The Pennsylvania State University

The Graduate School

College of Medicine

ANTIMALARIALS IN INCOMPLETE LUPUS ERYTHEMATOSUS – SIMULATION

STUDY OF AN ONGOING CLINICAL TRIAL

A Thesis in

Public Health Sciences

by

Danielle M. Feger

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

December 2017

The thesis of Danielle Feger was reviewed and approved* by the following:

Vernon M. Chinchilli

Distinguished Professor and Chair

Thesis Advisor

Duanping Liao

Professor and Vice Chair for Research

Nancy Olsen

Professor of Medicine

Kristen Kjerulff

Professor and Director

*Signatures are on file in the Graduate School.

Abstract

Incomplete lupus erythematosus (ILE) is believed to be a preclinical form of systemic lupus erythematosus (SLE)[1], a chronic and potentially debilitating autoimmune disease. Early treatment is necessary to limit the potential for irreversible organ damage[2,3]. Hydroxychloroquine (HCQ), an antimalarial drug, is widely used in the treatment of both diseases[1,4]. However, it is unknown whether HCQ slows the progression of ILE to systemic lupus. The Study of Antimalarials in Incomplete Lupus Erythematosus (SMILE), a 96-week randomized, double-blind, placebo-controlled clinical trial, recently was initiated in autumn 2017. This trial proposes to quantify the effectiveness of HCQ to slow or halt the progression of incomplete lupus into systemic lupus, as defined by the Systemic Lupus International Collaborating Clinics (SLICC) criteria.

The purpose of this thesis is to calculate statistical power and necessary sample size for the SMILE trial. Of primary interest is the risk of accumulating one or more SLICC criteria; secondary outcomes of interest include the risk of progressing from ILE to SLE, as well as changes in 52 soluble mediators and 13 autoantibodies over the course of the study. Original estimates for expected risk between treatment and placebo groups were provided by the principal investigators, and additional simulations to account for variance in risk estimates, inflation of sample size, and increased loss to follow-up were performed.

Forty percent of patients receiving placebo and 20% of patients receiving hydroxychloroquine are expected to accumulate at least 1 SLICC criteria throughout the course of the study. With a total of 180 subjects, there is 90% statistical power to detect a statistically significant difference with a two-sided, 0.05 level test between treatment and placebo groups. Total patient recruitment may be inflated to 240 subjects to allow for up to 25% loss to follow-up while maintaining 90% statistical power. Statistical power to detect a significant difference between treatment and placebo groups with a two-sided, 0.05 level test is decreased when the risk ratio between groups is decreased, and increased when the proportion of patients accumulating SLICC criteria increased.

Overall, the SMILE trial is well-powered to detect a significant difference in the risk of accumulating one or more SLICC criteria between treatment and placebo groups with a target sample size of 180 subjects needed to complete the trial. Statistically significant differences in secondary outcomes also are detectable at this sample size. However, previous clinical trials of lupus patients have shown relatively large loss to follow-up in this population. Therefore an enlarged target sample size of 240 subjects is appropriate for the SMILE trial.

Table of Contents

List of Figures

List of Tables

Chapter 1. Background and Scientific Rationale for SMILE Trial

Systemic lupus erythematosus (SLE) is a serious autoimmune disease that occurs most frequently in young women. It can present with a multitude of clinical and immunologic manifestations, making early diagnosis challenging. Symptoms may not all present at once and instead accumulate over years[5]. While some forms of lupus primarily involve the skin, systemic lupus can affect multiple organ systems and cause life-threatening damage[6]. SLE is a chronic disease and treatment is lifelong. Given the disease's potential to cause irreversible damage and limit quality of life for many years due to occurring in a young population, it is of great interest for both patients and physicians to initiate early medical intervention in persons at risk of developing SLE.

Incomplete lupus (ILE) is believed to be a heterogeneous form of preclinical systemic lupus[1]. Unfortunately, there is no one specific diagnostic test for ILE. A combination of both clinical and immunologic criteria are used to diagnose SLE; a patient is considered to have incomplete lupus if he/she does not accumulate enough criteria to obtain a SLE diagnosis. A variety of classification criteria methods, including the 2012 Systemic Lupus International Collaborating Clinics (SLICC) Criteria[7], and 1997 revised American College of Rheumatology (ACR) criteria[8], have been developed. A combination of factors is needed to define who is at risk[9].

Anti-nuclear antibody (ANA) positivity is required for both ILE and SLE diagnosis; however, ANA positivity is common in many other connective tissue diseases[10] and up to 20% of the healthy population has significant levels of ANA[11,12]. Antibodies such as anti-double-stranded DNA (dsDNA), anti-Smith, and antiphospholipid (anticardiolipin and anti-B$_2$-glycoprotien) are more specific to SLE and are included in the 2012 SLICC criteria, however a person with a combination of positive ANA, one autoantibody, and only one clinical criteria would not meet diagnostic criteria for SLE. Other autoantibodies may accumulate in both ILE and SLE individuals; the presence of specific combinations of autoantibodies may be able to predict which ILE patients progress to SLE[9].

Multiple organ systems are affected by lupus; common areas of involvement include the skin, kidneys, connective tissue, circulatory, and central nervous systems[6]. Effects can range from mild (e.g., butterfly rash) to life-threatening (e.g., renal failure, thrombosis). Clinical criteria used in the SLICC classification criteria include: acute or chronic cutaneous lupus, oral ulcers, non-scarring alopecia, synovitis, serositis, renal involvement, neurological involvement, hemolytic anemia, leukopenia or lymphopenia, and thrombocytopenia[7].

Treatment such as hydroxychloroquine (HCQ) and other antimalarial drugs are indicated for lupus and other rheumatologic disorders[4]. The efficacy of hydroxychloroquine in SLE patients has been well established. While the drug is known to have many beneficial effects, it is particularly effective at treating skin manifestations of lupus as well as arthralgias[13,14]. A large multiethnic lupus cohort, the LUMINA cohort, has provided much evidence regarding beneficial effects of HCQ. Five years after lupus diagnosis, 5% of patients taking hydroxychloroquine had skin damage while 24% of patients not taking hydroxychloroquine had skin damage[2]. Additionally, SLE patients not treated with hydroxychloroquine at the time of enrollment had significantly higher renal and neuropsychiatric disease damage at follow-up, while patients receiving HCQ had a reduced risk of accruing damage, particularly in those without damage at enrollment[3]. This suggests that early treatment with HCQ may help to prevent new damage.

Continuation of hydroxychloroquine treatment is also beneficial in established lupus patients. In a Canadian cohort of lupus patients being treated with hydroxychloroquine, patients switched to placebo were 2.5 times more likely to experience a clinical flare-up of disease compared to those who continued receiving HCQ[15]. Further analysis of the LUMINA cohort found that hydroxychloroquine use was protective of survival, even after controlling for demographic and clinical disease characteristics[16].

While it is also commonly prescribed to patients with ILE[1], few studies have been conducted to effectively show hydroxychloroquine's ability to slow or halt progression to systemic lupus. In a retrospective study of American military personnel who were diagnosed with SLE, those treated with hydroxychloroquine had a longer time from onset of first clinical symptom to systemic disease diagnosis[17]. However, to the best of the author's knowledge, no randomized double-blind placebo-controlled clinical trial on the efficacy of hydroxychloroquine in ILE has been conducted to date.

Chapter 2. SMILE Study Design

The purpose of the Study of Antimalarials in Incomplete Lupus Erythematosus (SMILE) trial is to address the gap in clinical knowledge and determine the effectiveness of hydroxychloroquine treatment in an incomplete lupus population. The SMILE Study is a randomized, double-blind, placebo-controlled clinical trial designed to establish whether hydroxychloroquine can effectively slow or prevent the progression from incomplete lupus to SLE. Each patient will be randomized to receive either an appropriate weight-based dosing of hydroxychloroquine sulfate or placebo treatment daily. Up to 240 patients with incomplete lupus will be enrolled across 6 study sites and followed at 12-week intervals for a period of 96 weeks. Additional data to be collected throughout the trial include disease activity, patient-reported outcomes, and laboratory biomarkers. Patients who meet diagnostic criteria for lupus as defined by the 2012 SLICC criteria at any point during the study will exit the trial. This study was approved by the Pennsylvania State University Institutional Review Board (IRB) as well as each participating site's IRB, with the Penn State Milton S. Hershey Medical Center serving as the lead clinical site.

Inclusion Criteria

To be eligible for the study, a patient must present with a positive ANA antibody at a titer of 1:80 or greater as detected by immunofluorescence assay (IFA), and 1 to 2 additional clinical or laboratory 2012 SLICC classification criteria. In addition, eligible patients must be between 15 and 45 years of age at visit 1 and provide written informed consent. Subjects under age 18 at time of enrollment must provide written informed assent and a parent or legal guardian must also provide consent in order to enroll in the trial.

Exclusion Criteria

Exclusion criteria in this study includes: diagnosis of SLE using the 2012 SLICC criteria or presence of positive ANA and biopsy-proven lupus nephritis; presence of any other autoimmune disorder excluding autoimmune thyroid conditions; fibromyalgia diagnosis; previous or current oral antimalarial agents including hydroxychloroquine, chloroquine, or quinacrine; current or previous treatment with immunosuppressive, immune modifying, or cytotoxic medications; use of any investigational agent within the previous 12 months; history of primary immunodeficiency; presence of active bacterial, viral, fungal, or opportunistic infection; known history of human immunodeficiency virus (HIV), or Hepatitis B or C infection; concomitant malignancy or history of malignancy with the exception of adequately treated basal or squamous cell carcinoma of the skin or carcinoma in situ of the cervix; findings on ophthalmological exam that prevent safe use of hydroxychloroquine; contraindications to treatment with hydroxychloroquine including pre-existing ocular disease, hepatic impairment, psoriasis, porphyria, or allergy to the drug or class; comorbidities requiring systemic corticosteroid therapy > 10 mg/day of prednisone or equivalent, or a change in corticosteroid dose within 4 weeks prior to visit 1; starting, stopping, or changing dose of over-the-counter (OTC) prescription non-steroidal anti-inflammatory drugs (NSAIDs) 4 weeks prior to visit 1; pregnant, breastfeeding, or unwilling to practice medically approved birth control throughout the study period; presence of a condition or abnormality that in the opinion of the investigator would compromise safety of the patient or quality of the data; or inability of comply with the study visit schedule and procedures. While the list of exclusion criteria is substantial, it ensures a well-defined study population and minimizes potential confounding effects.

Study Endpoints

The primary endpoint is the increase in clinical and laboratory features of SLE as defined by the 2012 SLICC classification criteria between week 0 and week 96. Secondary efficacy endpoints include: the proportion of patients who transition to SLE diagnosis using either the 2012 SLICC or 1997 revised ACR classification criteria, change in disease activity at every 3 months using the Modified SLEDAI (Systemic Lupus Erythematosus Disease Activity Index)-2K score and CLASI (Cutaneous Lupus Erythematosus Disease Area and Severity Index) cutaneous lupus score, the frequency of patients with anti-Smith, anti-RNP, anti-Ro, and anti-La antibodies every 6 months, the frequency of patients with abnormal dsDNA, C3, and C4 every 6 months, the change in disease damage scores using the SLICC/ACR damage index at baseline and 2 years, the proportion of patients who start, stop, or modify use of oral or topical corticosteroids during the study, and difference in response by race and ethnicity.

Data Collection

Up to 12 scheduled study visits will be performed for each patient over the 100-week period. Visit 1 serves as a screening visit to determine trial eligibility. Visit 2 consists of an ophthalmologic exam to ensure no pre-existing retinal pathology. Visits 3-11 will be conducted every 12 weeks and will allow for collection of biological samples, disease activity indices, and the completion of questionnaires. Visit 12 is a final ophthalmologic exam to ensure no hydroxychloroquine-related retinal damage occurred during the trial.

At visit 1, the patient will undergo informed consent, receive a screening number, and provide demographic and medical history for themselves, as well as autoimmune disease history for first-degree relatives. After a physical exam and collection of vital signs, blood and urine will be collected for the following laboratory tests: chemistry, hematology, pregnancy, urinalysis with microscopic examination, urine protein and creatinine concentration, complement 3 and 4 concentrations, direct anti-globulin test, G6PD level, 25-OH-vitamin D, anti-CCP3.1, and clinical autoantibodies.

Visit 2 consists solely of an ophthalmological exam for retinal pathology including dilated funduscopic examination (DFE), spectral domain ocular coherence tomography (SD-OCT), and Humphrey visual field (HVF) testing using a 10-2 pattern. Visits 3-11 are for sample and clinical data collection and occur every 12 weeks. The final visit, visit 12, is another ophthalmologic exam using the same procedures as visit 2. Each study site has a dedicated ophthalmologist who will perform all eye exams for visits 2 and 12.

At visit 3, after being checked for adverse events and concomitant medications, patients will undergo physical exams and have vital signs recorded. Any lupus disease activity will be measured using the SLICC, ACR, modified SLEDAI-2K and CLASI instruments, and any disease damage will be measured using the SLICC/ACR Damage Index. Patient self-assessments will be collected using the PROMIS and patient global assessment. Blood and urine will be collected for the following laboratory tests: chemistry, hematology, urinalysis with microscopic examination, urine protein concentration, anti-dsDNA antibodies by BioPlex, antibodies to extractable nuclear antigens by BioPlex, complement 3 and 4 concentration, multiplex cytokines, and autoantibody array. Blood and urine will be collected for DNA, RNA, serum, and plasma biobanking. Finally, the subject will be randomized to receive either hydroxychloroquine or placebo and dispensed the appropriate study drug.

Visits 4-6 and 8-10 are similar to visit 3, except for the following changes: The laboratory chemistry and pregnancy tests will not be performed, and the blood for multiplex cytokine and

autoantibody array measurements will not be collected. Additionally, medication compliance will be reviewed at each visit.

Visits 7 and 11 are similar to visit 3, except for the following changes: The laboratory tests for anti-dsDNA and antibodies to extractable nuclear antigens will not be performed, but the direct anti-globulin test will be performed. Blood for multiplex cytokines and autoantibody array measurements will not be collected. Medication compliance will be reviewed at each visit.

Feasibility

"Mock recruitment" for the SMILE study was conducted in 2015 during the study design phase in order to ensure an adequate potential patient sample. Patients who would meet eligibility criteria for the clinical trial were interviewed to provide their opinion and any suggestions for the proposed clinical trial. Approximately 60% of patients expressed potential interest in the trial, indicating that the projected number of patients screened will be approximately double the target enrollment number.

Power analyses for the primary and secondary outcomes were conducted to establish a target sample size for the trial. The rest of this thesis will expand on these power analyses.

Chapter 3. Methods

All power calculations and statistical analyses were performed in SAS/STAT® software version 9.4 (Cary, NC, USA). Particularly, PROC POWER was utilized for many of the power calculations. However for the primary outcome of interest, the difference in accumulation of SLICC criteria between the placebo and hydroxychloroquine treatment groups, no standard procedure built-in to the SAS software could calculate the necessary sample size for this study. Figures were produced using Python 3.5, particularly the matplotlib version 1.5.1 library[18].

To determine the required sample size for the primary outcome, a computer simulation of 1000 datasets was developed using predetermined assumptions about the proposed trial. Under the original assumptions of the clinical trial, it was expected that on average, 40% of placebo patients and 20% of HCQ patients (a relative risk of 2.0) would accumulate at least 1 SLICC criteria. Based on the above assumptions, SLICC criteria progression scores were randomly generated for each simulated participant. The difference in SLICC score progression between placebo and treatment groups was tested using a nonlinear mixed-effects model invoked by PROC NLMIXED for each dataset. The proportion of datasets, out of 1000, in which a significant difference at an alpha level = 0.05 was found approximates the statistical power to detect a difference using a two-sided, 0.05 significance level test at the given sample size.

Additional simulations were run to account for changes to the trial design. Three changes were implemented by the SMILE study team – an increase in total enrollment from 180 to 240 subjects, the addition of a 6th study site, and an allowance for up to 25% loss to follow-up. Therefore, additional simulations for 15%, 20%, and 25% loss to follow-up using the larger target sample size and 6th study arm were conducted. Other changes not under control of the study team, such as the effect of a change in the relative risk in accumulating SLICC criteria between the placebo and treatment group on statistical power, were tested. Simulations for an increased relative risk of 3.0 as well as a decreased relative risk of 1.5 for the placebo group were conducted at each level of loss to follow-up.

Secondary outcomes of interest included the probability of progressing from ILE to SLE over the 96-week period, the change in 52 soluble mediators over the 96-week period, and the change in 13 autoantibodies over the 96-week period. Power calculations were performed for each outcome, at each level of loss to follow-up. Additional simulations for the change in the 52 soluble mediators and 13 autoantibodies were performed using a Bonferroni adjustment for multiple comparisons. All power calculations for the secondary outcomes were performed using the SAS PROC POWER procedure.

Chapter 4. Results

Power Calculation for the Primary Outcome

Table 1. Power for Detecting Difference in Probability of Accumulating 1+ SLICC Score Criteria over 96-Week Period
Original Study Design

Placebo Probability	HCQ Probability	5 Study Sites (180 Subjects)		6 Study Sites (240 Subjects)	
		Power	95% CI	Power	95% CI
0.20	0.10	0.69	(0.66, 0.72)	0.82	(0.80, 0.84)
0.30	0.15	0.80	(0.78, 0.82)	0.90	(0.88, 0.92)
0.40	0.20	0.90	(0.88, 0.92)	0.97	(0.96, 0.98)
0.50	0.25	0.96	(0.95, 0.97)	0.99	(0.99. 0.99)
0.60	0.30	0.99	(0.98, 1.00)	1.00	(1.00, 1.00)

The original protocol assumed that 40% of placebo patients would accumulate at least 1 SLICC criteria over a 96-week period while 20% of patients treated with hydroxychloroquine would accumulate 1 or more SLICC criteria; this represents a relative risk (RR) of 2.0. With five study sites and 180 subjects enrolled and completing the trial, there is 90% statistical power to detect a statistically significant difference with a two-sided, 0.05 significance level test, given these assumptions. These calculations were made under the unlikely assumption that no subjects would be lost to follow-up.

With the addition of a 6th study site, the overall target enrollment was changed from 180 to 240 subjects. This was done so that up to 25% of subjects could be lost to follow-up; in this scenario, 180 subjects would remain in the study and the original 90% power calculation would remain valid. However, with six study sites and 240 subjects enrolled and completing the trial, there is 97% power to detect a statistically significant difference with a two-sided, 0.05 significance level test. With a 15% loss to follow-up, the power is 93%; at 20% loss, the power is 93%, and at 25% loss, the power is 90%.

All further power calculations for the primary outcome were conducted under the assumption of having 6 total study sites and a target subject enrollment of 240 subjects. Overall, as loss to follow-up increases, the power to detect a statistically significant difference between treatment groups decreases. Additionally, as the overall proportion of patients accumulating at least 1 SLICC criteria decreases (and necessarily the difference between groups becomes smaller), the power to detect a significant difference decreases. Additional tables of precise power calculation values are provided in the Appendix.

Figure 1. Statistical Power at Various Levels of Loss to Follow-Up
Relative Risk of 2.0
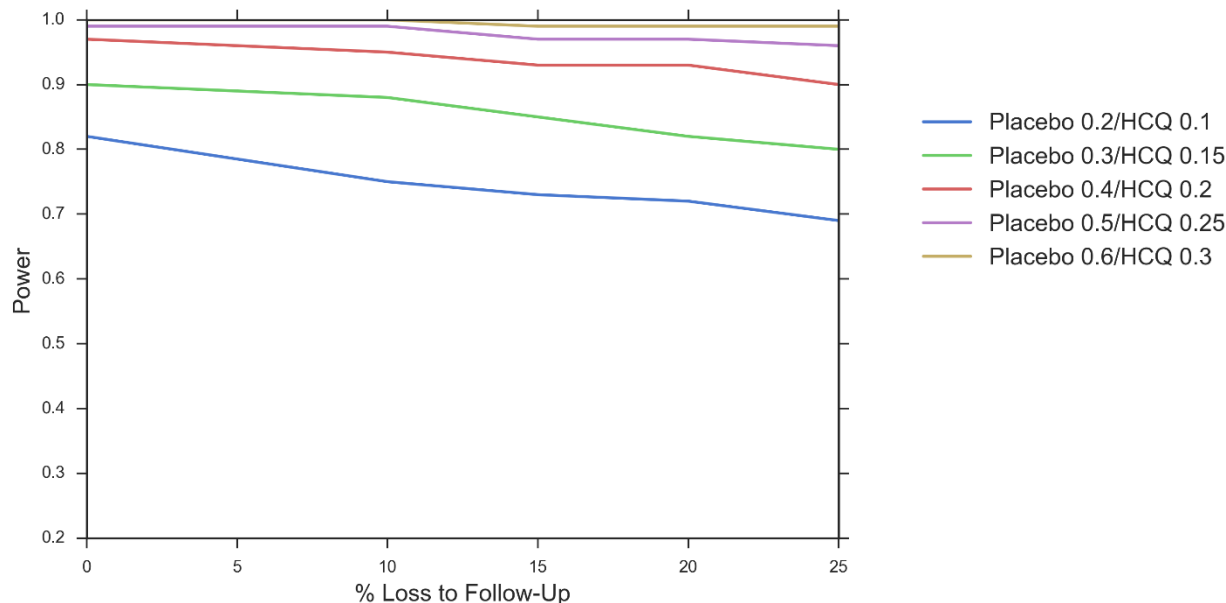
  With the risk ratio held constant at 2.0, should the proportion of subjects accumulating 1 or more SLICC criteria increase, the power to detect a statistical difference between groups quickly approaches 100%. This is true even with a large loss to follow-up. If 50% of placebo patients accumulate 1 or more criteria and 25% of patients receiving HCQ accumulate 1 or more SLICC criteria, then the power to detect a statistically significant difference using a two-sided, 0.05 level significance test is 96% when loss to follow-up is 25%. If 60% of placebo patients and 30% of HCQ-treated patients accumulate 1 or more SLICC criteria, even at 25% loss to follow-up, the power to detect a significant difference is 99%.

  However, if the proportion of patients accumulating 1 or more SLICC criteria is lower than expected, the power to detect a statistical difference between groups decreases. If 15% of patients overall (20% of the placebo group and 10% of the HCQ group) accumulate 1 or more SLICC criteria, at 25% loss to follow-up, the power to detect a statistically significant difference using a two-sided, 0.05 level significance test is 69%. A slight increase in the proportion of patients accumulating at least 1 SLICC criteria (30% of the placebo group and 10% of the HCQ group) at 25% loss to follow-up increases statistical power to 80%.

Figure 2. Statistical Power at Various Levels of Loss to Follow-Up
Relative Risk of 3.0
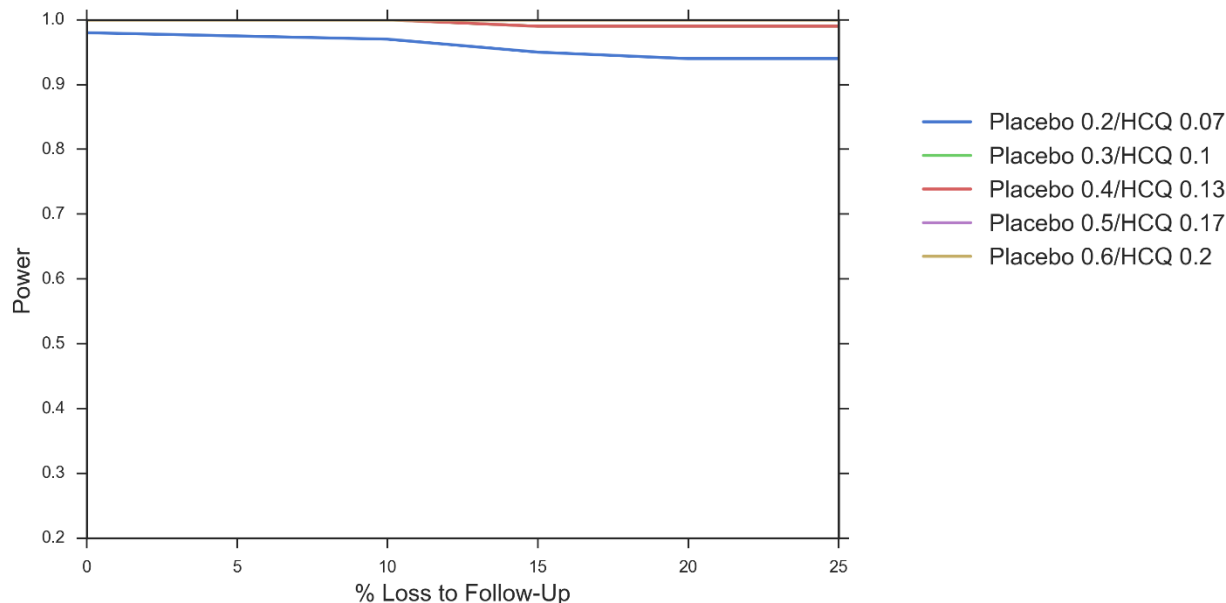
With a larger than expected risk ratio, the power to detect any significant difference between groups is greater than 90% regardless of loss to follow-up or the proportion of patients accumulating SLICC criteria within each group. Even with approximately 14% of patients accumulating 1 or more SLICC criteria (20% in the placebo group and 7% in the HCQ-treated group), the lowest overall proportion calculated, and 25% loss to follow-up, the power to detect a difference at a two-sided, 0.05 level of significance is 94%.

Figure 3. Statistical Power at Various Levels of Loss to Follow-Up
Relative Risk of 1.5
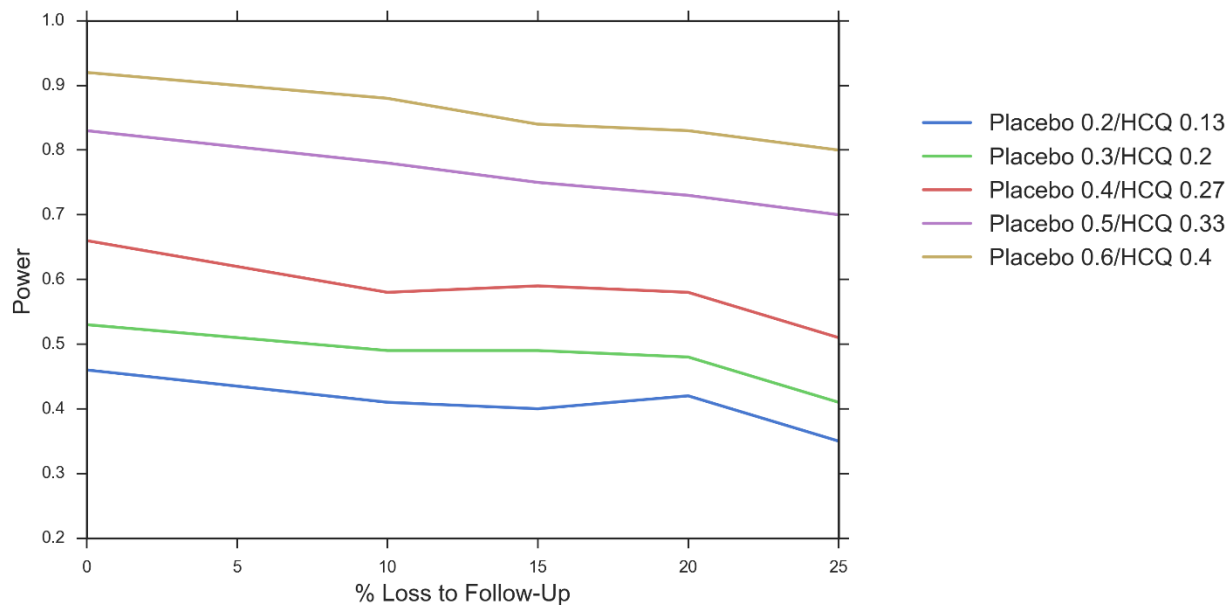
If the relative risk of accumulating 1 or more SLICC criteria is lower than expected for the placebo group, then the power to detect a significant difference is low. Using the original assumption that 40% of placebo patients will accumulate 1 or more SLICC criteria (and in this case 27% of patients in the treatment group accumulating any criteria), the power to detect a difference using a two-sided, 0.05 significance test is 66%. If 20% of placebo patients and 13% of treated patients accumulate any SLICC criteria, the power is 46%; if 30% of placebo patients and 20% of treated patients accumulate any SLICC criteria, the power is 53%; if 50% of placebo patients and 33% of treated patients accumulate any SLICC criteria, the power is 83%; if 60% of placebo patients and 40% of treated patients accumulate any SLICC criteria, the power is 92%. These were all calculated under the assumption that there will be no loss to follow-up. Higher levels of loss to follow-up lowered statistical power.

Power Calculation for the Secondary Outcomes

Table 5. Power for Detecting Difference in Probability of Progressing from ILE to SLE over 96-Week Period

Loss to Follow-Up	Old Protocol		New Protocol	
	Power	Remaining Subjects	Power	Remaining Subjects
0%	0.837	192	0.915	240
10%	0.789	173	0.881	216
15%	0.767	164	0.861	204
20%	0.736	154	0.831	192
25%	0.703	144	0.809	180

Assuming that 12% of patients treated with hydroxychloroquine progress to systemic lupus erythematosus (SLE) over a 96-week period, and that placebo patients are 2.5 times more likely to progress to SLE compared to the treatment group (30% of placebo patients progressing to SLE), then there is more than 80% power to detect a statistical difference with 180 patients completing the study; this represents a 25% loss to follow-up. This improves upon the previous protocol which only planned to enroll 192 patients total; with 25% loss to follow-up under this protocol, there is only 70% power to detect a statistical difference assuming the same proportion of subjects progress to systemic lupus diagnosis.

One measure commonly used to quantify the effectiveness of a drug is its number needed to treat (NNT). The NNT is a useful clinical measure and determines the average number of patients that need to be treated in order to prevent one from developing disease. It can be calculated as follows:

$$NNT = \frac{1}{p(c) - p(t)}$$

Where *p(c)* represents the proportion of control (or placebo) patients developing disease, and *p(t)* represents the proportion of treated patients progressing to disease. As NNT approaches 1 the effectiveness of the drug is said to increase, with NNT = 1 indicating perfect treatment. A larger NNT much greater than 1 indicates less effectiveness[19].

The number needed to treat (NNT) for this study is as follows:

$$NNT = \frac{1}{0.30 - 0.12}$$

$$NNT = 5.6$$

Therefore, for approximately every 6 subjects treated with hydroxychloroquine in this study, 1 will be prevented from developing systemic lupus.

Table 6. Change in Soluble Mediators and Autoantibodies over 96-Week Period

Loss to Follow-Up	Old Protocol		New Protocol	
	Power	Remaining Subjects	Power	Remaining Subjects
0%	0.834	192	0.906	240
10%	0.791	173	0.875	216
15%	0.772	164	0.856	204
20%	0.746	154	0.834	192
25%	0.717	144	0.809	180

A total of 13 autoantibodies and 52 soluble mediators are collected throughout the study and compared between those receiving hydroxychloroquine and those receiving placebo. The SMILE study has 80% statistical power to detect a mean difference of 0.425 standard deviation units between soluble mediators or autoantibodies with 180 subjects; this sample size is achievable even with 25% loss to follow-up over the course of the study. Before the protocol change, this study only had approximately 72% statistical power to detect the same mean difference at 25% loss to follow-up amongst subjects.

Because the number of soluble mediators and autoantibodies of interest is relatively large, resultant p-values must be adjusted for multiple comparisons. The Hochberg step-down procedure[20] will be used to determine the true effect size, which will be larger than the unadjusted estimate of 0.425 standard deviation units. A Bonferroni adjustment for multiple comparisons provides a conservative estimation of the Hochberg procedure and is shown below.

Table 7. Change in 52 Soluble Mediators over 96-Week Period (Bonferroni Adjusted)

Loss to Follow-Up	Old Protocol		New Protocol	
	Power	Remaining Subjects	Power	Remaining Subjects
0%	0.875	192	0.954	240
10%	0.817	173	0.923	216
15%	0.789	164	0.901	204
20%	0.749	154	0.875	192
25%	0.704	144	0.842	180

After correcting for multiple comparisons, a total of approximately 173 participants are needed to detect a mean difference of 0.65 standard deviation units at 80% statistical power. Under the new protocol design, this is achievable with a maximum 25% loss to follow-up; the old protocol design maintains at least 80% statistical power at just under 15% loss to follow-up.

Table 8. Change in 13 Autoantibodies over 96-Week Period (Bonferroni Adjusted)

Loss to Follow-Up	Old Protocol		New Protocol	
	Power	Remaining Subjects	Power	Remaining Subjects
0%	0.853	192	0.936	240
10%	0.798	173	0.902	216
15%	0.772	164	0.880	204
20%	0.735	154	0.853	192
25%	0.695	144	0.822	180

After correcting for multiple comparisons, just over 173 participants are needed to detect a difference of 0.575 standard deviation units at 80% power. This is achievable at a maximum 25% loss to follow-up under the new protocol design and at 10% loss to follow-up under the old protocol design.

Chapter 5. Discussion

The SMILE study has been appropriately designed to detect a difference in SLICC score accumulation between patients receiving hydroxychloroquine and patients receiving placebo. Initial power calculations showed that a target sample size of 180 patients would be sufficient to detect a statistical difference between the treatment and placebo groups if 40% of placebo patients and 20% of treated patients accumulated at least 1 SLICC criteria over the 96-week period.

Relative risk had an impact on statistical power. The initial assumption for the risk ratio for the SMILE trial was 2.0. Unsurprisingly, simulations using a larger risk ratio of 3.0 showed greater statistical power to detect a significant difference between SLICC score accumulation between the placebo and treatment groups. This is because statistical power to detect a difference between groups increases as the actual underlying difference between groups increases. Conversely, simulations assuming a lower relative risk ratio of 1.5 showed decreased statistical power to detect a significant difference between groups.

The overall proportion of patients accumulating 1 or more SLICC criteria also impacted statistical power. When relative risk was held constant, but the proportion of both patients receiving placebo or HCQ was low, the power to detect a statistically significant difference was decreased. Larger overall proportions of patients accumulating SLICC criteria increased statistical power; this effect helped to mitigate decreased statistical power estimates for the lower risk ratio of 1.5.

Sample size has a direct impact on statistical power and is under the control of the investigator, unlike the risk ratio and SLICC score accumulation. While increased sample size results in increased statistical power and protection against large loss to follow-up, it also results in increased cost. Therefore, study enrollment must be generous enough to allow for potential "significant" findings but small enough to remain financially feasible. A target sample size of 180 total subjects was determined to be adequate for the SMILE trial. However, concern for potential loss to follow-up resulted in an increase for the target recruitment of 240 subjects, or up to 25% loss to follow-up.

This is a much larger loss to follow-up margin than is typically allocated for in clinical trials[21]. However, cohorts of systemic lupus patients may exhibit unusually high loss to follow-up levels. In an outpatient clinic setting, 25.9% of SLE patients were lost to follow-up[22]. In the LUMINA cohort, an observational study, 29% of patients were lost to follow-up after 3.5 years, and 36% were lost to follow-up after 5 years[23]. Patients with higher disease activity[23], greater disease manifestations and lower socioeconomic status[24] were more likely lost to follow-up. Clinical trials had more reasonable loss to follow-up. In two European multi-site clinical trials, 7% of patients were lost to follow-up after 10 years in the Euro-Lupus Nephritis Trial[25] and 12% of patients were lost to follow-up after 10 years in the MAINTAIN trial[26].

In addition to maintaining follow-up throughout the study, successful completion of the SMILE trial is also dependent on SLICC score accumulation. It is important to consider the context in which SLICC criteria are used in this study, and its inherent relationship with the secondary outcome of progression to SLE. The SLICC criteria are used to classify individuals with systemic lupus; if a patient does not accrue any additional SLICC criteria, it is impossible for them to be classified with SLE based on this definition. Therefore, while the study is not well-powered to detect statistical differences with low overall levels of SLICC criteria accumulation, this may not detrimentally affect outcome. Given that the

study team has assumed that 30% of placebo patients will progress to SLE and 12% of patients treated with hydroxychloroquine will progress to SLE, the inherent assumption is that at least this proportion of patients (if not more) will accumulate 1 or more SLICC criteria. Of greater concern is the relative risk of accumulating 1 or more SLICC criteria between the treatment and placebo groups. If the risk of accumulating SLICC criteria is relatively higher in placebo patients compared to those treated with hydroxychloroquine, such as with a relative risk of 2.0 or higher, than the study is adequately powered at this sample size. However, if the RR is not as high comparatively, than the study is not adequately powered with this number of subjects.

Chapter 6. Conclusion

A target enrollment size of 240 subjects is appropriate for the SMILE trial. This sample size allows for relatively large loss to follow-up, as well as variation in the expected proportion of patients accumulating at least 1 SLICC criteria in both the hydroxychloroquine and placebo groups. Additional secondary outcomes of interest including risk of progressing from ILE to SLE, change in autoantibody profiles and change in soluble mediators are detectable at this sample size, using Hochberg's step-down procedure to adjust for multiple comparisons.

References

1. Bourn, R. & James, J. A. Pre-Clinical Lupus. *Curr. Opin. Rheumatol.* **27,** 433–439 (2015).

2. Pons-Estel, G. J. *et al.* Possible Protective Effect of Hydroxychloroquine on Retarding the Occurrence of Integument Damage in Lupus: Data from LUMINA, a Multiethnic Cohort. *Arthritis Care Res.* **62,** 393–400 (2010).

3. Fessler, B. J. *et al.* Systemic lupus erythematosus in three ethnic groups: XVI. Association of hydroxychloroquine use with reduced risk of damage accrual. *Arthritis Rheum.* **52,** 1473–1480 (2005).

4. Ruiz-Irastorza, G. & Khamashta, M. A. Hydroxychloroquine: the cornerstone of lupus therapy. *Lupus* **17,** 271–273 (2008).

5. Hallengren, C. S., Nived, O. & Sturfelt, G. Outcome of incomplete systemic lupus erythematosus after 10 years. *Lupus* **13,** 85–88 (2004).

6. Hochberg, M. C., Silman, A. J., Smolen, J. S., Weinblatt, M. E. & Weisman, M. H. Clinical features of systemic lupus erythematosus. in *Rheumatology* **2,** 1032–1044 (Elsevier Mosby, 2015).

7. Petri, M. *et al.* Derivation and Validation of Systemic Lupus International Collaborating Clinics Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheum.* **64,** 2677–2686 (2012).

8. Hochberg, M. C. Updating the American college of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40,** 1725–1725 (1997).

9. Olsen, N. J. *et al.* Autoantibody profiling to follow evolution of lupus syndromes. *Arthritis Res. Ther.* **14,** R174 (2012).

10. Kavanaugh, A., Russell, T., Reveille, J., Solomon, D. H. & Homburger, H. Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. *Arch. Pathol. Lab. Med. Northfield* **124,** 71–81 (2000).

11. Pisetsky, D. S. Antinuclear antibodies in healthy people: the tip of autoimmunity's iceberg? *Arthritis Res. Ther.* **13,** 109 (2011).

12. Wandstrat, A. E. *et al.* Autoantibody profiling to identify individuals at risk for systemic lupus erythematosus. *J. Autoimmun.* **27,** 153–160 (2006).

13. Dubois, E. L. Systemic lupus erythematosus: recent advances in its diagnosis and treatment. *Ann. Intern. Med.* **45,** 163–184 (1956).

14. Dubois, E. L. Antimalarials in the management of discoid and systemic lupus erythematosus. *Semin. Arthritis Rheum.* **8,** 33–51 (1978).

15. The Canadian Hydroxychloroquine Study Group. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. *N. Engl. J. Med.* **324,** 150–154

16. Alarcón, G. S. *et al.* Effect of hydroxychloroquine on the survival of patients with systemic lupus erythematosus: data from LUMINA, a multiethnic US cohort (LUMINA L). *Ann. Rheum. Dis.* **66,** 1168–1172 (2007).

17. James, J. *et al.* Hydroxychloroquine sulfate treatment is associated with later onset of systemic lupus erythematosus. *Lupus* **16,** 401–409 (2007).

18. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9,** 90–95 (2007).

19. Cook, R. J. & Sackett, D. L. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* **310,** 452–454 (1995).

20. Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–802 (1988).

21. Fewtrell, M. S. *et al.* How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Arch. Dis. Child.* **93,** 458–461 (2008).

22. Tien, Y.-C., Chiu, Y.-M. & Liu, M.-P. Frequency of Lost to Follow-Up and Associated Factors for Patients with Rheumatic Diseases. *PLoS ONE* **11,** (2016).

23. Bertoli, A. M. *et al.* Systemic lupus erythematosus in a multiethnic US cohort (LUMINA) XXXI: factors associated with patients being lost to follow-up. *Lupus* **15,** 19–25 (2006).

24. Uribe, A. G. *et al.* Relationship between adherence to study and clinic visits in systemic lupus erythematosus patients: data from the LUMINA cohort. *Lupus* **13,** 561–568 (2004).

25. Houssiau, F. A. *et al.* The 10-year follow-up data of the Euro-Lupus Nephritis Trial comparing low-dose and high-dose intravenous cyclophosphamide. *Ann. Rheum. Dis.* **69,** 61–64 (2010).

26. Tamirou, F. *et al.* Long-term follow-up of the MAINTAIN Nephritis Trial, comparing azathioprine and mycophenolate mofetil as maintenance therapy of lupus nephritis. *Ann. Rheum. Dis.* **75,** 526–531 (2016).

Appendix

Table 2. Power for Detecting Difference in Probability of Accumulating 1+ SLICC Score Criteria over 96-Week Period with a Relative Risk of 2.0 Accounting for Loss to Follow-Up

Placebo Prob.	HCQ Prob.	240 Subjects (0% Loss)		216 Subjects (10% Loss)		204 Subjects (15% Loss)		192 Subjects (20% Loss)		180 Subjects (25% Loss)	
		Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI
0.20	0.10	0.82	(0.80, 0.84)	0.75	(0.72, 0.78)	0.73	(0.70, 0.76)	0.72	(0.69, 0.75)	0.69	(0.66, 0.72)
0.30	0.15	0.90	(0.88, 0.92)	0.88	(0.86, 0.90)	0.85	(0.83, 0.87)	0.82	(0.80, 0.84)	0.80	(0.78, 0.82)
0.40	0.20	0.97	(0.96, 0.98)	0.95	(0.94, 0.96)	0.93	(0.91, 0.95)	0.93	(0.91. 0.95)	0.90	(0.88, 0.92)
0.50	0.25	0.99	(0.99, 0.99)	0.99	(0.98, 1.00)	0.97	(0.96, 0.98)	0.97	(0.96, 0.98)	0.96	(0.95, 0.97)
0.60	0.30	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	0.99	(0.98, 1.00)	0.99	(0.98, 1.00)	0.99	(0.98, 1.00)

Table 3. Power for Detecting Difference in Probability of Accumulating 1+ SLICC Score Criteria over 96-Week Period with a Relative Risk of 3.0 Accounting for Loss to Follow-Up

Placebo Prob.	HCQ Prob.	240 Subjects (0% Loss)		216 Subjects (10% Loss)		204 Subjects (15% Loss)		192 Subjects (20% Loss)		180 Subjects (25% Loss)	
		Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI
0.20	0.07	0.98	(0.97, 0.99)	0.97	(0.96, 0.98)	0.95	(0.94, 0.96)	0.94	(0.93, 0.95)	0.94	(0.93, 0.95)
0.30	0.10	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	0.99	(0.98, 1.00)	0.99	(0.99, 0.99)	0.99	(0.99, 0.99)
0.40	0.13	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)
0.50	0.17	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)
0.60	0.20	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)	1.00	(1.00, 1.00)

Table 4. Power for Detecting Difference in Probability of Accumulating 1+ SLICC Score
Criteria over 96-Week Period with a Relative Risk of 1.5
Accounting for Loss to Follow-Up

Placebo Prob.	HCQ Prob.	240 Subjects (0% Loss)		216 Subjects (10% Loss)		204 Subjects (15% Loss)		192 Subjects (20% Loss)		180 Subjects (25% Loss)	
		Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI	Power	95% CI
0.20	0.13	0.46	(0.43, 0.49)	0.41	(0.38, 0.44)	0.40	(0.37, 0.43)	0.42	(0.39, 0.45)	0.35	(0.32, 0.38)
0.30	0.20	0.53	(0.50, 0.56)	0.49	(0.46, 0.52)	0.49	(0.46, 0.52)	0.48	(0.45, 0.51)	0.41	(0.38, 0.44)
0.40	0.27	0.66	(0.63, 0.69)	0.58	(0.55, 0.61)	0.59	(0.56, 0.62)	0.58	(0.55, 0.61)	0.51	(0.48, 0.54)
0.50	0.33	0.83	(0.81, 0.85)	0.78	(0.75, 0.81)	0.75	(0.72, 0.78)	0.73	(0.70, 0.76)	0.70	(0.67, 0.73)
0.60	0.40	0.92	(0.90, 0.94)	0.88	(0.86, 0.90)	0.84	(0.82, 0.86)	0.83	(0.81, 0.85)	0.80	(0.78, 0.82)