hat{p}_{1k}(1-\hat{p}_{1k})}{n_{1k}-1} + \frac{\hat{p}_{2k}(1-\hat{p}_{2k})}{n_{2k}-1} \right)$$

and,

K = number of study centers

$\hat{p}_{1k}$ = observed proportion with the response of interest in Group 1 at study center k

$\hat{p}_{2k}$ = observed proportion with the response of interest in Group 2 at study center k

n_{1k} = sample size in Group 1 at study center k

n_{2k} = sample size in Group 2 at study center k

$N = \sum_{k=1}^{K} n_{1k} + n_{2k}$ = total sample size.

The critical value that F^* will be compared to is $F_{0.90}(f_1, f_2)$, the 90th percentile of the central F distribution with f_1 and f_2 degrees of freedom. As shown by Mehrotra [4], appropriate values of f_1 and f_2 are given by:

D. Statistical Methods (Cont.)

$$f_1 = \frac{\left(\sum_{k=1}^{K} n_k V_k - \frac{\sum_{k=1}^{K} n_k^2 V_k}{N}\right)^2}{\sum_{k=1}^{K} (n_k V_k)^2 + \left(\frac{\sum_{k=1}^{K} n_k^2 V_k}{N}\right)^2 - 2\frac{\sum_{k=1}^{K} n_k^3 V_k^2}{N}}$$

$$f_2 = \frac{\left[\sum_{k=1}^{K}\left(1-\frac{n_k}{N}\right) n_k V_k\right]^2}{\sum_{k=1}^{K}\left(1-\frac{n_k}{N}\right)^2\left(\frac{(n_k V_k)^2}{n_k - 1}\right)}$$

These interaction tests will be conducted at the $\alpha = 0.10$ significance level. If a statistically significant interaction is detected (p-value ≤ 0.10), the discussion of such an interaction will include an evaluation of its relevance in the context of the planned equivalence test. For example, the treatment effect differences will be estimated for each study center separately and will be displayed graphically. Observational and graphical assessment of the magnitude and nature of the interaction will be made in relationship to the prespecified margin for the equivalence tests (10 percentage points). The procedure to test for qualitative interaction proposed by Gail and Simon [7] will be used to characterize the observed interaction. This procedure will be adjusted for the equivalence margin (10 percentage points). If the interaction cannot be disregarded, certain study centers may be analyzed separately.

2) Interaction Assessment for Ration of GMTs

An ANOVA model will be used to evaluate whether or not the ratio of GMTs between 2 lots is consistent across study centers. In the ANOVA model, the natural log of the RIA titers will be the dependent variable and treatment group, study center and treatment-by-study center interaction will be the independent variables. The independent variables will be treated as fixed effects in the model.

D. Statistical Methods (Cont.)

These interaction tests will be conducted at the $\alpha = 0.10$ level. Regardless of whether or not a given interaction test is statistically significant, the treatment-by-study center interaction term will be included in the final ANOVA model. If a statistically significant interaction is detected, the discussion of such an interaction will include an evaluation of its relevance in the context of the planned test of equivalence. For example, the ratios of the GMTs between treatment groups will be estimated separately by study center and will be displayed graphically. Observational and graphical assessment of the magnitude and nature of the interaction will be made in relationship to the prespecified GMT ratio for the GMT comparison (0.5). The procedure to test for qualitative interaction proposed by Gail and Simon [7] will be used to characterize the observed interaction. This procedure will be adjusted for the equivalence margin (2-fold). If the interaction cannot be disregarded, certain study centers may be analyzed separately.

These interaction tests will only be performed for the per-protocol population (primary analyses) since the per-protocol and the All Type Specific HPV Naïve Subjects With Serology Data populations are not expected to differ substantially.

f. Pooling Subjects for Analysis

For the stratified analyses and treatment-by-study center interaction evaluations mentioned above, if some study centers have fewer than 5 subjects in 1 treatment group, the study centers concerned will be pooled with other sites that are in a similar geographic region to form a general "geographic site" with enough subjects in each treatment group (≥5 subjects per treatment group).

In addition, some study centers have multiple subsites within the main investigative site. Data from subsites will be combined with the main site for evaluating treatment-by-study center interaction and for any analyses which stratify by study center.

E. Multiplicity

There is only one primary efficacy hypothesis for the study. It will be tested with an interim analysis at the time that at least 19 endpoint cases have been observed and at a final analysis when at least 29 endpoint cases have been observed. In order to control the overall type I error rate for the efficacy study

E. Multiplicity (Cont.)

at $\alpha = 0.05$ two-sided, the α-levels used at the interim and final analyses must be adjusted to account for the multiple, correlated analyses. The effective α-levels at the interim and final analyses were computed using the power boundaries of Wang and Tsiatis [8]. Using an alpha boundary shape of 0.2 (where 0.0 represents the O'Brien-Fleming boundary and 0.5 represents the Pocock boundary), a two-sided alpha of 0.0204 will be spent at the interim analysis, and a two-sided alpha of 0.0411 will be spent at the final analysis.

There are 2 primary immunogenicity hypotheses for the consistency lot substudy. However, the success of the substudy requires that consistency be established for all 4 vaccine HPV types with respect to both rates and GMTs, so no multiplicity adjustment is necessary for the substudy.

The goals of the efficacy study and consistency lot substudy are considered independent. Therefore, no multiplicity adjustment will be made for the multiple hypotheses.

Note that no multiplicity adjustments will be made for the treatment group comparisons of multiple safety endpoints. Caution should be exercised when interpreting results when the overall type I error is not controlled.

F. Subgroup Analysis

All age-related analyses will be footnoted to acknowledge age differences among the study sites caused by the country-specific amendments to Protocol 015. Finland is primarily enrolling 16- to 17-year-old subjects, and Singapore is primarily enrolling subjects who are 20 to 26 years of age.

1. Efficacy

Primary efficacy estimates will be provided by geographical region (United States, Latin America, Europe, and Asia-Pacific). Due to the small number of expected primary endpoints, the primary efficacy estimates by race and age (grouped into 16 to 17, 18 to 19, 20 to 21, 22 to 23, and 24 to 26) will be provided in an integrated summary of efficacy results from across all of the Phase II/III studies. In addition, modeling techniques will be explored to assess the effects of various subgroups on the efficacy estimates in the integrated summary of efficacy.

2. Immunogenicity

Subgroup analyses will be performed to assess whether or not baseline subject characteristics are associated with Month 7 HPV responses. The

F. Subgroup Analysis (Cont.)

baseline subject characteristics that are of interest are race (Black, White, Asian, or Other), age, geographic region (United States, Europe, Latin America, and Asia), smoking status (current smokers or not), pregnancy status (ever pregnant or not), lifetime number of sexual partners, age of first sexual intercourse, and contraceptive use (hormonal or non-hormonal).

Multiple linear regression models will be constructed with the natural log of the Month 7 titers for a given vaccine HPV type as the response variable and the baseline subject characteristics as explanatory variables. The model will also include a term for treatment group. Separate models will be constructed for each HPV vaccine type. Simple linear regression models, which model the natural log of the Month 7 titers as a function of each of the above subject characteristics separately, will also be constructed to supplement the multiple regression model.

Since several of the baseline subject characteristics are expected to be highly correlated (for example, lifetime number of sexual partners and age at first sexual intercourse), multicollinearity may be observed in the multiple regression model. Correlations between baseline subject characteristics will be estimated. Pearson's correlation coefficient will be used if normality in the variables can be assumed. Kendall's tau correlation coefficient will be used otherwise. If there appear to be significant correlations among baseline subject characteristics, the interaction effects of the factors that are associated will be included in the multiple logistic regression models.

For any characteristic that is found to be significant ($p > 0.10$ will be used for reporting significant variables) in either the multiple regression model or the simple regression models, tabulations of the GMTs by treatment group and categories of the characteristic of interest may be constructed to further explore the relationship between the variables. In addition, RCDFs may be used to visualize the effect of the factor on RIA titers. Month 7 responses by race, age, and geographic region will be tabulated by treatment group regardless of whether or not the models show significance for these variables. Age will be tabulated using the following categories: 16 to 17, 18 to 19, 20 to 21, 22 to 23, 24 to 26. The GMTs at Months 12, 24 and 48 will also be tabulated by race, age, and geographic region.

Data permitting, Month 7 GMTs for each HPV type will also be tabulated for subjects who had cervicovaginal diseases at baseline. Baseline status (positive or negative) for infection with bacterial vaginosis, chlamydia,

F. Subgroup Analysis (Cont.)

gonorrhea, herpes, syphilis, or trichomonas separately will be summarized by treatment group. Tables will be footnoted to indicate that subjects in Denmark, Finland, Iceland, Norway, Poland, and Sweden did not undergo routine gonorrhea testing.

VIII. GROUP RULES FOR ANALYSIS

A. Definition of Time Points

Each of the subjects in each study will receive 3 doses of quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP vaccine or placebo. The vaccine or placebo will be administered at 0, 2, and 6 months. For the efficacy analyses, there will be no restrictions placed on the timing of the 3 vaccination visits as long as they occur within 1 year, or any of the follow-up visits with the exception of the Month 7 visit. The postvaccination clinical samples collected at the Month 7 visit will only be considered as valid for establishing the PCR status of a subject if they are obtained 14 to 72 days following the Month 6 vaccination.

For the per-protocol immunogenicity analyses, acceptable timing for the Month 2 vaccination visit will be 36 to 84 days (60 days ± 24 days), and acceptable timing for the Month 6 vaccination visit will be 148 to 218 days (183 days ± 35 days). The postvaccination clinical samples collected at Month 7 will be considered for the immunogenicity analyses if they are obtained 14 to 49 days (2 to 7 weeks) following the Month 6 vaccination. The acceptable day range for vaccination are slightly wider than the day ranges specified in the protocol to allow for differences in counting when defining a "month" (calendar month versus 30 days). The day range for the Month 7 visit is slightly wider than that specified in the protocol to ensure that all HPV protocols use consistent day ranges in their analyses. (The Phase I protocols specified slightly different ranges for the Month 7 visit [2 to 6 weeks postdose 3] than the Phase II/III studies [3 to 7 weeks postvaccination]).

The allowable day range for the Month 7 visit is narrower for the immunogenicity analyses than for the efficacy analyses because in the immunogenicity analyses, it is important to try to measure a Postdose 3 immune response that is as close to the peak response as possible. The goal of the Month 7 visit for the efficacy analyses is to provide an assessment of the subjects' "1 month Postdose 3" HPV status and mark the beginning of the follow-up period.

For the immunogenicity analysis of the persistence of antibody titers, only Month 24 serum samples collected 639 to 821 days (730 days ± 91 days)

A. Definition of Time Points (Cont.)

Postdose 1 and Month 48 serum samples collected 1369 to 1551 days (1460 days ± 91 days) Postdose 1 will be considered. These day ranges will be imposed on both the per-protocol and all type-specific HPV naïve subjects with serology data populations.

B. Definition of Which Value Within a Day Range Will be Used

For the efficacy analysis, with the exception of the Month 7 visit, multiple measurements within a day range are not possible, as no day ranges are imposed on the visits, scheduled or unscheduled. With respect to the Month 7 visit and the day ranges imposed on the Month 24 and 48 visits for the immunogenicity analysis, multiple visits within a day range are possible. However, since the visits in Protocol 015 are labeled, if multiple efficacy measurements are taken within the acceptable day range for a given visit, the values for the samples that are labeled with that visit by the study site will be the values that are used for the analysis. For example, if a subject comes to the clinic for a visit on Day 276 and again on Day 453, and the Day 276 visit is labeled Month 12 while the Day 453 visit is labeled Month 18, then the Day 276 visit will be used for the Month 12 visit, and the subject will be considered outside the day range for the Month 18 visit. As part of the routine screening process, the data will be screened prior to analysis to make sure that only 1 visit for each subject is labeled with each visit label.

C. Definition of Baseline Value

For serum samples, the baseline values for a study participant are the assay results obtained from the sample that was collected from that subject on the day of the initial vaccination. The PCR results that will be used to establish that subjects are negative for HPV 6, 11, 16, and/or 18 at baseline and throughout the vaccination phase of the study (through Month 7) are the HPV 6, 11, 16, and 18 results for the (b)(4) and (b)(4) swab and the (b)(4) swab at Day 1 and Month 7 and any biopsies collected between Day 1 and Month 7, inclusive.

D. Description of Data Handling Procedures Prior to Analysis

All data will be screened and cleaned prior to unblinding. This includes identification of subjects excluded from the per-protocol efficacy and immunogenicity analyses, the modified intention-to-treat efficacy analyses, and the "all type specific HPV naïve subjects with serology" immunogenicity analyses. All data handling guidelines and actions will occur prior to unblinding according to Merck's Standard Operating Procedure (SOP) for "in-

D. Description of Data Handling Procedures Prior to Analysis (Cont.)

house blinded" studies, i.e., blinding for study participants, investigators, and internal Merck personnel. The in-house unblinded database will be "frozen" in order to ensure that analyses of data in response to regulatory queries will be performed on the same data as were used for that submission.

E. Description of Protocol Violations

The per-protocol analyses of efficacy and immunogenicity will exclude data according to the following rules:

1. **Violations With Respect to the Vaccination Regimen or Sample Collection Regimen**

   a. Subjects who received any of their 3 vaccinations outside of the day ranges specified in Section VIII.A will be excluded from the immunogenicity analyses only. Subjects who did not receive all 3 vaccinations within a year will be excluded from the efficacy analysis.

   b. Subjects who received the incorrect clinical material or an incorrect dose of the correct clinical material at any of the 3 vaccination visits will be excluded.

   c. Incorrectly randomized subjects will be excluded.

2. **Inclusion/Exclusion Criteria Violations**

   a. Subjects who are younger than 16 or who are older than 26 years of age will be included.

   b. Subjects who have engaged in sexual intercourse within 48 hours prior to a scheduled visit that includes a pelvic examination will be included.

   c. Subjects who become pregnant prior to Month 6 and subsequently choose to continue the pregnancy (i.e., not terminate the pregnancy) will be included in the efficacy analyses as long as they received all 3 vaccinations within 1 year. These subjects will likely be excluded from the immunogenicity analyses as they most likely missed vaccinations due to the pregnancy.

   d. Subjects who have a lifetime history of more than 4 male or female sexual partners will be included.

E. Description of Protocol Violations (Cont.)

e. Subjects who are concurrently enrolled in clinical studies of investigational agents or studies involving collection of cervical specimens may or may not be included. These subjects will be judged on a case-by-case basis by the Clinical Monitor while the Clinical Monitor is still blinded and without knowledge of each individual's case status.

f. Subjects with a history of any prior abnormal Pap test showing SIL, ASC-US, ASC-H, or biopsy showing CIN will be included in the analyses of HPV type-specific endpoints as long as they are negative for the appropriate HPV types by an MRL assay. These subjects will be excluded from analyses involving HPV types for which MRL has no available type-specific assay.

g. Subjects with a history of genital warts or genital warts present at baseline will be included in the analyses if they were enrolled in the study.

h. Subjects with any known or suspected immune disorders may or may not be included. These subjects will be judged on a case-by-case basis by the Clinical Monitor while the Clinical Monitor is still blinded to each subject's treatment allocation and case status.

i. Subjects who are immunocompromised or have been diagnosed as having HIV infection will be excluded.

3. Concomitant/Prior Therapy Violations

a. Subjects who have received any immune globulin (including RhoGAM™) or blood derived products at any time through Month 7 of the study will be excluded.

b. Subjects receiving any immunosuppressives will be excluded with the exception of subjects using topical or inhaled steroids.

c. Subjects who receive any nonstudy inactivated vaccine within 14 days of a study vaccine will be excluded as well as subjects who receive any nonstudy live virus vaccine within 21 days before or 14 days after a study vaccine.

d. Subjects who have had prior vaccination with an HPV vaccine will be excluded.

IX. PRESENTATION AND FORMAT OF RESULTS

A. Outline of Results Section

The following is an outline of the anticipated sections that will appear in the "Results" section of the CSR.

Results

A. Study Subjects and Data Sets Analyzed
B. Efficacy Evaluation and Results
C. Immunogenicity Evaluation and Results
D. Safety Evaluations

B. Format of Data Within Tables

1. Decimals

All means, standard errors, percentages, and confidence intervals will be reported to 1 decimal place.

2. Rounding

Means, standard errors, percentages, and confidence intervals endpoints will be rounded up to the next decimal place if the second decimal place is ≥5 and down to the next decimal place if the second decimal place is <5.

C. Notation and Ordering of Treatment Groups

1. Efficacy and Safety

Group 1: Quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP Vaccine	Group 2: Placebo

2. Immunogenicity

Group 1: Quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP Vaccine - Consistency Lot 1	Group 2: Quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP Vaccine - Consistency Lot 2	Group 3: Quadrivalent HPV (Types 6, 11, 16, 18) L1 VLP Vaccine - Consistency Lot 3

D. Reporting of Statistical Significance

All p-values will be rounded to the nearest thousandth (3 decimal places). All statistical tests with the exception of the consistency lot and interaction tests will be conducted at a one-sided α-level of 0.025. The consistency lot comparisons will be conducted at a one-sided α-level of 0.05, and the interaction tests will be conducted at an α-level of 0.10. A p-value less than the prespecified α-level reflects statistical but not necessarily clinical significance.

X. REFERENCES

1. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med 1985;4:213-26.

2. Martin DO, Austin H. Exact estimates for a rate ratio. Epidemiology 1996;7:29-33.

3. Wiens B, Heyse J, Matthews H. Similarity of three treatments, with application to vaccine development. ASA Proceedings of the Biopharmaceutical Section 1996;203-06.

4. Mehrotra DV. Improving the Brown-Forsythe solution to the generalized Behrens-Fisher problem. Communications in Statistics-Simulation and Computation 1997;23:1139-45.

5. Mehrotra DV, Chan ISF. Testing for treatment by stratum interaction in stratified comparative binomial trials. August 13, 2000. ASA presentation, Indianapolis, IN.

6. Brown MB, Forsythe AB. The small sample behavior of some statistics which test the equality of several means. Technometrics 1974;16(1):129-32.

7. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics 1985;41:361-72.

8. Wang SK, Tsiatis, AA. Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 1987;43:193-99.

XI. APPENDICES