CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

204017Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	204017 / SN58
Drug Name:	Twirla® (levonorgestrel/ethinyl estradiol 120/30 mcg/day) transdermal contraceptive delivery system
Indication(s):	Prevention of Pregnancy
Applicant:	Agile Therapeutics, Inc.
Date(s):	Submission Date: 5/16/2019
	PDUFA Due Date: 02/20/2020
Review Priority:	Class II Resubmission
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Yun Tang, Ph.D.
Concurring Reviewers:	Mahboob Sobhan, Ph.D., Acting Statistical Team Leader, Division of Biometrics III
	Laura Lee Johnson, Ph.D., Division Director, Division of Biometrics III
Medical Division:	Division of Bone, Reproductive and Urologic Drug Products
Clinical Team:	Nneka McNeal-Jackson, M.D., Medical Reviewer
	Gerald D. Willett, M.D., Clinical Team Leader
Project Manager:	Jeannie Roule

Keywords: NDA review, clinical study, prevention of pregnancy

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The Applicant, Agile Therapeutics, Inc., is seeking FDA approval of Twirla® (hereafter referred to as AG200-15), for the third time after receiving two Complete Response (CR) letters in 2013 and 2017. AG200-15 is an ethinyl estradiol (EE) and levonorgestrel (LNG) transdermal system (TDS) intended for the prevention of pregnancy in women of reproductive age. The original NDA was submitted in 2012 with data from two Phase 3 trials, Study ATI-CL12 (hereafter referred to as Study 12) and Study ATI-CL13 (hereafter referred to as Study 13), and FDA issued a CR letter citing clinical and study conduct deficiencies among others. Following FDA's recommendations to provide further evidence of efficacy with a new study, the Applicant resubmitted their NDA in 2017 with data from a third Phase 3 study (Study ATI-CL23; hereafter referred to as Study 23). A second CR letter was issued for product quality and an unacceptable manufacturing site. In that CR letter, the Division also mentioned that there were concerns about reduced effectiveness of the product, but it was unclear whether these deficiencies were related to the adhesive properties of the transdermal system.

In 2018, the Applicant submitted Formal Dispute Resolution Requests to FDA's Office of Drug Evaluation III and the Office of New Drugs (OND); in both cases their appeals were denied. The basis for the OND denial was the lack of clinical data that demonstrates acceptable adhesion properties of the TDS.

In 2019, the Applicant resubmitted the NDA with new clinical data to resolve adhesion deficiencies noted in the 2017 CR letter, but without additional clinical efficacy data to resolve reduced efficacy concerns. In this resubmission, the Applicant reiterated that AG200-15 is effective for preventing pregnancy in women of reproductive age based on Study 23 with a Limitation of Use (LOU). Specifically, the Applicant is proposing an indication for the overall reproductive age population with a LOU for obese women (Body Mass Index (BMI) ≥ 30 kg/m^2).

As noted in the 2017 CR letter, AG200-15 demonstrated reduced overall effectiveness based on data from Study 23. This is based on the Division's advice[1] to Agile in October 2013 that the Division has never approved a combined hormonal contraceptive (CHC) with an upper bound of the 95% confidence interval (CI) for the Pearl Index that exceeds 5.

The review of Study 23 in the current review cycle noted AG200-15's worsened efficacy with increasing BMI. The effectiveness of AG200-15 overall and in each of the BMI subgroups did not meet the Division's criteria of an upper 95% CI bound of 5, even in normal weight women (Pearl Index of 3.5 with an upper bound of 5.2).

The highlights of efficacy and safety findings pertaining to this resubmission are:

[1] See meeting minutes dated on October 10, 2013.

- AG200-15 demonstrated reduced effectiveness in both non-obese (BMI < 30 kg/m^2) and obese women (BMI ≥ 30 kg/m^2): Pearl Indices of 4.34 (95% CI: 2.86 to 5.82) for non-obese and 8.64 (95% CI: 5.79 to 11.50) for obese women, respectively.

- There was a trend towards an increasing Pearl Index with increasing BMI (details in section 4): Pearl Indices of 3.46 (95% CI: 1.77 to 5.16) for normal weight women (BMI < 25 kg/m^2), 5.69 (95% CI: 2.99 to 8.40) for overweight women (BMI ≥ 25 to < 30 kg/m^2), and 8.64 (95% CI: 5.79 to 11.50) for obese women (BMI ≥ 30 kg/m^2), respectively.

- The estimated incidence rate of venous thromboembolic events (VTEs) for AG200-15 was recalculated during this review cycle and was approximately 28 per 10,000 women-years in Study 23. This was higher than any other approved CHCs approved in the past 10 years.

There remains clinical uncertainty whether AG200-15's effectiveness in real world use could potentially be worse than seen in Study 23. In Study 12, where pregnancy information was not adequately captured, the Pearl Index for women with BMI < 32 kg/m^2 was much higher (7.5 (95% CI: 5.0 to 10.0))[2]. If this study had adequately captured pregnancies, it could have revealed a pregnancy rate that was equivalent to a non-hormonal contraceptive in normal weight women. This is unacceptable for a combined hormonal contraceptive with risks of thromboembolism.

At this time in the review cycle, a remaining key issue is whether AG200-15 can be approved for a subgroup of women based on subgroup analyses by BMI in Study 23 when overall effectiveness was considered deficient in both the previous and current review cycle. We recommend that AG200-15 is studied further to reduce the uncertainty around the effectiveness and safety. Therefore, we do not recommend approval based on the lack of new clinical efficacy data.

[2] Statistical review of NDA204017 SN0000.

2 INTRODUCTION

2.1 Overview

Application History

The Applicant, Agile Therapeutics, Inc., is seeking FDA approval of AG200-15, a TDS containing EE and LNG for the prevention of pregnancy in women of reproductive age. AG200-15 is 28 cm^2 matrix type TDS designed to deliver 120 mcg of LNG and 30 mcg of EE per day. The proposed dosing regimen for one cycle of AG200-15 is one TDS applied to either the abdomen, buttock, or upper torso every seven days for three consecutive weeks followed by one TDS-free week.

AG200-15 is undergoing its third NDA review cycle by the FDA. The Applicant submitted the original NDA on April 12, 2012. The original NDA submission was based on the safety and efficacy data from Study 12 and Study 13, each of which compared AG200-15 to an approved oral CHC. The following issues were identified during the first review cycle:

- The two Phase 3 trials failed to demonstrate acceptable evidence of efficacy (the Pearl Index in the larger 13-cycle study (Study 12) was 7.50 with an upper bound of the 95% CI of 9.97. The Pearl Index in the smaller 6-cycle study (Study 13) was 8.19 with an upper bound of the 95% CI of 16.19)[3].

- There were significant problems with study conduct (e.g., high rates of premature trial discontinuations and lost to follow-up), as well as product quality.

The FDA issued a CR letter on February 13, 2013 recommending that the Applicant address the deficiencies outlined above by conducting a new Phase 3 trial to demonstrate efficacy of AG200-15. In response to the 2013 CR letter, the Applicant resubmitted the NDA in June 2017 with data from a new Phase 3 study (Study 23). The FDA issued a second CR letter on December 21, 2017 due to the following concerns:

- Deficiencies in adhesion and manufacturing quality.

- Concerns whether the reduced effectiveness of AG200-15 were related to adhesion properties of the product.

In June 2018 and again in August 2018, the Applicant submitted Formal Dispute Resolution Requests to FDA's Office of Drug Evaluation III and the Office of New Drugs, respectively. In both cases the FDA denied these appeals. Regarding the noted adhesion deficiencies, the FDA recommended that the Applicant conduct a head-to-head comparative clinical non-inferiority

[3] Study 13 was not considered a pivotal efficacy Phase 3 study (see statistical review of NDA204017 SN0000); it was supportive for safety purposes.

wear study against Xulane, the only marketed contraceptive TDS with adequate adhesion, to demonstrate adequate adhesion of AG200-15.

The Applicant resubmitted the NDA for AG200-15 in May 2019 including clinical data that addressed our concerns on the adhesion noted in the 2017 CR letter. Refer to Office of Product Quality review dated 14-Aug-19 for details of the adhesion assessment of AG200-15.

The current resubmission did not contain any new efficacy data. Therefore, the efficacy of AG200-15 in the current review cycle is evaluated based on Study 23 and Study 12 (see section below). In addition to the effectiveness of AG200-15 in the overall population, this review paid attention to the effectiveness of AG200-15 in different BMI subgroups and the incidence rate of VTE in Study 23.

Key design features of Study 23 are outlined in Table 1.

Table 1: Summary of Key Features of Study 23

Study Number (No. of Sites / Country) Dates of Study Conduct	Subject Population	Treatments	Sample Size (Enrolled /Safety)	Duration of Treatment	Design[1]
ATI-CL23 (102 / U.S.) 9-23-14 to 11-02-16	Heterosexually active, 18-40 years old females who were at risk of pregnancy with no weight or BMI restriction	AG200-15 (3 weekly TDS, 7 days TDS-free)	2,032/2,031	Up to 13 cycles	SA, OL, MC

[1]SA = Single Arm, OL = Open Label, MC = Multicenter

Approach to the Review

The focus of this review is on providing labeling that will allow effective and safe use of AG200-15. The Applicant proposes a LOU claim and subgroup analyses of effectiveness by BMI in Study 23, including modeling based on recommendations by the Bone, Reproductive, and Urologic Advisory Committee (BRUDAC) meeting on October 30, 2019.

This review will address the following outstanding questions:

- Whether labeling adequately describes the effectiveness of AG200-15;
- Whether exploratory modeling can be used to assist providers and patients in making decisions on whether to use AG200-15;
- A corrected VTE incidence rate associated with use of AG200-15.

Although there were significant study conduct issues with Study 12, effectiveness data from Studies 12 and 23 will be discussed in this review to obtain the statistical perspective on the results.

2.2 Data Sources

The study reports and the datasets for Study 23 were submitted electronically to the Electronic Document Room. The SAS datasets were complete and well documented.

The study datasets and SAS programs for Study 23 are located at:

\\CDSESUB1\evsprod\NDA204017\0038\m5\datasets\ati-cl23\analysis\adam
The original study reports for Study 23 submitted in the 2017 review cycle are located at:

\\CDSESUB1\evsprod\NDA204017\0038\m5\53-clin-stud-rep\535-rep-effic-safety-stud\contraception\5351-stud-rep-contr\ati-cl23

The study report addendum for Study 23 submitted in the current review cycle are located at:

\\CDSESUB1\evsprod\NDA204017\0058\m5\53-clin-stud-rep\535-rep-effic-safety-stud\contraception\5351-stud-rep-contr\ati-cl23

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design

Study 23

Study 23 was a U.S. only, single-arm, open-label, 13-cycle, multicenter Phase 3 clinical trial evaluating the contraceptive efficacy, safety, and tolerability of AG200-15. A total of 2,032 U.S. women aged 18 years or older at risk for pregnancy and desiring to use contraception were enrolled at 102 investigational sites. Per the Statistical Analysis Plan, the sample size was determined as follows:

"The Study is sized to provide 90% power to establish that if the underlying Pearl Index is no larger than 3.5, the Pearl Index will have an upper limit of a two-sided 95% confidence interval no larger than 5. Assuming that each of 1900 enrolled subjects 18 to ≤ 35 years of age will on average provide 8.5 cycles on treatment, the study will generate approximately 16,000 cycles of exposure to AG200-15 in this age group. The calculations also assume that roughly 21% of these cycles will not be included in the primary evaluation of efficacy because of use of back-up contraception or absence of sexual activity. Thus, a total of approximately 12,675 cycles was used in the power calculations." (pages 9-10)

In order to maximize the compliance for study participation of Study 23's population, the study protocol specified the following inclusion and exclusion criteria:

Inclusion criterion:

- A subject should demonstrate at least 90% compliance with electronic Diary (eDiary) entry and return two check-in phone calls from the investigator site during the two-week Run-In Period.

Exclusion criterion:

- A subject should not participate in the study if the investigator deemed the subject might have poor compliance with the study protocol and procedures.

For information related to study visits and use of eDiary in Study 23, refer to Section 3.1.1.1 in the statistical review of NDA204017 SN0038.

Study 12

Study 12 was the pivotal Phase 3 efficacy study contained in the original NDA submission to support the approval of AG200-15. It was a U.S. only, multicenter, open-label, randomized study

conducted in healthy, sexually active women requesting contraception. As noted in the 2013 CR letter (and as previously stated) Study 12 had substantial problems with study conduct, including low completion rates and issues with subject follow-up of pregnancy data and overall data collection. Therefore, there remains concerns with the Study 12's data quality and results.

A comparator product was included in Study 12. Because of the study conduct and the limited number of subjects who received the comparator, data from the comparator will not be discussed.

Refer to the statistical and clinical reviews of NDA204017 SN0000 for details of Study 12.

3.1.2 Subject Populations and Analysis Datasets (Study 23)

The following analysis populations were defined in the protocol.

Safety population: All subjects who wore at least one TDS for the study period.

Contraceptive efficacy population (CEP): All subjects who wore at least one TDS and were documented to have a negative enrollment serum ß-human chorionic gonadotropin (β-hCG).

Intent-to-treat (ITT) efficacy dataset: All complete or incomplete cycles in which intercourse occurred and no backup contraception was used. A cycle was defined as a 28-day period consisting of 21 days on treatment (consecutive administration of three 7-day TDSs) followed by 7 days off treatment (the TDS-free week).

3.1.3 Study Endpoints and Statistical Methodologies (Study 23)

The primary efficacy endpoint defined by the Applicant was the pregnancy rate measured by the Pearl Index in women who were ≤ 35 years of age at study entry irrespective of BMI. The Pearl Index, defined as the pregnancy rate per 100 women-years of drug exposure is calculated as follows:

$$\text{Pearl Index} = \frac{\text{number of on-treatment pregnancies x 13 cycles}}{\text{number of evaluable cycles}} \text{ x } 100$$

where

- "[O]n-treatment pregnancy" was defined as the pregnancy with an estimated date of conception between the date of first AG200-15 system application through 7 days after the last system removal.

- "[E]valuable" cycles were defined as the complete or incomplete 28-day cycles in which vaginal intercourse occurred and no back-up contraception was used based on eDiary data.

The pre-specified secondary efficacy endpoints included the Pearl Indices by BMI (<30 kg/m^2, <25 kg/m^2, ≥25 to <30 kg/m^2, and ≥30 kg/m^2), self-identified race (White, Black, and Other), and ethnicity (Hispanic or Latino and Not Hispanic or Latino).

Life table analysis was conducted as a supportive efficacy analysis to estimate the cumulative pregnancy rate through cycle 13 in women ≤ 35 years of age. In this review, life table analyses were performed on both the CEP (i.e., no exclusion of cycles for lack of vaginal intercourse and use of back-up contraception) and ITT (i.e., exclusion of cycles for lack of vaginal intercourse and use of back-up contraception) populations. For each population, the life table analyses were conducted on overall population and by subgroups of BMI, self-identified race, and ethnicity.

In addition to point estimates of the Pearl Index and the life table cumulative probabilities of pregnancy, two-sided 95% CIs are provided.

In the current review, the primary efficacy evaluation was based on the ITT population and the supportive efficacy evaluation was based on the ITT population and CEP. Patient disposition, demographics, baseline characteristics, and safety evaluation were based on the Safety Population.

3.1.4 Subject Disposition, Demographic and Baseline Characteristics (Study 23)

Subject Disposition

Table 2 summarizes the subject disposition information in Study 23, including reasons for study discontinuation. The Safety Population included 2,031 women. Of these 2,031 women, 1,042 (51.3%) subjects dropped out of the study prematurely. The top three reasons for study discontinuation were “subject’s decision” (15.3%), “lost to follow-up” (11.2%), and “adverse event” (10.9%).

The Applicant’s results presented in Table 2 were obtained based on their classifications of reasons for trial discontinuation. However, after reviewing the Applicant’s Information Request responses dated December 18, 2019, the clinical reviewer did not agree with the Applicant’s classifications for the following two subjects:

- Subject (b) (6): the reason for trial discontinuation for this subject was classified as “Lost to follow-up” by the Applicant despite her positive pregnancy result. The clinical reviewer adjudicated that this subject was discontinued due to “Pregnancy” since her pregnancy was confirmed by ultrasound.

- Subject (b) (6): the reason for trial discontinuation for this subject was classified as “Pregnancy” even though her pregnancy was confirmed negative by ultrasound and two serum pregnancy tests. After reviewing the subject’s narratives, the clinical reviewer

believed that subject decision (i.e., withdrawal of consent) may have been a more appropriate classification.

Note that the primary efficacy evaluation was not changed by the Applicant's misclassifications of the above two subjects' reasons for trial discontinuation. The Applicant correctly deemed the pregnancy from subject (b) (6) as on-treatment pregnancy and included it in their efficacy evaluation. Subject (b) (6) was not included in the primary efficacy evaluation because she was over 35 years of age.

Table 2. Subject Disposition: Study 23

Category	Safety Population (N=2,031) n (%)	
Completed the study	989 (48.7)	
Discontinued the study	1,042 (51.3)	
	Applicant	**FDA**
Reason for discontinuation		
Adverse event	222 (10.9)	222 (10.9)
Non-compliance	116 (5.7)	116 (5.7)
Lost to follow-up	**229 (11.3)**	**228 (11.2)**
Subject's decision	**310 (15.3)**	**311 (15.3)**
Pregnancy	73 (3.6)	73 (3.6)
Protocol violation	14 (0.7)	14 (0.7)
Investigator decision	17 (0.8)	17 (0.8)
Sponsor decision	18 (0.9)	18 (0.9)
Sponsor decision (study termination)	2 (0.1)	2 (0.1)
Other reasons	41 (2.0)	41 (2.0)

Note. Denominator for % calculation is the number of subjects in the Safety Population (N=2,031).
Source: Clinical Study Report and Reviewer's Analysis

Study 23 subjects withdrew rapidly starting at cycle 2. For more details on subject disposition by treatment cycle, refer to Section 3.1.4 in the statistical review of NDA 204017 SN0038.

The 51.3% subject withdrawal rate is consistent with recent contraceptive trials submitted to the Division. For the comparison of dropout rates for Study 23 relative to those for Study 12 and registration trials for oral CHCs approved between 2007 and 2017, refer to Section 3.1.4 in the statistical review of NDA 204017 SN0038.

Demographics and Baseline Characteristics

The mean BMI in the safety population in Study 23 was 28.3 kg/m^2. As shown in Table 3 below, Study 23 included 35.3% obese women (BMI $\geq$ 30 kg/m^2), 39.4% normal weight women (BMI < 25 kg/m^2), and 25.3 % overweight women (BMI $\geq$ 25 kg/m^2 to < 30 kg/m^2).

Table 3. BMI Characteristics: Study 23

BMI[1] (kg/m^2)	Safety Population (N=2,031) n (%)
<30 (Non-obese)	1313 (64.7)
<25 (Normal)	800 (39.4)
≥25 to <30 (Overweight)	513 (25.3)
≥30 (Obese)	717 (35.3)

[1]BMI subpopulations (Normal, Overweight and Obese) add up to N = 2,030: Subject (b) (6) had no BMI information.
Note. Denominator for % calculation is the number of subjects in the Safety Population (N=2,031).
Source: Clinical Study Report Addendum and Reviewer's Analysis

For information on other demographics and baseline characteristics, refer to Section 3.1.4 in the statistical review of NDA 204017 SN0038.

3.1.5 Results and Conclusions (Study 23)

Table 4 presents an overview of efficacy populations for Study 23. Of the 2,031 women in the Safety Population who wore at least one TDS for the study period, 2,024 had a negative enrollment serum β-hCG and were included in the CEP. Of the women in the CEP, 1,932 women contributed at least one evaluable cycle in which vaginal intercourse occurred and no backup contraception was used. These 1,932 women comprise the ITT efficacy dataset. Of the women in the ITT dataset, 1,736 were ≤ 35 years at baseline and comprise the *primary analysis population* for effectiveness, namely, women who were 35 years of age or younger, wore at least one AG200-15 system, had a negative enrollment serum β-hCG pregnancy test, and had at least one evaluable cycle.

Table 4: Overview of Efficacy Populations: Study 23

Population	Subjects (N = 2,031) (%)
Contraceptive Efficacy Population (CEP)	2,024 (99.7)
ITT	1,932 (95.1)
Primary Analysis Population (Age ≤ 35 in ITT)	1,736 (85.5)

Note. Denominator for % calculation is the number of subjects in the Safety Population (N=2,031).
Source: Clinical Study Report Addendum and Reviewer's Analysis

The primary efficacy endpoint was the pregnancy rate estimated by the Pearl Index in subjects less than 35 years old.

For the remainder of the review, "ITT population" and "CEP" will refer to data from women ≤ 35 years at baseline unless otherwise noted.

Table 5 presents primary efficacy results based on the ITT population. In 15,165 evaluable cycles of AG200-15 use, FDA identified 68 on-treatment pregnancies. As shown in Table 5, the estimated overall Pearl Index in women less than or equal to 35 years old was 5.83 (95% CI: 4.45 to 7.21).

Table 5. Pearl Index in Subjects ≤ 35 Years of Age (ITT): Study 23

N	Number of On-Treatment Pregnancies	Number of Evaluable Cycles	Pearl Index (95% CI)
1,736	68	15,165	5.83 (4.45, 7.21)

Source: Clinical Study Report Addendum and Reviewer's Analysis

Similar results were observed using life table analysis based on the ITT population (Table 6) and CEP (Table 7). There were some slight numerical differences in the life table results based on CEP reported by the Applicant and the FDA. As shown in Table 7, the Applicant reports an estimated cumulative pregnancy rate of 5.29 (95% CI: 4.17 to 6.70), while the Agency's estimated cumulative pregnancy rate was 5.32 (95% CI: 4.20 to 6.74).

Table 6. Cumulative Pregnancy Rate in Subjects ≤ 35 Years of Age (ITT): Study 23

	N	Number of On - Treatment Pregnancies	Number of Cycles[1]	Cumulative Pregnancy Rate (95% CI)
FDA	1,736	68	15,165	5.48 (4.32, 6.04)

Source: Reviewer's Analysis

Table 7. Cumulative Pregnancy Rate in Subjects ≤ 35 Years of Age (CEP): Study 23

	N	Number of On - Treatment Pregnancies	Number of Cycles[1]	Cumulative Pregnancy Rate (95% CI)
Applicant[1]	1,816	68	-	5.29 (4.17, 6.70)
FDA	1,816	68	16,330	5.32 (4.20, 6.74)

[1]The Applicant did not provide the number of cycles used in their life table analysis in their submissions.
Source: Clinical Study Report Addendum and Reviewer's Analysis

3.2 Evaluation of Safety (Study 23)

Thromboembolism, and more specifically VTE is the serious adverse event of greatest concern with CHC use. Given the potential lethality and morbidities associated with VTEs, this section focused on the evaluation of the VTE risk with the use of AG200-15 as requested by the clinical review team. For the complete and detailed evaluation of safety, refer to the clinical review of NDA 204017 SN0058.

Across Studies 12, 13, and 23, VTE events (either deep venous thrombosis (DVT) or pulmonary embolism (PE)) were reported in six women: three women had a DVT, two women had a PE, and one woman had both a DVT and a PE. By study, there were six VTE events (3 DVT and 3 PE) in Study 23, one in Study 12, and zero in Study 13 (Table 8). Of all seven observed VTE events, only six from five subjects (one subject enrolled in Study 12 and four subjects enrolled in Study 23) were considered probably drug-related. The DVT experienced by subject (b) (6) in

Study 23 was considered not related to the study drug. Both the Applicant and the Agency excluded the subject (b) (6) from the calculation of the drug-related VTE incidence rate. *In Study 23, all five subjects with VTEs had a BMI > 30 kg/m^2.*

Table 8. A Summary of Subjects with VTEs in Studies 12, 13, and 23

Subjects	Age	BMI (kg/m^2)	VTE Event(s)	Relatedness to Drug
Study 12				
Subject (b) (6)	26	19.9	DVT	Probably related
Study 23				
Subject (b) (6)	25	31.8	PE	Probably related
Subject	26	34.3	PE	Probably related
Subject	24	35.7	DVT	Not related
Subject	33	36.3	DVT/PE	Probably related
Subject	35	39.0	DVT	Probably related

Source: FDA and the Applicant's Backgrounders for the 2019 BRUDAC Meeting / Clinical Review of NDA 204017 SN0000

Table 9 presents the estimated incidence rates reported by the Applicant and the FDA. As shown, the Applicant reported an estimated drug-related VTE incidence rate of 22 per 10,000 woman-years by pooling safety data from Studies 12, 13, and 23. The Division did not agree with the Applicant's pooling of VTE data from all three studies because the Division had concerns about the quality of data collection in Studies 12 and 13. Since the VTE events may have been underreported in Studies 12 and 13, pooling data from all three studies may underestimate the actual risk of VTE. Therefore, the Division reported an estimated drug-related VTE incidence rate of 28 (95% CI: 8 to 71) per 10,000 woman-years based solely on Study 23. Given the small number of subjects who experienced VTEs in Study 23, there remains considerable uncertainty about the magnitude of VTE risk with AG200-15 as shown by the wide 95% CI. Nevertheless, four subjects in one trial of this size and duration experiencing VTEs is concerning to the Division and represents a safety signal, particularly in obese women as these four subjects with VTEs had a BMI > 30 kg/m^2.

Table 9. VTE Incidence Rate with the Use of AG200-15

	Studies Included in Analysis	Number of Subjects with VTEs	Number of Cycles	VTE Incidence Rate (95% CI[1]) (Per 10,000 Woman-Years)
Applicant	12, 13, 23	5	29,900	22 (7, 51)[2]
FDA	23	4	18,841	28 (8, 71)

[1]The VTE incidence was assumed to follow a Poisson distribution.
[2]The Applicant did not provide the 95% CI of 7 to 51, which was calculated by the FDA reviewer.
Source: Clinical Overview and Reviewer's Analysis

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

This review first evaluated the effectiveness by BMI subgroups in Studies 23 and 12. Additional exploratory analyses to explore the relationship between pregnancy rate and continuous BMI were conducted as recommended by some advisory committee members at the 2019 BRUDAC[4] meeting.

Table 10 presents the Pearl Indices based on subgroup analyses of women ≤ 35 years of age in the ITT population for Study 23. As shown, the estimated AG200-15 pregnancy rate was almost doubled in the obese subgroup (BMI ≥ 30 kg/m^2) compared to the non-obese subgroup (BMI < 30 kg/m^2): the estimated Pearl Indices were 4.34 (95% CI: 2.86 to 5.82) for non-obese women and 8.64 (95% CI: 5.79 to 11.50) for obese women, respectively. The estimated Pearl Indices for normal weight (BMI < 25 kg/m^2) and overweight (BMI ≥25 kg/m^2 to <30 kg/m^2) women were 3.46 (95% CI: 1.77 to 5.16) and 5.69 (95% CI: 2.99 to 8.40), respectively.

Table 10. Pearl Index in Subjects ≤ 35 Years of Age by Subgroup (ITT): Study 23

Population	N	# On-Treatment Pregnancies	# Evaluable Cycles	Pearl Index (95% CI)
BMI[1] (kg/m^2)				
<30 (Non-obese)	1,123	33	9,888	4.34 (2.86, 5.82)
<25 (Normal)	684	16	6,007	3.46 (1.77, 5.16)
≥25 to <30 (Overweight)	439	17	3,881	5.69 (2.99, 8.40)
≥30 (Obese)	612	35	5,264	8.64 (5.79, 11.50)
Race				
White	1,159	46	10,281	5.82 (4.14, 7.49)
Black	418	17	3,454	6.40 (3.36, 9.43)
Other	159	5	1,430	4.55 (0.57, 8.52)
Ethnicity				
Hispanic or Latino	330	12	2,851	5.47 (2.38, 8.56)
Not Hispanic or Latino	1,406	56	12,314	5.91 (4.37, 7.46)

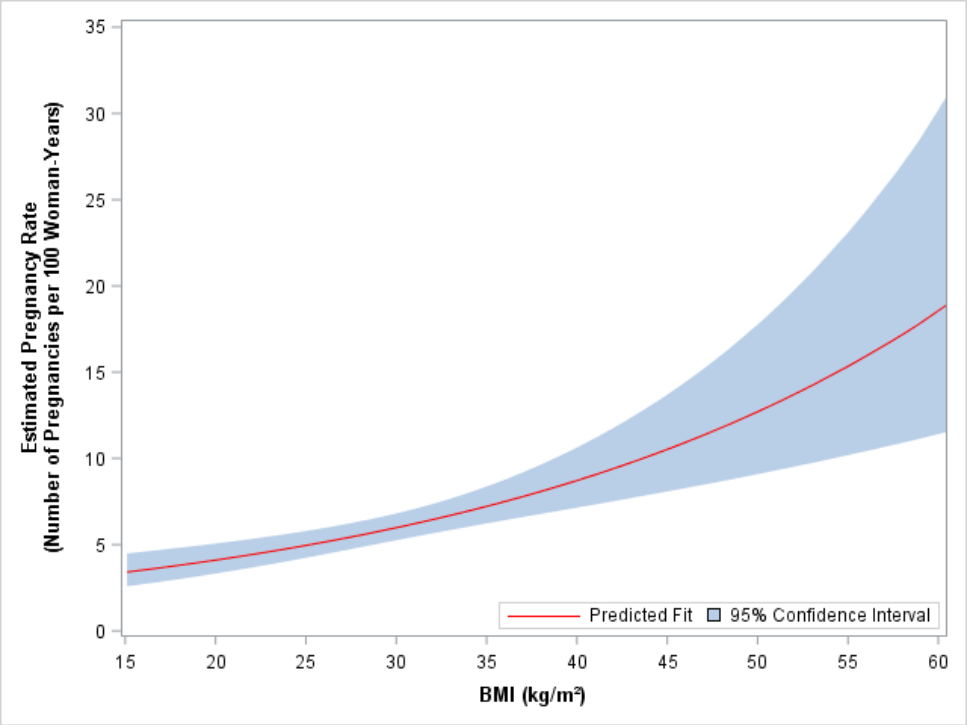
[1]BMI subpopulations (Normal, Overweight and Obese) add up to N = 1,735: Subject (b) (6) had no BMI information.
Source: Clinical Study Report Addendum and Reviewer's Analysis

Study 12 results, in Appendix 2, included an estimated Pearl Index of 7.37 (95% CI: 4.82 to 9.92) in non-obese women. Study 12 conduct issues may have resulted in an undercount of pregnancies, suggesting this may be an underestimate of the Pearl Index.

[4] Some advisory committee members at the BRUDAC meeting on October 30, 2019 recommended to include the relationship between the pregnancy rate and continuous BMI in the label.

A model using the primary analysis population from Study 23 shows a monotonic increase in the estimated pregnancy rate (i.e., the number of pregnancies per 100 woman-years) as BMI increases (Figure 1). The 95% confidence intervals drastically widen for BMI > 45 kg/m^2. Few women in the primary analysis population of Study 23 had BMI above 45 kg/m^2 (~ 3%) and they accounted for a low number of pregnancies (~ 3%) among all pregnancies in the primary analysis population (N = 1,735[5]). A description of the procedures used to identify the Poisson model to create the plot in Figure 1 are provided in Appendix 3.

Figure 1: Pregnancy Rate by BMI for Subjects ≤ 35 (ITT)



In Study 23, there were some slight numerical differences in Pearl Index estimates among the racial and ethnic subgroups. Among subjects ≤ 35 years of age, the Pearl Index estimate was slightly lower in White subjects compared to Black subjects (5.82; 95% CI 4.14, 7.49 vs. 6.40; 95% CI 3.36, 9.43), and in Hispanic or Latino subjects compared to non-Hispanic or Latino subjects (5.47; 95% CI 2.38, 8.56 vs. 5.91; 95% CI 4.37, 7.46).

The life table results based on the ITT population and CEP are similar, and their results are presented in Table 11 and Table 12 respectively in the Appendix 1 of this review. As shown, regardless if based on the ITT population or CEP, the results of the life table analysis by BMI and race were similar to the Pearl Indices. Consistent with the Pearl Index trends, among subjects ≤ 35 years of age the cumulative pregnancy rate generally increased with increasing BMI. Similar to Pearl Index analyses by race, the cumulative pregnancy rate through cycle 13 was

[5] Only N = 1,735 in the primary analysis population had BMI information, so the Poisson analysis was conducted on N = 1,735.

slightly lower in White subjects compared to Black subjects. However, inconsistent with the Pearl Index results by ethnicity, the life table estimates of pregnancy rate based on both the ITT population and CEP were slightly greater in Hispanic or Latino subjects compared to non-Hispanic or Latino subjects. Note that there were some very slight numerical differences between the review team's life table results and the Applicant's results when life table analyses were based on CEP. Despite these differences, the overall pattern of life table results was consistent with Pearl Indices. This is expected for the majority of CHC products that do not evaluate products beyond 13 treatment cycles (approximately one year).

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

As noted in the 2017 review cycle, FDA review team identified 12 additional on-treatment pregnancies in Study 23 that the Applicant failed to include in the efficacy evaluation. The Applicant agreed with the FDA's adjudication and included the additional 12 pregnancies in the efficacy evaluation contained in their current NDA submission. Based on the FDA confirmed 68 on-treatment pregnancies and 15,165 cycles, AG200-15 has demonstrated an overall Pearl Index of 5.83 with an upward trend by BMI categories of normal, overweight, and obese women with corresponding Pearl Indices of 3.46, 5.69, and 8.64, all their upper bounds of 95% CI exceeding 5. Life table results were similar to Pearl Index results.

5.2 Conclusions and Recommendations

The Applicant's claim that "AG200-15 is effective in the prevention of pregnancy in women of reproductive age"[6] has not been demonstrated through data from Study 23, despite the Applicant's efforts to enroll a highly compliant population. FDA analyses did not support the above efficacy claim as the effectiveness of AG200-15 in the general population does not meet the Division's previously communicated criteria (i.e., upper bound of Pearl Index estimate ≤ 5). We also notice that AG200-15's efficacy worsened with increasing BMI. The effectiveness of AG200-15 in the pre-specified BMI subgroups did not meet the Division's criteria. Therefore, the Applicant's LOU claim cannot be justified.

The Applicant noted that the majority of CHC products approved to date have been obtained from women who have normal weight. When looking at this subgroup, the upper bound is close to but higher than the Division's requested upper bound of 5.

There remains uncertainty whether the AG200-15 effectiveness in real world use could potentially be worse than seen in Study 23 given the results of Study 12, where the Pearl Index for women with BMI < 32 kg/m^2 was much higher (7.50 (95% CI: 5.02 to 9.97)). Finally, the VTE incidence observed in Study 23 represents a safety signal for AG200-15, particularly among obese women.

At this time in the review cycle, a remaining key issue is whether AG200-15 can be approved for a subgroup of women based on subgroup analyses by BMI in Study 23 when overall effectiveness was considered deficient in both the previous and current review cycle. We recommend that AG200-15 is studied further to reduce the uncertainty around the effectiveness and safety. Therefore, we do not recommend approval based on the lack of new clinical efficacy data.

[6] Source: ATI-CL23 Efficacy Supplement, page 20.

APPENDICES

Appendix 1

Life Table Results (Study 23)

Table 11. Cumulative Pregnancy Rate in Subjects ≤ 35 Years of Age by Subgroup (ITT): Study 23

Population	N	Number of On-Treatment Pregnancies	Number of Cycles[1]	Cumulative Pregnancy Rate (95% CI)
BMI[1] (kg/m^2)				
<30 (Non-obese)	1,123	33	9,888	4.08 (2.89, 5.74)
<25 (Normal)	684	16	6,007	3.06 (1.87, 5.00)
≥25 to <30 (Overweight)	439	17	3,881	5.59 (3.47, 8.94)
≥30 (Obese)	612	35	5,264	8.08 (5.82, 11.17)
Race				
White	1,159	46	10,281	5.36 (4.02, 7.14)
Black	418	17	3,454	6.24 (3.85, 10.02)
Other	159	5	1,430	4.65 (1.91, 11.07)
Ethnicity				
Hispanic or Latino	330	12	2,851	5.77 (3.27, 10.07)
Not Hispanic or Latino	1,406	56	12,314	5.42 (4.17, 7.03)

[1]BMI subpopulations (Normal, Overweight and Obese) add up to N = 1,735: Subject (b) (6) had no BMI information.
Source: Reviewer's Analysis

Table 12. Cumulative Pregnancy Rate in Subjects ≤ 35 Years of Age by Subgroup (CEP): Study 23

Population	N	Number of On-Treatment Pregnancies	Number of Cycles[1]	Cumulative Pregnancy Rate (95% CI)
Applicant				
BMI[2] (kg/m²)				
<30 (Non-obese)	1,177	33	-	3.97 (2.81, 5.58)
<25 (Normal)	721	16	-	2.98 (1.82, 4.88)
≥25 to <30 (Overweight)	456	17	-	5.45 (3.38, 8.70)
≥30 (Obese)	638	35	-	7.72 (5.57, 10.67)
Race				
White	1,212	46	-	5.19 (3.89, 6.90)
Black	436	17	-	5.99 (3.71, 9.60)
Other	168	5	-	4.44 (1.84, 10.50)
Ethnicity				
Hispanic or Latino	350	12	-	5.58 (3.17, 9.74)
Not Hispanic or Latino	1,466	56	-	5.23 (4.03, 6.78)
FDA				
BMI[2] (kg/m²)				
<30 (Non-obese)	1,177	33	10,619	3.98 (2.82, 5.60)
<25 (Normal)	721	16	6,464	3.01 (1.83, 4.91)
≥25 to <30 (Overweight)	456	17	4,155	5.45 (3.38, 8.70)
≥30 (Obese)	638	35	5,698	7.79 (5.62, 10.77)
Race				
White	1,212	46	11,014	5.21 (3.91. 6.94)
Black	436	17	3,757	6.06 (3.75, 9.73)
Other	168	5	1,559	4.44 (1.84, 10.50)
Ethnicity				
Hispanic or Latino	350	12	3,083	5.61 (3.19, 9.79)
Not Hispanic or Latino	1,466	56	13,247	5.27 (4.05, 6.83)

[1]The Applicant did not provide the number of cycles used in their life table analysis.
[2] BMI subpopulations (Normal, Overweight and Obese) add up to N = 1,815: Subject (b) (6) had no BMI information.
Source: Clinical Study Report Addendum and Reviewer's Analysis

Appendix 2

Pearl Index in Subjects ≤ 35 Years of Age: Study 12

Table 13. Pearl Index in Subjects ≤ 35 Years of Age: Study 12

Population	N	# On-Treatment Pregnancies	# Evaluable Cycles	Pearl Index (95% CI)
Overall	1,060	45	7,685	7.61 (5.39, 9.83)
BMI < 32 kg/m^2	827	35	6,070	7.50 (5.02, 9.97)
BMI < 30 kg/m^2	767	32	5,645	7.37 (4.82, 9.92)

Note. The pre-specified primary endpoint for Study 12 was the Pearl Index in women ≤ 35 Years of Age with BMI < 32 kg/m^2.
Source: Statistical Review of NDA 204017 SN0000

Appendix 3

Poisson Regression Model – Assessment of Pregnancy Rate and Continuous BMI

The plot in Figure 1 is based on a univariate Poisson regression model fit to the primary analysis population in Study 23. A series of regression models were performed to identify the model that could best describe the relationship between pregnancy rate and continuous BMI in Study 23. Univariate Poisson regression model was first considered for two reasons: 1) the number of pregnancies in a clinical contraceptive trial generally follows the Poisson distribution[7] and 2) Poisson model can account for the differing exposure times between subjects. Due to underdispersion in the data, the negative binomial model was considered but it failed to converge. Adding additional baseline covariates[8] into the Poisson model also failed to handle underdispersion in the data. In order to account for the excessive zeroes (> 90%) in the data, a zero-inflated model and zero-inflated binomial models were considered. However, the former was unable to improve the model fit compared to the univariate Poisson model and the latter failed to converge. A quadratic Poisson regression model and regression splines to model continuous BMI in the Poisson model also were unable to improve the model fit relative to the univariate Poisson model. Therefore, the final model used to fit the data was a univariate Poisson model with the count of on-treatment pregnancies as the response variable and continuous BMI as the predictor variable, with a scale parameter[9] based on deviance statistics being used to account for the underdispersion in the data. Bootstrapping[10] was used to validate the chosen model.

[7] Gerlinger et al. (2003). Recommendation for confidence interval and sample size calculation for the Pearl Index. *The European Journal of Contraception and Reproductive Health Care*, *8*, 87-92.

[8] Other than BMI, investigated baseline variables include Race, Ethnicity, Age, Education Level, Smoking Status, Alcohol Use in Past 12 Months, and Prior Use of Hormonal Contraceptives. Of all these investigated covariates, only age (years) and BMI are continuous covariates; all others were categorical covariates. None of these additional covariates were found to have statistical interaction with BMI.

[9] McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Edition, Chapman and Hall, London.

[10] 5,000 bootstrap samples were generated by repeatedly sampling observations with replacement from the primary analysis population (N = 1,735). Each bootstrap sample of size N = 1,735 was fit with the chosen univariate Poisson model. The mean over all 5000 BMI beta coefficient estimates was 0.0372 (SD = 0.0108), a 1% change from the beta coefficient estimate ($\beta = 0.0377$) from the original sample. The bootstrap percentile CI (2.5th percentile and the 97.5th percentile) of the 5,000 bootstrap BMI coefficient estimates was (0.0149, 0.0574).

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

YUN TANG
02/10/2020 09:47:23 PM

MAHBOOB SOBHAN
02/10/2020 09:56:09 PM

LAURA L JOHNSON
02/10/2020 09:59:56 PM

US Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics
Division of Biometrics VI

STATISTICAL REVIEW AND EVALUATION

NDA NO.	204017
SERIAL NO.	0058
DATE RECEIVED BY THE CENTER	05/16/2019
DRUG NAME	AG200-15
DOSAGE FORM	TDS
INDICATION	Contraception
SPONSOR	Agile Therapeutics, Inc.
REVIEW FINISHED	8/7/2019
STATISTICAL REVIEWER	Chao Wang, Ph.D.
SECONDARY REVIEWER	Meiyu Shen, Ph.D.
PROJECT MANAGER	Jeannie Roule

Chao Wang, Ph.D., Mathematical Statistician, CDER/OTS/OB/DBVI
Meiyu Shen, PhD., Lead Mathematical Statistician, CDER/OTS/OB/DBVI

Concur: _____

Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DBVI

CC List:

Jeannie Roule, CDER/OND/ODEIII/DBRUP

Contents

1 Executive summary

I evaluated the applicant's in vivo adhesion data collected in the clinical studies ATI-CL25 and ATI-CL26 for NDA 204017.

For ATI-CL25:

- This is a randomized, open-label, single-dose, two-treatment, comparative crossover adhesion study of AG200-15 and Xulane in healthy female volunteers.
- No significant issues regarding the study design and conduct that may adversely impact the data quality were identified.
- A total of 83 subjects were randomized in the study. Seventy-eight (78) subjects were included in the per protocol population.
- The difference between the means of the paired mean adhesion score between AG200-15 transdermal delivery system (TDS) and Xulane were -0.24, and the 95% upper confidence limit for the mean difference is -0.16, which is less than the non-inferiority margin, 0.15. Hence, AG200-15 TDS passed the non-inferiority test.
- The probability that a randomly selected AG200-15 TDS maintains at least 75% adhesion throughout its entire wear period is estimated to be 0.99 and its 95% lower confidence limit is 0.95.

For ATI-CL26:

- This is a single-dose, open-label, non-comparative adhesion study of AG200-15 TDS in healthy female volunteers.
- The probability that a randomly selected AG200-15 TDS maintains at least 75% adhesion throughout its entire wear period is estimated to be 0.93 and its 95% lower confidence limit is 0.83.

2 Introduction

On May 19, 2019, FDA Office of Pharmaceutical Quality, Office of New Drug Products requested the Division of Biometrics VI, Office of Biostatistics to evaluate the applicant's in vivo adhesion data of a transdermal delivery system (TDS) product in the clinical studies ATI-CL25 and ATI-CL26 for NDA 204017 submitted on May 19, 2019.

In this evaluation, I reviewed the study and analyzed the in vivo adhesion data.

3 Statistical reviewer's analysis

3.1 ATI-CL25

3.1.1 Study design and conduct

This is a randomized, open-label, single-dose, two-treatment comparative crossover adhesion Study of AG200-15 and Xulane in Healthy Female Volunteers. No significant issues regarding the study design and conduct that may adversely impact the data quality were identified.

3.1.2 Analysis Datasets

I analyzed the adhesion endpoints on the per protocol (PP) population. The PP population included all patches except those intentionally removed early, for example due to unacceptable irritation or those on subjects who were discontinued prior to the end of the labeled duration of wear for reasons unrelated to adhesion (e.g., due to a protocol violation). The PP population also excluded patches with prolonged water exposure (> 10 consecutive minutes on any day). A randomized subject who discontinued the study prior to receiving any patch was excluded from the PP population for both treatment periods.

There were 83 subjects randomized in the study. Two subjects in each treatment group discontinued. More specifically,

- Subject (b) (6) had prolonged water exposure during both treatment periods, so the data for neither treatment were included in the analysis.
- Subject (b) (6) had prolonged water exposure during treatment Period 1 (Xulane). Subject (b) (6) was counted in the PP population for AG200-15 and not for Xulane. However, Subject (b) (6) is not included in the pair-wise non-inferiority comparison as only data for one product is available.
- Subject (b) (6) discontinued the study earlier (only partial data for Xulane was collected). Data for neither treatment were included in the analysis.
- Subject (b) (6) discontinued the study earlier after 2 days. Data for neither treatment were included in the analysis.
- Subject (b) (6) discontinued the study earlier after 2 days. Data for neither treatment were included in the analysis.
- Subject (b) (6) discontinued the study earlier after 1 day. Data for neither treatment were included in the analysis.

Therefore, there were 78 subjects in the per protocol population for AG200-15 and 77 subjects in the per protocol population for Xulane. However, when analyzing the primary endpoint via hypothesis testing, Subject (b) (6) could not be included because her data only contributed to one period and could not be analyzed using a paired t-test.

Patch adhesion was assessed by trained study site personnel at 24-hour intervals (0 hr, 24 hrs, 48 hrs, 72 hrs, 96 hrs, 120 hrs, 144 hrs, and 168 hrs (± 2 hrs). Photographs of the patch application site were also obtained at each adhesion assessment.

All analyses used the worst score carried forward method, such that the highest adhesion score for a subject using the five-point adhesion scale assessed at any time point after baseline is used for subsequent time points until a higher score was assessed for that subject.

3.1.3 Exploratory data analysis

The frequency of adhesion scores (AS) at each time point of measurement for AG200-15 and Xulane are summarized in Tables 1 and 2 respectively. The tables show that in general both products perform better at the early stage after application than at the end of wear period. In addition, AG200-15 seems to adhere better than Xulane, as less number of scores 1 and higher occurred for AG200-15.

It is also of interest to see the first time for a TDS to be observed with a given non-zero score. These frequencies are summarized in Tables 3 and 4 for AG200-15 and Xulane respectively.

	AS=0	AS=1	AS=2	AS=3	AS=4
Day=0	78	0	0	0	0
Day=1	77	1	0	0	0
Day=2	75	3	0	0	0
Day=3	72	6	0	0	0
Day=4	68	9	1	0	0
Day=5	63	14	1	0	0
Day=6	60	17	1	0	0
Day=7	58	19	1	0	0

Table 1: Summary of frequency of adhesion scores at each time point of measurement after application for AG200-15.

	AS=0	AS=1	AS=2	AS=3	AS=4
Day=0	77	0	0	0	0
Day=1	72	5	0	0	0
Day=2	69	7	1	0	0
Day=3	64	12	1	0	0
Day=4	51	24	2	0	0
Day=5	38	35	4	0	0
Day=6	34	35	8	0	0
Day=7	29	37	11	0	0

Table 2: Summary of frequency of adhesion scores at each time point of measurement after application for Xulane.

3.1.4 Non-inferiority analysis

Here I compared the adhesion of AG200-15 (denoted by T) with that of Xulane (denoted by R) through a non-inferiority (NI) test. The paired mean adhesion score (MAS) differences between between AG200-15 and Xulane are illustrated in Figure 1. Let μ_T and μ_R denote the MAS for T and R respectively. The NI test concerns the following hypotheses, $H_0 : \mu_T - \mu_R \geq 0.15$ versus $H_1 : \mu_T - \mu_R < 0.15$. The mean and standard deviation of the paired mean adhesion score (MAS) differences are -0.24 and 0.46 respectively. Assuming that the the pair differences follow a normal distribution, the 95% upper confidence limit for the mean paired MAS difference is -0.16, which is less than the NI margin 0.15. Thus AG200-15 passed the NI test.

	AS=1	AS=2	AS=3	AS=4
Day=0	0	0	0	0
Day=1	1	0	0	0
Day=2	2	0	0	0
Day=3	3	0	0	0
Day=4	3	1	0	0
Day=5	5	0	0	0
Day=6	3	0	0	0
Day=7	2	0	0	0

Table 3: Summary of frequency of time of a TDS firstly observed with a given adhesion score after application for AG200-15.

	AS=1	AS=2	AS=3	AS=4
Day=0	0	0	0	0
Day=1	5	0	0	0
Day=2	3	1	0	0
Day=3	5	0	0	0
Day=4	13	1	0	0
Day=5	12	2	0	0
Day=6	3	4	0	0
Day=7	4	3	0	0

Table 4: Summary of frequency of time of a TDS firstly observed with a given adhesion score after application for Xulane.

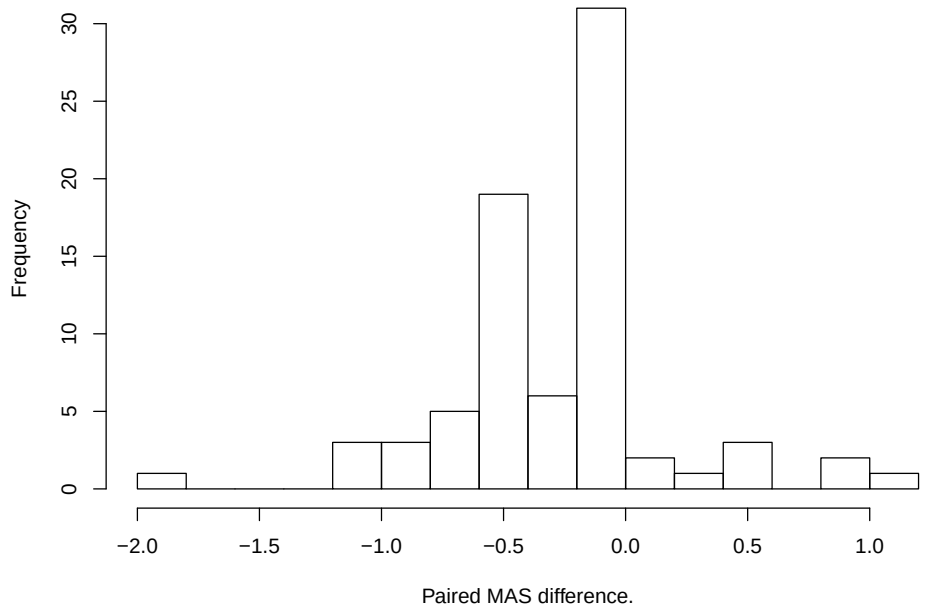


Figure 1: Histogram of paired differences of MAS between AG200-15 and Xulane for all subjects.

3.1.5 The probability of TDS maintaining at least 75% adhesion during the entire wear period

In addition, I summarized the estimated probability of a randomly selected TDS maintaining at least 75% adhesion (corresponding to adhesion score 1 or less) for the entire wear period and Jeffreys and second-order-corrected (SOC) 95% lower confidence limits (LCL) (Cai, 2005) for AG200-15 and Xulane in Table 5. As the applicant also submitted percent adhesion data, to assess the robustness of the data, I also considered the probability of the TDS maintaining at least 70% adhesion and the results are shown in Table 6. The results for AG200-15 did not change.

Product	Sample size	Number of sucesses	Estimated probability	Jeffery 95% LCL	SOC 95% LCL
AG200-15	78	77	0.99	0.95	0.95
Xulane	77	66	0.86	0.78	0.78

Table 5: Summary of the probability that a randomly selected TDS mainatains at least 75% adhesion for its entire wear period.

Product	Sample size	Number of sucesses	Estimated probability	Jeffery 95% LCL	SOC 95% LCL
AG200-15	78	77	0.99	0.95	0.95
Xulane	77	71	0.92	0.86	0.86

Table 6: Summary of the probability that a randomly selected TDS mainatains at least 70% adhesion for its entire wear period.

3.2 ATI-CL26

In addition to ATI-CL25, the applicant also submitted a second study, ATI-CL26. This is a single-dose, open-label, non-comparative adhesion study of AG200-15 in healthy female volunteers. This study is similar to ATI-CL25 except that there is no comparative product. The study result is summarized in Table 7.

Sample size	Number of sucesses	Estimated probability	Jeffery 95% LCL	SOC 95% LCL
30	28	0.93	0.83	0.83

Table 7: Summary of the probability that a randomly selected TDS mainatains at least 75% adhesion for its entire wear period for Study ATI-CL26.

4 Conclusion and Recommendation

In conclusion, my independent evaluation showed that the data in the controlled study ATI-CL25 support the non-inferiority of AG200-15 to Xulane in terms of in vivo adhesion. In addition, a standalone evaluation of the data for AG200-15 showed that the probability that a randomly selected AG200-15 maintains at least 75% adhesion throughout its entire wear period in the controlled setting is estimated to be 0.99 and the 95% lower confidence limit is 0.95. Using the data in the single-arm, noncomparative study ATI-CL26, the probability is estimated to be 0.98 and the 95% lower confidence limit is 0.83, where the smaller 95% lower confidence limit is likely due to the smaller sample size.

References

Cai, T. T. (2005), "One-sided confidence intervals in discrete distributions," *Journal of Statistical planning and inference*, 131(1), 63–88.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

CHAO WANG
08/13/2019 03:21:16 PM

MEIYU SHEN
08/13/2019 03:24:04 PM

YI TSONG
08/14/2019 07:43:43 AM

US Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics
Division of Biometrics VI

STATISTICAL REVIEW AND EVALUATION

NDA NO.	204017
SERIAL NO.	
DATE RECEIVED BY THE CENTER	7/19/2017
DRUG NAME	Twirla (levonorgestrel/ethinyl estradiol 120/30 mcg/day) Transdermal Patch
DOSAGE FORM	TDS
INDICATION	
SPONSOR	Agile
REVIEW FINISHED	11/13/2017
STATISTICAL REVIEWER	Chao Wang, Ph.D.
SECONDARY REVIEWER	Meiyu Shen, Ph.D.
PROJECT MANAGER	Thao Vu

Chao Wang, Ph.D., Mathematical Statistician, CDER/OTS/OB/DBVI
Meiyu Shen, PhD., Lead Mathematical Statistician, CDER/OTS/OB/DBVI

Concur: _____

Yi Tsong, Ph.D., Division Director, CDER/OTS/OB/DBVI

CC List:

Lillian Patrician, CDER/OTS/OB
Thao Vu, CDER/OPQ/OPRO/DRBPMI/RBPM

Contents

1 Statistical evaluation of evidence

1.1 Executive summary

The statistical reviewer evaluated the sponsor's in vivo adhesion data collected in the clinical study ATI-CL23 for NDA 204017. The adhesion data was collected from 55892 patches used by 2022 subjects.

Below is the summary of our review.

- Data quality:
  - The study protocol specified that the subjects should try to reapply a patch if it is detached partially, with no specific instruction on the timing of measuring adhesion. The reapplication and uncertainty in assessment timing created inconsistency.
  - An investigator would assess adhesion during scheduled a subject's in-person clinic visit. However, any patch could be assessed by an investigator at most once and thus the investigator score could not represent the adhesion of a patch for its entire 7-day wear period. The statistical reviewer used this dataset to compare the subject data. Since 91% investigator data were the same with the corresponding subject data, statistical reviewer used subject data to evaluate adhesion.
  - 22% of adhesion data reported by subjects were missing.
  - Out of 55892 patches, 5% patches detached completely before the end of the 7-day wear period and 2.1% patches detached completely on day 1.
  - 54.4% subjects experienced at least one patch complete detachment during the study. 25% subjects experienced at least one patch complete detachment in cycle 1, and this proportion decreased to 6.1% in cycle 13.
  - For all patches used in all cycles, 5% patches detached completely. 9.9% patches detached completely in cycle 1, and the proportion of detached patches decreased to 2.4% in cycle 13.
- Statistical analysis of subject data:
  - CMC reviewers recommended that a patch have acceptable adhesion with adhesion score less than 2 throughout the entire wear period. We quantified the adhesion of a patch by its first time to a score of 2 or greater, and performed a hypothesis test with the null hypothesis being *the probability that the first time a patch reaches a score of 2 or greater is beyond the proposed wear period is less than 0.9* with significance level 0.05, using the data for all patches of all cycles. The point estimate of the probability that the first time a patch reaches a score of 2 or greater is beyond the proposed wear period is 88.7%, with the 95% lower confidence limit 85.33%. Thus the null hypothesis could not be rejected and the test result does not support that there are at least 90% patches showing daily adhesion score less than 2 throughout the wearing period with 95% confidence.
  - We also performed the above hypothesis test with patch data grouped by cycles. None of the cycle-wise data supported acceptable adhesion.
  - The study setting differs from that of a typical clinical study for adhesion, which usually consists of 50 to 100 subjects who would stay in a clinic, wear a single patch for 7 days and have an investigator assess adhesion daily. In order to see the results for a typical study, at within-subject patch sequence number 1, 10, 20, 30, and 40, we subsampled the patch data for 50 subjects and performed the above hypothesis test with 10^4 replicated subsamples. Nearly none of the subsampled data supported acceptable adhesion.
- Overall, the data do not support that the patch adhesion is acceptable and we recommend the sponsor conduct another clinical study which is specially designed for adhesion.

1.2 Introduction

On 6/19/2017, FDA Office of Pharmaceutical Quality, Office of New Drug Products requested the Division of Biometrics VI, Office of Biostatistics to evaluate the sponsor's in vivo adhesion data of a transdermal delivery system (TDS) product in the clinical study, ATI-CL23, in NDA 204017.

The adhesion of the TDS is a critical quality attribute for drug delivery. A TDS with poor adhesion could adversely affect drug delivery. In this review, we analyzed the data from both investigators and subjects and provided a criterion for evaluating the adhesion performance of a new TDS product. We also recommended a new adhesion study.

1.3 Statistical Reviewer's analysis

1.3.1 Comments on the study design and the impact on data quality

The adhesion data were collected in an open-label, multicenter Phase 3 clinical study with around 2000 subjects enrolled for 1 year (thirteen 28-day cycles) of treatment at up to 70 investigation sites. Patch adhesion was measured in terms of adhesion score (AS) reported by both subjects and investigators (*subject data* and *investigator data*).

All subjects were scheduled to complete a total of 8 in-person clinic visits spread across the entire study period, during which the investigators measured AS according to the scoring system defined by the FDA guidance, *Assessing Adhesion with Transdermal Delivery Systems and Topical Patches for ANDAs* (hereafter referred to as the Guidance):

- 0: $\geq 90\%$ adhered (essentially no lift off the skin),
- 1: $\geq 75\%$ to $< 90\%$ adhered (some edges only lifting off the skin),
- 2: $\geq 50\%$ to $< 75\%$ adhered (less than half of the TDS lifting off the skin),
- 3: $> 0\%$ to $< 50\%$ adhered (not detached, but more than half of the TDS lifting off the skin without falling off),
- 4: $= 0\%$ adhered (TDS detached; completely off the skin).

The subjects self-evaluated the AS daily and recorded the score into an eDiary system in which they also entered other information such as patch change and removal day. The subjects were instructed to follow a simplified version:

- 0: No lifting or small amount of lifting at the edges of the patch,
- 1: More than a small amount of lifting at the edges up to one-quarter of the patch lifting off,
- 2: More than one quarter of the patch lifting off up to half of the patch lifting off,
- 3: More than half of the patch is lifting off, but the patch is still on,
- 4: Patch has completely come off.

Since there were two sources of adhesion data, an important question is which dataset we should use. It is generally believed that the investigator measurement is more accurate and objective and it is ideal if the investigator data could be used. To assess the overall adhesion for a TDS, it is crucial to monitor adhesion each day throughout the entire wear period of 7 days for this TDS. Because investigator data were measured during irregular time points and for a patch it could be assessed by an investigator at most once, investigator

data were not very useful. On the other hand, the subjects reported AS daily and thus provided more information. However, subject data suffered the following issues:

- The subjects were instructed to measure the adhesion daily. However, there was no specification at what time a subject should measure, creating inconsistency in the time interval between two measurements.
- The fact that this product is re-adhesible also makes the adhesion study nontraditional. In a typical adhesion study the objective is to assess the adhesion throughout the wear period without human intervention, the subject should not try to reapply the patch once it is applied. In Section 8.4.3 of the study protocol amendment dated December 18, 2014, it specified the following. "If the patch partially or completely detaches from the skin, the subject should attempt to reapply the same patch. If it cannot successfully be fully reapplied, the patch should be immediately replaced with a new patch; supplemental adhesives or wraps should not be used to hold the patch in place." It implies that subject intervention on patch adhesion can happen. Since our objective is to study adhesion, if a patch need to be replaced by a new patch whenever it is cannot fully reapplied, we cannot observe the patch adhesion performance if it cannot fully reapplied after observing even partial detachment. From the submitted data, we observed within-patch AS fluctuation, and a score of greater than 0 implies that the subject would reapply the patch if the above cited protocol is fully complied. The reapplication of a patch which has detached partially could also make the adhesion in subsequent days better, compared with no reapplication.
- There existed other cases in which a patch could not be used until the end of the 7-day period which resulted in missing data. The possible cases are listed below:
  - A subject might forget to assess the patch and input the record.
  - A patch which detached completely and could not be fully reapplied must be replaced with a new patch.
  - If the subject experienced skin irritation and/or itching that she felt warranted patch removal, the patch should be removed and discarded, and a new (reserve) patch should be immediately applied. The subject should remain on the same patch change schedule. In this case, neither of the two patches would be used for 7 days.
  - If during any cycle, the subject went for a period of < 24 hours without a patch on or a patch became detached and remained detached for < 24 hours, the subject should replace it with a new patch immediately. The subject's patch change day would remain the same. In this case, neither of the two patches would be used for 7 days.
  - If during any cycle, the subject forgot to change her patch on the scheduled patch change day and it had been < 48 hours from the time of the scheduled change, the subject should apply a new patch immediately. The next patch should be applied on the usual patch change day. The new patch would be used for less than 7 days.

1.3.2 Imputing missing subject data

We imputed the missing data in subject data before conducting our analysis. From a regulatory point of review, we used the imputation method recommended by the Guidance. For a patch, a missing record for which there was no preceding record was replaced by the earliest available record, and a record with preceding record(s) was replaced by the maximum score of all preceding record(s). Table 1 summarizes the data after imputation.

	Count	Percentage (%)
Imputed	86701	21.85
Original	310122	78.15

Table 1: Summary of original and imputed data.

There might be doubt in whether the scores reported by the subjects were accurate. We addressed this issue by comparing subjects data with investigator data to confirm the accuracy of subject data.

1.3.3 Comparison between subject data and investigator data

Since investigator data were of smaller amount than subject data, for each score measured by an investigator, we matched it with the subject score sharing the same subject ID and analysis date, then we computed the differences between the matched subject and investigator scores. The differences are summarized by Table 2, showing that 90.91% scores were exactly the same, and 6% scores were different by 1. There were also scores with difference of 4 which might be due to the different timings of measurements by subjects and investigators in the case that a patch completely fell off and a new patch was applied. The summary of the differences shows that the scores reported by subjects were largely reliable and subject data could be used to assess the patch adhesion.

	-4	-3	-2	-1	0	1	2	3	4
Count	25	5	17	209	9257	426	79	48	117
Empirical probability (%)	0.25	0.05	0.17	2.05	90.91	4.18	0.78	0.47	1.15

Table 2: Summary of differences between matched subject and investigator data.

1.3.4 Summary statistics of subject data

We first provided some summary statistics for subject data. In Table 3 we summarize the AS on different patch days. 2.1% patches fell off completely on day 1 and 3.0% patches on day 4. The percentage of patches with score 0 decreased from 90% on day 1 to 79% on day 7. The percentages of patches with scores other than 1 increased with patch day.

Patch day \ AS	0	1	2	3	4
1	51140 (90%)	3188 (5.6%)	741 (1.3%)	424 (0.75%)	1196 (2.1%)
2	53784 (95%)	1784 (3.2%)	307 (0.54%)	164 (0.29%)	650 (1.2%)
3	52338 (92%)	2888 (5.1%)	460 (0.81%)	253 (0.45%)	750 (1.3%)
4	50288 (89%)	4075 (7.2%)	759 (1.3%)	413 (0.73%)	1154 (2.0%)
5	48285 (85%)	5406 (9.5%)	1018 (1.8%)	541 (0.95%)	1439 (2.5%)
6	46545 (82%)	6446 (11%)	1373 (2.4%)	714 (1.3%)	1611 (2.8%)
7	44996 (79%)	7435 (13%)	1681 (3.0%)	884 (1.6%)	1693 (3.0%)

Table 3: Summary of AS on different patch days for all patches. For each patch day, the number of patches with given AS is followed by percentages in round brackets.

We summarized the within-patch maximum AS (max(AS)) of all 55892 patches in Table 4. Overall 72% patches had score 0 for all 7 days, 17% patches had a maximum score of 1, and 5% patches fell off within the wear period. For the within-patch max(AS) grouped by cycles, 9.9% patches in cycle 1 detached completely and the proportion decreased to 2.4% in cycle 13. The cycle-wise summary for within-patch max(AS) being 1, 2, 3, and 4 are also illustrated in Figure 1.

max(AS) / Cycle	0	1	2	3	4	Total
1	3709 (58%)	1417 (22%)	385 (6.1%)	215 (3.4%)	630 (9.9%)	6356
2	3767 (65%)	1144 (20%)	295 (5.1%)	167 (2.9%)	466 (8.0%)	5839
3	3658 (69%)	1010 (19%)	212 (4.0%)	137 (2.6%)	306 (5.7%)	5323
4	3498 (71%)	868 (18%)	190 (3.9%)	125 (2.5%)	228 (4.6%)	4909
5	3274 (72%)	784 (17%)	182 (4.0%)	111 (2.4%)	205 (4.5%)	4556
6	3198 (74%)	695 (16%)	156 (3.6%)	97 (2.2%)	171 (4.0%)	4317
7	3036 (75%)	649 (16%)	146 (3.6%)	84 (2.1%)	142 (3.5%)	4057
8	2967 (77%)	587 (15%)	117 (3.0%)	60 (1.6%)	134 (3.5%)	3865
9	2845 (78%)	506 (14%)	119 (3.3%)	73 (2.0%)	103 (2.8%)	3646
10	2632 (76%)	510 (15%)	125 (3.6%)	65 (1.9%)	126 (3.6%)	3458
11	2560 (77%)	480 (15%)	110 (3.3%)	51 (1.5%)	106 (3.2%)	3307
12	2517 (79%)	426 (13%)	102 (3.2%)	59 (1.8%)	100 (3.1%)	3204
13	2440 (80%)	386 (13%)	96 (3.1%)	61 (2.0%)	72 (2.4%)	3055
Overall	40101 (72%)	9462 (17%)	2235 (4.0%)	1305 (2.3%)	2789 (5.0%)	55892

Table 4: Summary of within-patch max(AS). For each cycle or overall, the number of patches with given AS is followed by percentages in round brackets.

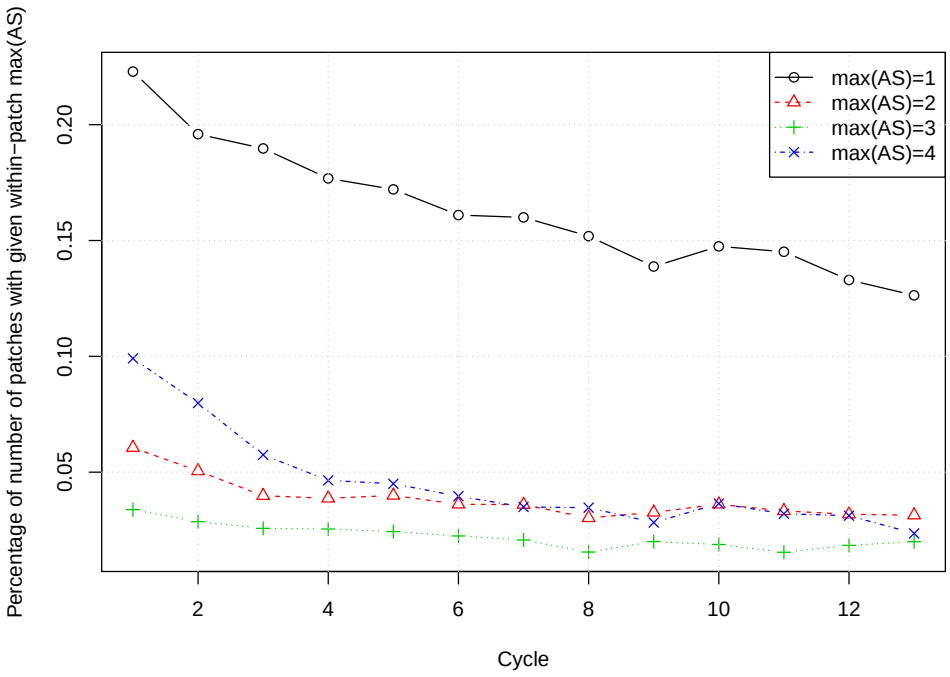


Figure 1: Plot of percentage of number of patches with given within-patch max(AS) for each cycle.

It is also of interest to see the worst score for each subject, so we summarized the within-cycle and overall max(AS) for all subjects in Table 5. The within-cycle max(AS) for any subject is the maximum AS for all patch scores of the subject for the given cycle. The overall max(AS) for a subject is the maximum AS of all AS for the subject. The table shows that 54.4% subjects experienced complete detachment during the study; cycle-wise, 24.8% subjects experienced complete detachment in cycle 1 and the proportion decreased to 6.1% in cycle 1. The cycle-wise summary for within-cycle max(AS) being 1, 2, 3, and 4 is also shown in Figure 2.

Cycle \ max(AS)	0	1	2	3	4	Total
1	686 (34%)	520 (26%)	183 (9.1%)	129 (6.4%)	501 (24.8%)	2019
2	772 (41%)	460 (25%)	155 (8.3%)	97 (5.2%)	382 (20.5%)	1866
3	819 (47%)	471 (27%)	116 (6.7%)	77 (4.5%)	246(14.2%)	1729
4	836(52%)	382(24%)	117(7.3%)	78 (4.9%)	195(12.1%)	1608
5	799(53%)	351(23%)	106(7.0%)	73(4.9%)	176 (11.7%)	1505
6	794(56%)	331(23%)	85(6.0%)	54(3.8%)	156 (11.0%)	1420
7	794(59%)	285(21%)	81(6.0%)	56(4.2%)	128 (9.5%)	1344
8	748(59%)	301(24%)	62(4.9%)	39 (3.1%)	117 (9.2%)	1267
9	748 (62%)	250(21%)	67(5.5%)	51 (4.2%)	93 (7.7%)	1209
10	689(60%)	229(20%)	71(6.2%)	39(3.4%)	113 (9.9%)	1141
11	668(61%)	238(22%)	58(5.3%)	36(3.3%)	92 (8.4%)	1092
12	673(63%)	211(20%)	62(5.8%)	34 (3.2%)	84 (7.9%)	1064
13	668(65%)	202(20%)	49(4.8%)	41(4.0%)	62(6.1%)	1022
Overall	230 (11%)	357(18%)	174(8.6%)	161 (8.0%)	1100 (54.4%)	2022

Table 5: Summary of within-cycle max(AS). For each cycle or overall, the number of patches with given AS is followed by percentages in round brackets.

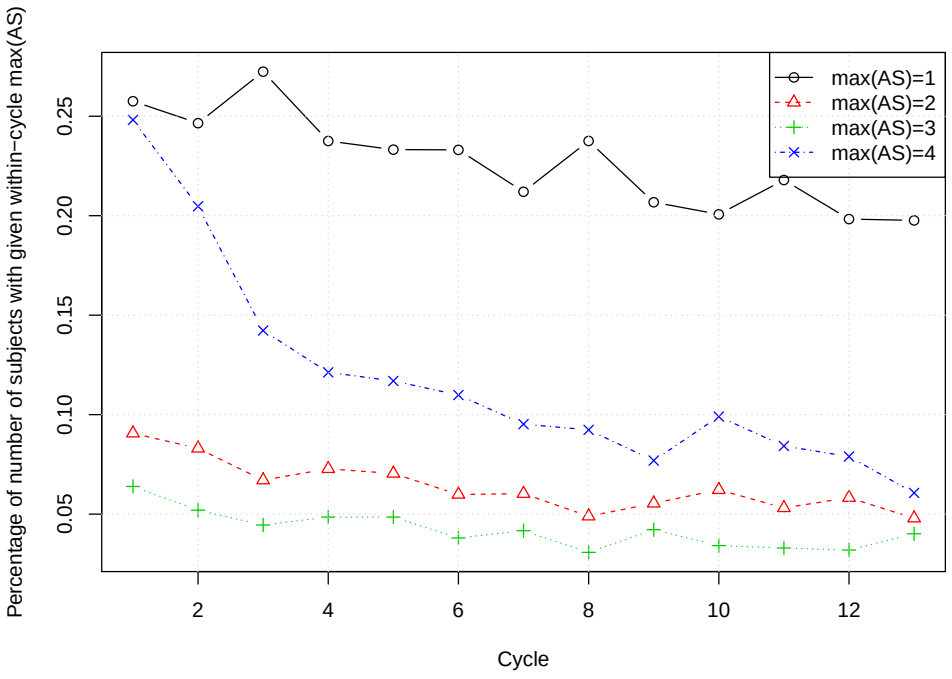


Figure 2: Plot of percentage of number of subjects with given within-cycle max(AS).

1.3.5 How to quantify adhesion performance

In order to evaluate patch adhesion from the perspective of quality control, we need to establish a criterion to determine whether the adhesion is acceptable. This is a study consisting of one single TDS, and there is no existing FDA guidance in assessing the adhesion of a single TDS. We used a criterion for assessing adhesion in patch level. More specifically, let θ be the minimum score such that the patch adhesion is not acceptable. Our primary endpoint is the first time (measured in days) to score θ, denoted by $t(\theta)$, for a patch. A reasonable requirement is that the probability of a patch with $t(\theta)$ greater than the number of scheduled days of application, which is 7 for this TDS, should be with high probability, or mathematically, $P(t(\theta) > 7) > p_0$. For instance, $p_0 = 90\%$ implies that a user can expect that more than 90% of patches can be used for 7 days, while less than 10% of patches will achieve a score equal to or greater than θ within 7 days.

We used the following hypothesis test to determine if the adhesion quality of a TDS is acceptable:

$$H_0 : P(t(\theta) > d_{end}) \leq p_0,$$

versus

$$H_1 : P(t(\theta) > d_{end}) > p_0,$$

where $d_{end} = 7$ and $p_0 = 90\%$.

This is a binomial trial. Let n be the total number of patches, $A_{s,c,i}(t)$ be the AS of the patch i on day t in cycle c for subject s, and $\hat{p}$ be the sample estimate of $P(t(s) > d_{end})$. Then $\hat{p} = \frac{1}{n}\sum_{s,c,i}\prod_{t=1}^{d_{end}} 1\{A_{s,c,i}(t) < \theta\}$. Let $\alpha = 0.05$ be the significance level of the test. Since the sample size is very large, we reject H_0 if

$$t_{stat} = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} > z_{1-\alpha},$$

where $z_{1-\alpha}$ be the $(1-\alpha)$-quantile of standard normal distribution.

Alternatively, H_0 is rejected if its 95% lower confidence limit, given by $\hat{p} - z_{1-\alpha}\sqrt{\hat{p}(1-\hat{p})/n}$, is greater than p_0.

Through the discussion with CMC reviewers, we determined that a patch of AS less than 2 is acceptable, i.e., $\theta = 2$. Thus, in order for a patch to have acceptable adhesion, it should have score less than 2 throughout the entire wearing period, i.e., its corresponding time-to-event variable $t(\theta)$ should be greater than 7. The values of $t(\theta)$ are $1, 2, \ldots, 7$, and Inf which represents censored value beyond 7. Assuming that there is not significant within-subject correlation, the time-to-event variable $t(\theta)$ for each patch can be assumed to be independent and identically distributed. Its empirical distribution based on subject data is summarized in Table 6, in which in addition to $t(2)$, the distributions of $t(3)$ and $t(4)$ are also summarized.

θ \ AS	1	2	3	4	5	6	7	Inf
2	2361 (4.2%)	369 (0.65%)	425 (0.75%)	742 (1.3%)	778 (1.4%)	893 (1.6%)	836 (1.5%)	50285 (89%)
3	1620 (2.9%)	296 (0.52%)	272 (0.48%)	455 (0.8%)	502 (0.89%)	513 (0.9%)	478 (0.84%)	52553 (93%)
4	1196 (2.1%)	249 (0.44%)	173 (0.31%)	308 (0.54%)	339 (0.6%)	312 (0.55%)	236 (0.42%)	53876 (95%)

Table 6: Summary of time to score $\theta = 2, 3, 4$. The frequency counts are followed by percentages in round brackets.

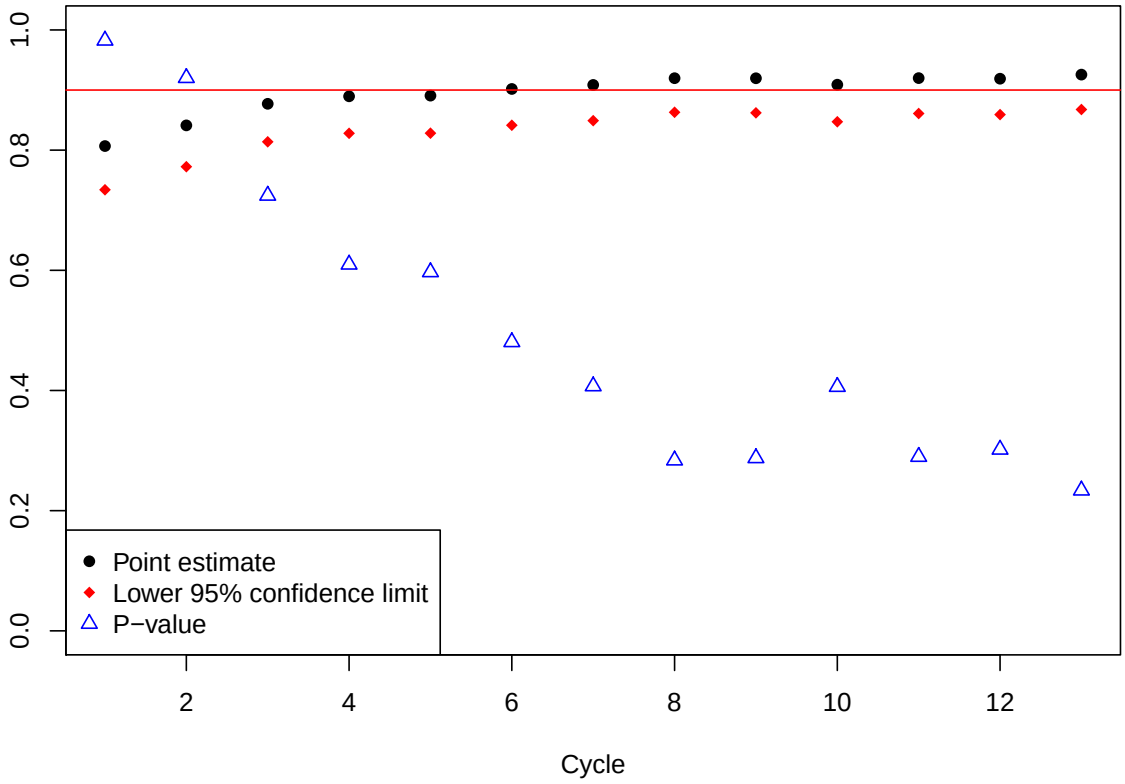


Figure 3: Plot of test results with data grouped by cycles. For each cycle, the point estimate, lower 95% confidence limit, and p-value are illustrated by black, red, and blue filled circle, respectively.

The estimated probability of failure (having a score of 2 or greater within 7-day wear period) is 88.7% with 95% lower confidence limit 85.31%. Thus H_0 cannot be rejected, i.e., *there is no strong evidence that there are at least 90% patches showing daily AS less than 2 throughout the wearing period.*

It may be conjectured that the patch adhesion will improve as subjects become more experienced. To see whether this is true, we performed the above test with the data grouped by cycles. Figure 3 illustrates the test result, where for each cycle, the point estimate, 95% lower confidence limit and p-value are given. The null hypothesis could not be rejected for any cycle, although the increasing point estimates and lower 95% confidence limits for the probability of a patch passing the adhesion test suggested improving adhesion performance.

1.3.6 Test of subsamples mimicking a typical clinical study

Since this study differs from a typical clinical adhesion study in which each of 50 to 100 subjects stays in a clinical site and wears a single patch for 7 days so that an investigator can assess the adhesion daily, it is of interest to see the adhesion result in the typical clinical setting. This can be done by performing the

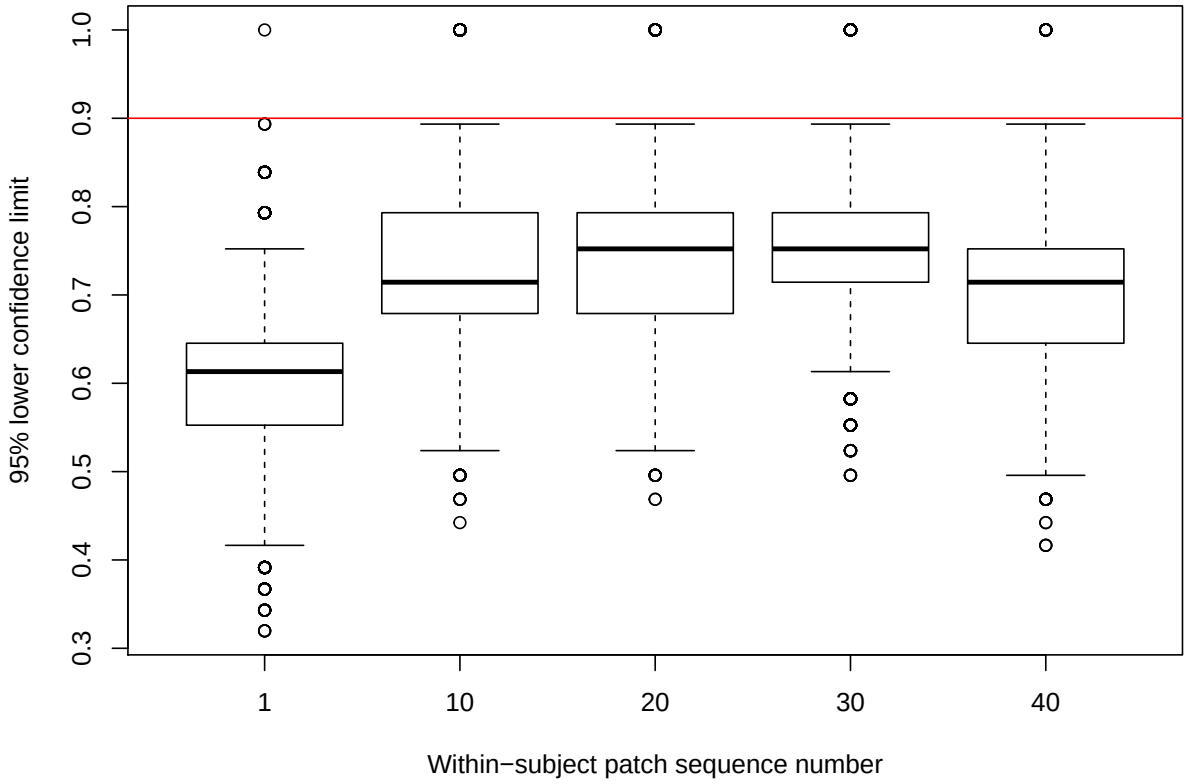


Figure 4: Box-plots for the 95% lower confidence limits of the tests computed from 50 subjects at within-subject patch sequence number 1, 10, 20, 30, and 40.

hypothesis test with subsample of the original data. We first randomly sampled 50 subjects. To control for the "learning" effect due to subjects' increasing experience, for each selected subject, we selected the patch adhesion data for i-th patch that the subject used, with i being 1, 10, 20, 30, or 40. For each i, with the 7-day adhesion data from i-th patches that 50 subjects used, we performed the same hypothesis test. To eliminate the randomness due to different subjects subsampled from the population, for each i we repeated the random sampling procedure for 10^4 times. In this way, it is equivalent to conducting 10^4 clinical studies for each i. To summarize the results, for each i, we computed the boxplot of 10^4 estimated 95% lower confidence limits for p_0, shown in Figure 4. Since the null hypothesis is rejected only when the 95% lower confidence limit is greater than 0.9, the plot indicates that only 258 (0.52%) out of 5×10^4 simulated clinical studies supported acceptable adhesion.

1.4 Conclusion and Recommendation

Assuming that a patch should have AS less than 2 (at least 75% adhered) throughout the entire 7-day wear period to ensure acceptable adhesion, the analysis of the subject data shows that the probability of a patch

passing the adhesion test is 0.89 with 95% lower confidence limit 85.31%, indicating that *there is no strong evidence that there are at least 90% patches showing daily AS less than 2 throughout the wearing period*. Nearly all tests based on random subsample of subject data mimicking typical clinical study for adhesion could not reject the null hypothesis and thus do not support acceptable adhesion as well.

We recommend the sponsor conduct another clinical study which is specially designed for adhesion.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

CHAO WANG
06/27/2018
This is a duplicate of a previously checked-in review. This version fixed some typos and the display of a figure. No significant change has been made.

MEIYU SHEN
06/27/2018

YI TSONG
06/28/2018

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA Serial Number:	204017 / SN38
Drug Name:	Twirla® (2.6 mg levonorgestrel / 2.3 mg ethinyl estradiol) transdermal delivery system
Indication(s):	Prevention of Pregnancy
Applicant:	Agile Therapeutics, Inc.
Date(s):	Submission Date: 06/26/2017
	PDUFA Due Date: 12/26/2017
Review Priority:	Class II Resubmission
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Kate Li Dwyer, Ph.D.
Concurring Reviewer:	Mahboob Sobhan, Ph.D., Team Leader
Medical Division:	Division of Bone, Reproductive and Urologic Drug Products
Clinical Team:	Nneka McNeal-Jackson, M.D., Medical Reviewer
	Catherine Sewell, M.D., Acting Team Leader
Project Manager:	Charlene Williamson

Keywords: NDA review, clinical studies, compliance

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

AD	Adhesion Score
ß-hCG	beta-Human Chorionic Gonadotropin
BMI	Body Mass Index
CEP	Contraceptive efficacy population
CI	Confidence Interval
CRL	Complete Response Letter
eCRF	electronic Case Report Form
eDiary	Electronic Diary
EE	Ethinyl Estradiol
FDA	Food and Drug Administration
HC	Hormonal Contraceptives
ITT	Intent-to-Treat
LNG	Levonorgestrel
MF	Method Failure
OC	Oral Contraceptive
PI	Pearl Index
PPI	Per-Protocol-Instructions
TCDS	Transdermal Contraceptive Delivery System

1 EXECUTIVE SUMMARY

The Applicant is seeking approval of Twirla® (hereafter referred to as AG200-15), a 7-day transdermal contraceptive delivery system for the prevention of pregnancy. The original NDA was submitted on April 12, 2012 with one Phase 3 study (Study ATI-CL12) considered to be pivotal. On February 13, 2013, the Division issued a Complete Response Letter (CRL) to cite clinical and study conduct deficiencies among others.

To address the clinical deficiency, the Division recommended the Applicant to "conduct a new preapproval Phase 3 study in a representative sample of women in the U.S. who are seeking hormonal contraception. This study will need to demonstrate an acceptable Pearl Index and upper bound of the 95% confidence interval" in the CRL. At the End of Review meeting and in subsequent discussions with the Applicant, the Division reiterated the need to "optimize continuation in the study and compliance with use of study drug, and to minimize loss to follow-up and missing data".

Study ATI-CL23 was conducted to address the contraceptive efficacy, safety and tolerability, and patch adhesion of the product. Study ATI-CL23 was a single-arm, open-label, multicenter Phase 3 study of the contraceptive efficacy, safety and tolerability of AG200-15 transdermal contraceptive delivery system (TCDS). The study was conducted entirely in the US with no restriction on the weight or Body Mass Index (BMI) of study participants. The primary efficacy objective of the study was to evaluate the contraceptive efficacy of AG200-15 irrespective of BMI. The pre-specified primary analysis was the pregnancy rate assessed by the Pearl Index for all in-treatment pregnancies in women aged ≤ 35 years at enrollment with intercourse and no other use of birth control methods.

This review focused on assessing whether the Applicant address the clinical and study conduct deficiencies cited in the CRL. During the early review stage, the clinical team has identified twelve additional pregnancies that were not included in the Applicant's original submission. Therefore, the efficacy assessment in this review included these additional pregnancies. The following are the highlights of findings noted in this submission:

- Although the Applicant had made efforts to maximize subject continuation and retention, the discontinued due to subject's decision remains high (15.3%). In general, early withdrawal in this study was 10% higher compared to approved hormonal contraceptives (HC) since 2008.
- Based on the FDA confirmed pregnancies, the Pearl Index in study ATI-CL23 was 5.83 with the upper bound of 95% CI of 7.2 irrespective of BMI. The Pearl Index in non-obese subjects (BMI < 30 kg/m^2) was 4.34 with the upper bound of 95% CI of 5.82. These Pearl Indices and the upper bounds of their associated 95% confidence intervals remains substantially higher than that seen in the registration trials for any of the approved hormonal contraceptives.

- The Pearl Index in obese subjects (BMI ≥ 30 kg/m^2) was 8.64 with the upper bound of 95% CI of 11.52 which was twice as much compared to non-obese subjects.

In summary, despite improvement in the study conduct and patch quality, the Applicant does not provide sufficient evidence to address the clinical deficiency stated in the CRL with an acceptable Pearl Index and upper bound of the 95% confidence interval. From a statistical perspective, given

the conduct of the study, poor patch quality and high pregnancy rate based on pre-specified analyses and the comparison to recent proved HC drugs, the evidence from this newly conducted trial (ATI-CL23) is not sufficient to support the approval of AG200-15 in the prevention of pregnancy.

2 INTRODUCTION

2.1 Overview

The Applicant, Agile Therapeutics, Inc., submitted original NDA on April 12, 2012 to seek approval for Twirla®, refer to AG200-15(levonorgestrel and ethinyl estradiol) (LNG and EE) transdermal delivery system (TDS), for the prevention of pregnancy. The Pearl Index (PI) in the original pivotal efficacy study for AG200-15 (Study ATI-CL12) was substantially higher than the Pearl Indices in the pivotal efficacy studies of approved hormonal contraceptives (HC). Note that there is no pre-determined acceptable Pearl Index and upper bound in FDA guidance or regulations that establishes what an acceptable Pearl Index is for purposes of contraceptive drug approval.

This Class 2 resubmission contains a new Phase 3 study (ATI-CL23) entitled: A single-arm, open-label, multicenter phase 3 study of the contraceptive efficacy, safety and tolerability of the AG200-15 transdermal contraceptive delivery system (TCDS). Study ATI-CL23 was conducted to address the deficiencies noted in the Complete Response (CR) letter. Table 1 presents key features of the above study.

Table 1: Key Features: Study ATI-CL23

Study Number (No. of Sites / Country) Dates of Study Conduct	Subject Population	Treatments	Sample Size[1] (Safety/CEP)	Duration of Treatment	Design[2]
ATI-CL23 (102 / U.S.) 9-23-14 to 11-03-16	Heterosexually active, 18-40 years old females who were at risk of pregnancy with no weight restriction	AG200-15 (3 weekly patches, 7 days patch-free)	2031/2024	Cycles 1-13	SA, OL, MC

[1] CEP = Contraceptive Efficacy Population
[2] SA = Single Arm, OL = Open Label, MC = Multicenter

2.2 Data Sources

The study reports and the data sets were submitted electronically to the Electronic Document Room. The SAS data sets were complete and well documented.

The study reports are located at:

\\CDSESUB1\evsprod\NDA204017\0038\m5\53-clin-stud-rep\535-rep-effic-safety-stud\contraception\5351-stud-rep-contr\ati-cl23

The datasets and programs for study ATI-CL23 are located at:

\\CDSESUB1\evsprod\NDA204017\0038\m5\datasets\ati-cl23

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

3.1.1.1 Study Design

Study ATI-CL23 was a single-arm, open-label, multicenter Phase 3 study of the contraceptive efficacy, safety and tolerability of AG200-15 transdermal contraceptive delivery system (TCDS). The study was conducted entirely in the US with no restriction on the weight or BMI of study participants with the intention to represent the diverse population in women seeking hormonal contraception.

AG200-15 was designed to deliver daily hormone exposure, as measured by an area under the curve (AUC) of EE and LNG like oral doses of 120µg of LNG and 30µg of EE. AG200-15 is used in a 4-week (28-day) treatment cycle: a patch is applied and replaced every 7 days for 3 consecutive weeks, followed by a 1-week "patch-free" period.

Approximately 2,100 sexually active subjects at risk for pregnancy and desiring to use contraception, including approximately 1,900 sexually active women aged 18 to 35 years and approximately 200 women aged > 35 years, were enrolled for 1 year (thirteen 28-day cycles) of treatment at 102 investigational sites. The sample size target of 200 women aged >35 years was achieved by using centralized stratified enrollment.

The study consisted of an initial Screening Visit followed by a Run-In Visit and subsequent Run-In Period during which subjects were required to demonstrate ability to use and comply with an eDiary and other requirements for communication with the investigative site. Subjects who successfully completed the Run-In Period were enrolled for a treatment period of 1 year, or thirteen 28-day cycles, during which they were to complete a total of 8 scheduled in-person clinic visits and 6 telephone visits. Urine pregnancy testing was performed during each clinic visit.

Subjects utilized eDiary in which they entered daily information on: patch wear, patch adhesion, exposure of the patch to water, patch-related skin irritation/itching, and vaginal bleeding/spotting. Subjects also entered weekly patch use information, including patch change and removal day, anatomic site of patch application, sexual activity and use of back-up methods of contraception (including reasons for back-up contraception if applicable).

Subjects recorded patch adhesion in the eDiary daily using adhesion score (AS) from 0 to 4, where 0 = no lifting or small amount of lifting at the edges of the patch; 1 = up to one-quarter of the patch lifting off; 2 = up to half of the patch lifting off; 3 = more than half of the patch lifting off but the patch is still on; and 4 = patch has completely come off.

Current and prior hormonal contraception users were evaluated to classify as:

1. Naïve Users: subjects who had not previously used any hormonal contraceptive
2. Former Users: subjects with previous use of hormonal contraceptives but not in the 6 months prior to enrollment

3. Recent Users: subjects who were not currently using a hormonal contraceptive, but had used a hormonal contraceptive within 6 months prior to enrollment
4. Current Users: subjects who were currently using a hormonal contraceptive.

3.1.1.2 Removal of Patients from Therapy or Assessment

Every subject had the right to refuse further participation in the study at any time without being required to provide reasons for the decision and without prejudice for further treatment. A subject's participation was terminated immediately upon her request. A subject could also be withdrawn from the study at any time at the discretion of the Investigator or Sponsor. If, at the time of discontinuation, at least one dose of study medication had been administered, the subject was asked to follow up for a Final Visit.

Per the Applicant, due to a known increased risk of noncompliance-related pregnancy, subject discontinuation was considered for significant non-compliance as described in the categories below. Investigators considering subject discontinuation for these reasons contacted the Sponsor to discuss the specific case prior to withdrawal of the subject.

- Noncompliance with study medication or eDiary
- Noncompliance with Scheduled Visits (Loss to Follow-up)
- Noncompliance with Study Procedures
- Sexual Activity and Back-Up Contraception Use

If a subject was prematurely discontinued from the study, the reason for discontinuation was selected from the list below and entered in the relevant eCRF.

- Adverse Event
- Death
- Lost to Follow-Up
- Subject Decision
- Pregnancy
- Noncompliance with study medication
- Protocol Violation (violation of inclusion/exclusion criteria, prohibited CMEDs, etc.)
- Investigator Decision
- Sponsor Decision
- Other (appropriate details were to be provided in the eCRF).

3.1.1.3 Endpoints

The primary objective of the study is to evaluate the contraceptive efficacy of AG200-15. The primary efficacy endpoint of this study is on-treatment pregnancy.

Any subject with a positive urine pregnancy test underwent pelvic examination for assessment of uterine size (weeks), serum quantitative ß-hCG testing and prompt ultrasound evaluation to determine the estimated date of conception. Based on the estimated date of conception in relation to

the date of first patch application and/or date of last patch removal, each confirmed pregnancy was classified as:

Pre-treatment pregnancy: estimated date of conception prior to the date of first patch application

On-treatment pregnancy: estimated date of conception from the date of first patch application through Day 7 after the last patch removal

Post-treatment pregnancy: estimated date of conception after Day 7 following removal of the last study patch.

3.1.2 Subject Populations and Analysis Datasets

Safety population included all subjects who wore at least one patch for any period.

Contraceptive efficacy population (CEP) included all subjects who wore at least one patch and were documented to have a negative enrollment serum ß-hCG.

A cycle was defined as a 28-day period consisting of 21 days on treatment (consecutive administration of three 7-day patches) followed by 7 days off treatment (the patch-free week). All complete or incomplete cycles in which the study patch was worn were included in these datasets:

Intent-to-treat (ITT) Efficacy Dataset: All complete or incomplete on-therapy cycles in which intercourse occurred and no back-up contraception was used.

Per-Protocol-Instructions (PPI) Efficacy Dataset: All complete or incomplete on-therapy cycles in which intercourse occurred, excluding two cohorts of cycles:

Cohort 1: cycles in which a back-up method of contraception was used for reasons other than the protocol-specified procedures for missed days of patch use, unless pregnancy occurs; and

Cohort 2: cycles in which the subject missed ≥ 1 day of patch use and did not adhere to the recommended procedures for missed days of patch use.

The PPI efficacy dataset was intended to support assessment of pregnancy rates during treatment cycles in which subjects were compliant with protocol-specified instructions for patch use. On-treatment pregnancies with an estimated date of conception occurring during a cycle in Cohort 2 were therefore not included in the PPI Pearl Index calculation.

Method Failure Efficacy (MF): All complete or incomplete on-therapy cycles in which intercourse occurred, excluding three cohorts of cycles:

Cohort 1: cycles in which a back-up method of contraception was used for reasons other than the protocol-specified procedures for missed days of patch use, unless pregnancy occurs; and

Cohort 2: cycles in which the subject missed ≥ 1 day of patch use and did not adhere to the recommended procedures for missed days of patch use.

Cohort 3: the cycles immediately following Cohort 2 cycles.

The MF dataset was intended to address the increased probability of contraceptive failure in cycles immediately following non-compliant cycles. On-treatment pregnancies with an estimated date of conception occurring during cycles in any of cohorts 1, 2, or 3 were therefore not included in the MF Pearl Index calculation.

3.1.3 Statistical Methodologies

Contraceptive efficacy (pregnancy rates) was evaluated through calculation of Pearl Indices and life table analyses. The Pearl Index, defined as the number of on-treatment pregnancies times 1300 divided by the number of on-therapy cycles, provides an estimate of the number of pregnancies per 100 woman-years of product use.

Primary efficacy analysis

The primary efficacy analysis is based on ITT efficacy dataset which excluded all cycles in which no intercourse occurred and cycles with documented use of back-up contraception, unless the subject became pregnant during the cycle. All on-treatment pregnancies, including pregnancies conceived during cycles in which back-up contraception was used or in which no sexual activity was reported, were included in the analysis. Subjects who turned 36 years of age during the study were not censored from the primary efficacy cohort. In addition to the Pearl Index point estimates, two-sided 95% confidence intervals were also reported.

Secondary Efficacy Analyses

The most important secondary efficacy analyses defined in the protocol were the following:

1. The Pearl Index in subjects aged ≤ 35 years, with BMI < 30 kg/m^2, utilizing the ITT efficacy cycle dataset (unless the subject became pregnant during the cycle)
2. The Pearl Index in subjects aged ≤ 35 years, with BMI < 30 kg/m^2 utilizing the PPI efficacy cycle dataset (unless the subject became pregnant during the cycle)
3. The Pearl Index in subjects aged ≤ 35 years, irrespective of BMI, utilizing the PPI efficacy cycle dataset.

Supportive life table analyses were also used to estimate pregnancy rates for all subjects who were in the contraceptive efficacy population (CEP). The endpoint of primary interest was the cumulative probability of pregnancy at the end of cycle 13. All cycles were included in the life table analysis for subjects who applied at least one patch and were documented to have a negative enrollment serum ß-hCG.

Pearl Indices and cumulative probabilities of pregnancy were evaluated for all contraceptive efficacy datasets (ITT, PPI, and MF) for all subjects in addition to subjects aged ≤ 35 years and for subjects aged 18 to 40 years. Contraceptive efficacy endpoints were also stratified by race/ethnicity, previous contraceptive use status, BMI, and patch application site.

The study was sized to provide 90% power to establish that an underlying Pearl Index no larger than 3.5 and would have an upper bound of a two-sided 95% confidence interval not exceeding 5. It was assumed that each of 1900 enrolled subjects would on average provide 8.5 cycles on treatment, so that the study would generate approximately 16,000 cycles of exposure to AG200-15 in women aged 18 to ≤ 35 years. It was also assumed that close to 21% of these cycles would not be included in the primary evaluation of efficacy due to use of back-up contraception or absence of sexual activity. Thus, a total of approximately 12,675 cycles was used in the power calculations.

As the study proceeded, Agile (or designee) monitored study discontinuation and completion rates and could, if necessary, increase the number of subjects enrolled to achieve an adequate number of efficacy evaluable cycles for estimation of pregnancy rates.

3.1.4 Patient Disposition, Demographic and Baseline Characteristics

As presented in Table 2, a total of 2031 women wore at least one patch for Study ATI-CL23. Of those subjects in the Safety Population, 2,024 had a negative enrollment serum ß-hCG and were included in the Contraceptive Efficacy Population (CEP). Of these, 989 subjects completed the study based on eCRF data. The primary reasons for study discontinuation are "subject decision" (15.3%), "lost to follow-up" (11.3%), and "adverse event" (10.9%). Note that data from study ATI-CL12 is not part of this submission, but only used as reference to show how current study faired with respect to study conduct and deficiencies noted.

Compared to Study ATI-CL12, there are 5.7% more patients in Study ATI-CL23 who completed the study due to the improvement on reducing the lost to follow-up subjects, but the discontinuation due to subject's decision remained high (15.3%). Further, the Sponsor terminated 2% more patients compared to Study ATI-CL12 due to Non-Compliance (5.7% in ATI-CL23 vs. 3.7% in ATI-CL12) and another 18 (0.9%) patients due to Sponsor Decision.

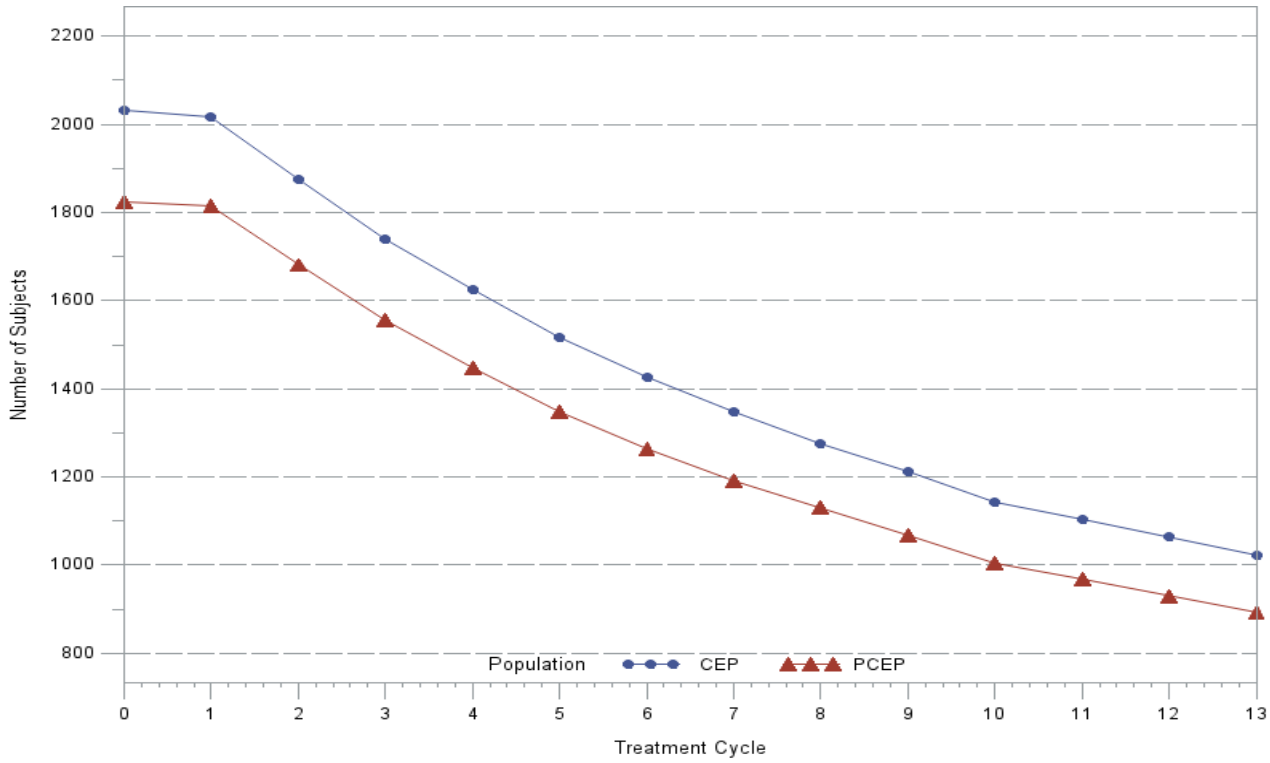
Table 2: Subject Disposition: Study ATI-CL12 and ATI-CL23

Category/Study	ATI-CL12		ATI-CL23	
Safety Population	1,129		2,031	
Contraceptive Efficacy Population	998	88.4%	2,024	99.7%
Completed the Study	485	43.0%	989	48.7%
Discontinued the Study	644	57.0%	1042	51.3%
Reason for Discontinued				
Adverse event	123	10.9%	222	10.9%
Non-Compliance	42	3.7%	116	5.7%
Lost to follow-up	229	20.3%	229	11.3%
Subject's Decision	183	16.2%	310	15.3%
Pregnancy	34	3.0%	73	3.6%
Protocol Violation	11	1.0%	14	0.7%
Investigator Decision	13	1.2%	17	0.8%
Sponsor Decision		0.0%	18	0.9%
Sponsor Decision (Study Termination)		0.0%	2	0.1%
Others	9	0.8%	41	2.0%

* Percentage based on the total number of treated subjects within each corresponding treatment group.
(Source: Clinical Report and Reviewer's Analysis)

Subject disposition by treatment cycle is depicted in Figure 1, showed that subjects withdrew rapidly starting at cycle 2 and more than 50% of the subjects dropped out before the end of the study.

Figure 1 – Subject Disposition by Treatment Cycle: Study ATI-CL23

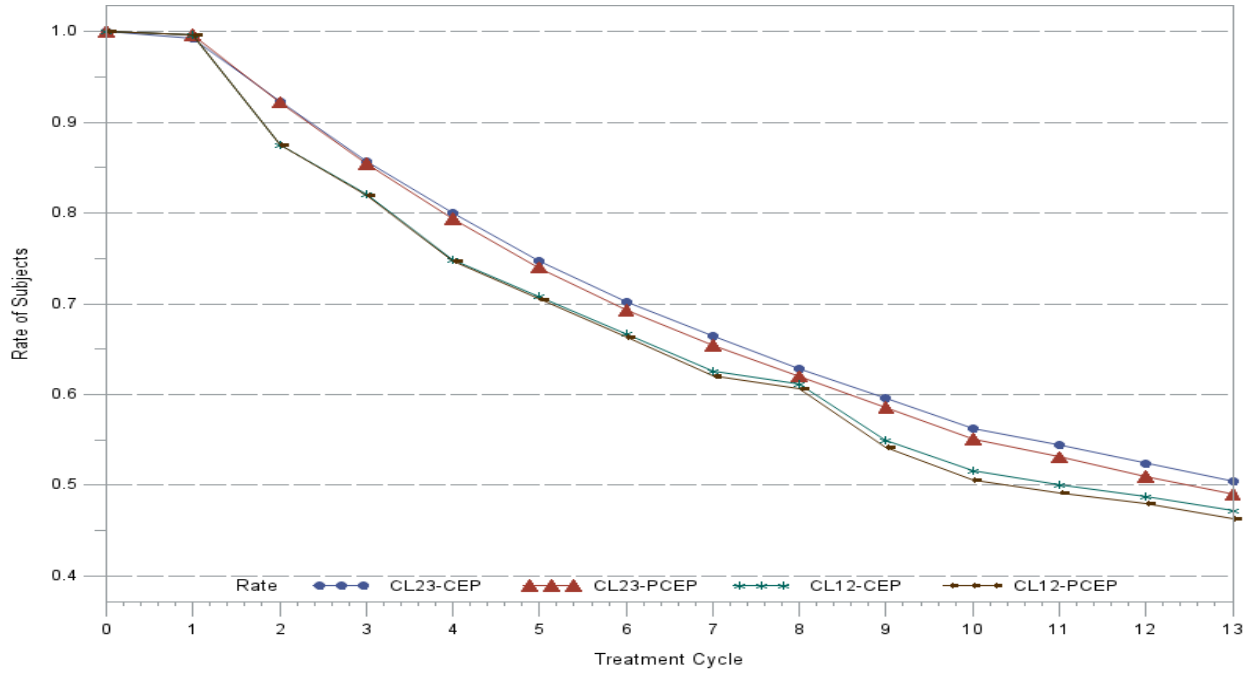


CEP= Contraceptive Efficacy Population, PCEP = CEP with age <=35

(Source: Reviewer's Analysis)

Further, dropout rates in Study ATI-CL23 are compared to previous Study ATI-CL12. As depicted in Figure 2, the Applicant has made slight progress in the retention of subjects. In cycle 2, there are about 5% less dropouts in the current study compared to ATI-CL12. However, the dropout rate remains almost same starting at cycle 3 until the end of the study.

Figure 2 – Completion Rates by Treatment Cycle: Study ATI-CL12 and ATI-CL23



CEP= Contraceptive Efficacy Population, PCEP = CEP with age <=35

(Source: Reviewer's Analysis)

In addition, we compared the early subject withdrawal with those approved hormonal contraceptives (HC) in the past 10 years. As showed in Table 3, among these registration trials for the approved oral contraceptives except for Natazia, Study ATI-23 had relatively high total early subject withdrawal (around 10% more) and high dropout due to Subject Decision (or Withdraw Consents).

Table 3: Early Subject Withdrawal in Twila and Pivotal Study of Recent Approved Oral Contraceptives (Last 10 Years)

Drug/Study	Year Approved	Early Subject Withdrawal			
		Total	Loss to Follow-up	Subject Decision	Others
Study ATI-CL12	NA	57.0%	20.3%	16.2%	0.8%
Study ATI-CL23	NA	51.3%	11.3%	15.3%	2.0%
Quartette	2013	40.4%	13.3%	6.0%	0.8%
Lo Loestrin Fe	2010	41.7%	13.7%	8.9%	5.8%
Natazia	2010	51.3%	13.0%	9.8%	11.6%
Generess	2010	40.9%	16.2%	8.9%	4.5%
LoSeasonique	2008	37.1%	14.1%	8.6%	1.4%

(Source: Reviewer's Analysis)

Exploratory analysis has been conducted by this reviewer to check whether the high dropout rate was due to patch quality. Per CMC reviewer's recommendation, for a patch to have acceptable adhesion, the adhesion score (AS) should be less than 2 throughout the entire wear period. Thus, we defined that a subject had bad patch if the subject had a patch with AS greater than 2 in any wear period. As presented in Table 4, there are about 71% subjects who had bad experience with patch adhesion. However, the percentages of the bad patch are similar between the completers and early withdrawals (70.5% in Completers versus 71.2% in Non-completers). For details regarding patch quality, please see CMC review.

Table 4: Patch Quality by Completer Status: Study ATI-CL23

Category	Number	Bad Patch	% Bad Patch
Contraceptive Efficacy Population	2,022	1,433	70.9%
Completed the Study	988	697	70.5%
Discontinued the Study	1034	736	71.2%

(Source: Reviewer's Analysis)

Demographic and baseline characteristics of Study ATI-CL23 are shown in Table 5. Study ATI-CL23 included 24.3% black subjects and 35.3% obese subjects with BMI $\geq$ 30 kg/m^2. Most subjects were current, recent, or former contraceptive users. Although the Division recommended to enroll sufficient number of truly naïve users of HC, there were only 9.4% subjects considered naïve users. There were 82.7% of the subjects considered to be Naïve to Use of HC Patch.

Table 5: Demographic Characteristics: Study ATI-CL23

Demographic Parameters	Number (N = 2031)	Percent
Age (years)		
≤ 35	1830	90.1%
> 35	201	9.9%
BMI (kg/m^2)		
Non-Obese (<30)	1313	64.6%
Obese (≥ 30)	717	35.3%
Race		
White	1358	66.9%
Black or African American	493	24.3%
Asian	65	3.2%
Other	115	5.7%
Ethnicity		
Hispanic or Latino	400	19.7%
Not Hispanic or Latino	1631	80.3%
Previous HC Use Status		
Naïve	190	9.4%
Former	875	43.1%
Recent	262	12.9%
Current	704	34.7%
Naïve to Use of HC Patch	1680	82.7%

* Percentage based on the total number of treated subjects within each corresponding treatment group.
(Source: Reviewer's Analysis)

3.1.5 Results and Conclusions

The primary efficacy of pregnancy was evaluated by calculating the pregnancy rates by Pearl Index and life table method in subjects less than 35 years old. During the early review of this NDA, Clinical reviewer identified 12 more pregnancies that the Applicant did not include in the efficacy analysis. The summary of the pregnancies is presented in Table 6.

Table 6: Pregnancy Summary for All Treated Subjects: Study ATI-CL23

Confirmed Pregnancies	Applicant	FDA
Pre-Treatment	4	4
On-Treatment	62	74
Post-Treatment	18	8
Overall	84	86

(Source: Reviewer's Analysis)

This reviewer recalculated the efficacy results using the FDA confirmed pregnancies. Table 7 shows both the FDA and the Applicant's analysis results based on ITT efficacy population. For primary analysis, the Applicant reported 56 pregnancies with a Pearl Index of 4.80 (95% Confidence interval: 3.55 to 6.06), while FDA analysis counted 68 (in 18 to <35 years of age) pregnancies with a Pearl Index of 5.83 (95% C.I.: 4.45 to 7.21). For the key secondary analysis in non-obese subjects with BMI < 30 kg/m^2, the Applicant reported 30 pregnancies with a Pearl Index of 3.94 (95% Confidence interval: 2.53 to 5.35), while FDA analysis counted 33 pregnancies with a Pearl Index of 4.34 (95% C.I.: 2.86 to 5.82). For obese subjects with BMI ≥ 30 kg/m^2, the Pearl Index of 8.64 (95% Confidence interval: 5.79 to 11.50) is almost doubled than these of non-obese subjects.

As noted earlier, the Applicant designed this study to demonstrate Pearl Index (PI) no larger than 3.5 within upper bound of a two-sided 95% confidence interval not exceeding 5. As reported in Table 7, even the non-obese women had PI of 4.35 with an upper bound of a two-sided 95% confidence interval of 5.82, both significantly higher than what was expected.

Table 7: Pregnancy Rates for All AG200-15 Treated Subjects (ITT): Study ATI-CL23

Population		N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
Applicant	Age ≤ 35 years	1,736	56	15,165	4.80	(3.55, 6.06)
	Age ≤ 35 with BMI < 30 kg/m^2	1,123	30	9,888	3.94	(2.53, 5.35)
	Age ≤ 35 with BMI ≥ 30 kg/m^2	612	26	5,264	6.42	(3.96, 8.88)
	All Subjects	1,932	62	17,126	4.71	(3.54, 5.88)
Reviewer	Age ≤ 35 years	1,736	68	15,165	5.83	(4.45, 7.21)
	Age ≤ 35 with BMI < 30 kg/m^2	1,123	33	9,888	4.34	(2.86, 5.82)
	Age ≤ 35 with BMI ≥ 30 kg/m^2	612	35	5,264	8.64	(5.79,11.50)
	All Subjects	1,932	74	17,126	5.62	(4.34, 6.89)

(Source: Clinical Report and Reviewer's Analysis)

Further sensitivity analyses were conducted using only FDA confirmed pregnancies on PPI and MF populations by excluding non-compliant and method failure subjects, respectively. As shown in Table 8, Pearl Indices by BMI were twice as much in non-obese (BMI<30 kg/m^2) women compared to obese women (BMI >30 kg/m^2) in both analysis populations. Overall, sensitivity analyses showed that none of the populations had either Pearl Index less than 3.5 or an upper bound of a two-sided 95% confidence interval less than 5.

Table 8: FDA Sensitivity Analysis: Study ATI-CL23

	Population	N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
PPI	Age ≤ 35 years	1,714	57	14,118	5.25	(3.89, 6.61)
	Age ≤ 35 with BMI < 30 kg/m²	1,109	28	9,241	3.94	(2.48, 5.40)
	Age ≤ 35 with BMI ≥ 30 kg/m²	604	29	4,867	7.75	(4.94,10.56)
	All Subjects	1,906	63	15,949	5.14	(3.87, 6.40)
MF	Age ≤ 35 years	1,684	49	13,073	4.87	(3.51, 6.23)
	Age ≤ 35 with BMI < 30 kg/m²	1,094	25	8,583	3.79	(2.30, 5.27)
	Age ≤ 35 with BMI ≥ 30 kg/m²	589	24	4,480	6.96	(4.19, 9.74)
	All Subjects	1,874	55	14,794	4.83	(3.56, 6.11)

(Source: Reviewer's Analyses)

Similar results were shown using Life Table analysis. Table 9 shows the cumulative pregnancy failure rate of 4.22 (95% C.I.: 3.37 to 8.88) by the Applicant, while FDA reports a rate of 5.32 (95% C.I.: 4.20 to 6.74) with an additional twelve pregnancies.

Table 9: Life Table Analysis: Study ATI-CL23

	Population	# of On-Treatment Pregnancies	Cumulative Pregnancy Rate	95% Confidence Interval
Applicant	Age ≤ 35 years	56	4.22	(3.37, 8.88)
	Age ≤ 35 with BMI < 30 kg/m^2	30	3.54	(2.47, 5.07)
	Age ≤ 35 with BMI ≥ 30 kg/m^2	26	5.48	(3.74, 8.00)
	All Subjects	62	4.17	(3.25, 5.34)
FDA	Age ≤ 35 years	68	5.32	(4.20, 6.74)
	Age ≤ 35 with BMI < 30 kg/m^2	33	3.98	(2.82, 5.60)
	Age ≤ 35 with BMI ≥ 30 kg/m^2	35	7.79	(5.62,10.77)
	All Subjects	74	5.13	(4.09, 6.44)

(Source: Clinical Report and Reviewer's Analysis)

The Applicant also compared the pregnancy rates with recent approved hormonal contraceptives (HC) in the past ten years. As showed in Table 10, compared to all the HC which were approved since 2008, PIs or upper bound of 95% CI in Study ATI-23 with or without BMI restriction were much higher: Pearl Index (5.83 versus up to 3.19 with no BMI restriction, 4.34 versus up to 2.92 with BMI restriction) and 95% upper bound (7.21 versus up to 4.03 with no BMI restriction, 5.82 versus up to 4.21 with BMI restriction). Until recently, no HC drug has been approved with Pearl Index (PI) larger than 3.5 or an 95% upper bound CI greater than 5.

Table 10: Pregnancy Rates in Twila and Pivotal Study of Recent Approved Oral Contraceptives (Last 10 Years)

Drug/Study	Year Approved	PI	95% CI	BMI (kg/m^2)
Study ATI-CL12	NA	7.61	(5.39, 9.83)	No Restriction
Study ATI-CL23	NA	5.83	(4.45, 7.21)	No Restriction
Study ATI-CL23	NA	4.34	(2.86, 5.82)	BMI < 30
Quartette	2013	3.19	(2.49, 4.03)	No Restriction
Lo Loestrin Fe	2010	2.92	(1.94, 4.21)	BMI ≤ 35
Natazia	2010	1.64	(NA, 3.82)	BMI ≤ 35
Generess	2010	2.01	(1.21, 3.14)	BMI ≤ 35
LoSeasonique	2008	2.74	(1.92, 3.78)	No Restriction

(Source: Reviewer's Analyses)

In summary, both Pearl Index and life table analyses using either all AG200-15 treated subjects with or without BMI restriction consistently demonstrated that AG200-15 had much higher pregnancy rate than other approved HC drugs in the past ten years.

3.2 Evaluation of Safety

We defer to clinical reviewer's report for safety information.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The efficacy results were based on women aged up to 35 only so the subgroup analyses by gender and age were not necessary for this indication. The phase 3 studies were conducted in the US so no by-region analysis was necessary.

Subgroup analyses by race/ethnicity and previous HC use status using FDA confirmed pregnancies on women age 35 and younger in ITT population were conducted. As presented in **Table 11**, pregnancy rate in current HC users were lower than that of other HC users with a Pearl Index of 3.16 (95% Confidence interval: 1.51 to 4.82). However, no conclusions could be drawn due to limited sample size.

Table 11: Pregnancy Rates by HC use Status (ITT with age ≤ 35): Study ATI-CL23

Population	N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
Race					
White	1159	46	10281	5.82	(4.14, 7.49)
Black or African American	418	17	3454	6.40	(3.36, 9.43)
Other	159	5	1430	4.55	(0.57, 8.52)
Ethnicity					
Hispanic or Latino	330	12	2851	5.47	(2.38, 8.56)
Not Hispanic or Latino	1406	56	12314	5.91	(4.37, 7.46)
Previous HC Use Status					
Naïve	171	5	1415	4.59	(0.57, 8.61)
Former	726	35	6107	7.45	(4.99, 9.91)
Recent	227	14	1888	9.64	(4.61,14.67)
Current	612	14	5755	3.16	(1.51, 4.82)

(Source: Reviewer's Analyses)

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Although the Applicant made efforts to maximize subject continuation and retention, the rates of subject early withdrawal remained high compared with approved hormonal contraceptives (HC) since 2008. The percentage of the women experiencing bad patch were very high (71%), even though the percentages are similar between the completers and early withdrawals. The review team also identified twelve additional pregnancies that the Applicant failed to include in the efficacy evaluation.

Based on the FDA confirmed pregnancies, both Pearl Index and life table analyses using either all AG200-15 treated subjects or non-obese subsects with BMI < 30 kg/m^2 consistently demonstrated that AG200-15 had high pregnancy rate ranging from 4.34 to 5.83 with the upper bound of 95% CI for the point estimate ranging from 5.82 to 7.21. The Pearl indices in obese subjects with BMI > 30 kg/m^2 were approximately two-fold compared to non-obese subjects.

5.2 Conclusions and Recommendations

The Applicant reported efficacy results based on pre-specified endpoint and statistical methods. Both the Pearl Index and life table method consistently showed much higher pregnancy rate in this population than seen in support of other hormonal contraceptives. Note that despite more efforts had been made by the Applicant to minimize subject early withdrawal and improve patch quality, the contraceptive failure rate continues to remain high in the current study ATI-CL23, although rates are lower than these of previous study ATI-CL12. The Applicant failed to demonstrate an acceptable Pearl Index and upper bound of the 95% confidence interval compared to recent approved HC drugs.

From a statistical perspective, given the conduct of the study, poor patch quality and high pregnancy rate based on pre-specified analyses and the comparison to recent approved HC drugs, the evidence from this newly conducted trial (ATI-CL23) is not sufficient to support the approval of AG200-15 in the prevention of pregnancy.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

KATE L DWYER
12/09/2017

MAHBOOB SOBHAN
12/09/2017

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA Serial Number:	204017 / N000
Drug Name:	AG200-15 (2.60 mg levonorgestrel / 2.30 mg ethinyl estradiol) transdermal delivery system
Indication(s):	Prevention of Pregnancy
Applicant:	Agile Therapeutics, Inc.
Date(s):	Submission Date: 4/13/2012
	PDUFA Due Date: 2/13/2013
Review Priority:	Standard
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Kate Dwyer, Ph.D.
Concurring Reviewer:	Mahboob Sobhan, Ph.D., Team Leader
Medical Division:	Division of Reproductive and Urologic Drug Products
Clinical Team:	Dan Davis, M.D., Medical Reviewer
	Lisa Soule, M.D., Team Leader
Project Manager:	Charlene Williamson

Keywords: NDA review, clinical studies, sensitivity analyses, compliance

Table of Contents

LIST OF TABLES

1 EXECUTIVE SUMMARY

The Applicant is seeking approval of AG200-15, a 7-day transdermal contraceptive delivery system for the prevention of pregnancy. To support the efficacy and safety, two studies (ATI-CL12 and ATI-CL13) were submitted in this application. Study ATI-CL12 was conducted to evaluate the contraceptive efficacy, while study ATI-CL13 was conducted to compare the safety profile compared to a low dose oral contraceptive. This review only evaluates the efficacy data from a statistical perspective in order to determine the robustness of the efficacy of AG200-15 in the prevention of pregnancy.

Study ATI-CL12 was a multi-center, open-label, randomized, comparative, parallel group study conducted in healthy, sexually active women requesting contraception. AG200-15 is a 7-day transdermal (patch) contraceptive delivery system (TCDS) containing 2.60 mg levonorgestrel (LNG) and 2.30 mg ethinyl estradiol (EE) for the prevention of pregnancy in women. The patch is replaced every 7 days for 3 weeks, followed by a 1-week "patch-free" period. The primary efficacy objective of the study was to evaluate the contraceptive efficacy of AG200-15 compared to a low-dose oral contraceptive (OC). The pre-specified, primary analysis was based on the Pearl Index for all in-treatment pregnancies in women aged 17 to 35 years with a BMI less than 32 kg/m^2 with intercourse and no other use of birth control methods. Because no between-treatment efficacy comparisons were planned, this review focuses only on the AG200-15 treatment arm.

There were no statistical issues identified in this submission. However, two clinical issues regarding the pregnancy outcome and analysis population were noted during the review cycle. First, the clinical team has identified eight additional pregnancies that were not included in the Applicant's analysis. Secondly, the Applicant amended the analysis datasets to exclude approximately 15% of the laboratory-verified non-compliant subjects from the analysis datasets. The Applicant's argument was that these subjects did not reflect the relevant study population who are actively seeking hormonal contraception. However, this amendment was neither pre-specified in the statistical analysis plan nor agreed to by the Division. Therefore, this review reports efficacy results based on additional pregnancies identified by FDA without excluding any non-compliant subjects *post-hoc*.

In study ATI-CL12, the Pearl Index, based on women aged 17 to 35 years with a BMI less than 32 kg/m^2 , with no intercourse and other use of birth control methods was 7.50 (95% confidence interval: 5.02 to 9.97). The point estimate of 7.50 for the Pearl Index and the upper bound of its 95% confidence interval of 9.97 are larger than those seen in recently approved OC products. From a statistical perspective, the pre-specified analyses based on the Pearl Index are appropriate and the results are valid. However, the clinical utility and the approvability of AG200-15 in the prevention of pregnancy based on such high failure rate (Pearl Index) is the Clinical Division's decision.

2 INTRODUCTION

2.1 Overview

The Applicant, Agile Therapeutics, Inc., is seeking approval of AG200-15, a 7-day transdermal contraceptive delivery system (TCDS) containing 2.60 mg levonorgestrel (LNG) and 2.30 mg ethinyl estradiol (EE) for the prevention of pregnancy.

The efficacy and safety of AG200-15 was assessed in two phase 3 studies. Studies ATI-CL12 and ATI-CL13 were both multi-center, open-label, randomized, comparative, parallel group studies conducted in healthy, sexually active women requesting contraception. Study ATI-CL12 was considered the pivotal efficacy study, while study ATI-CL13 was supportive for safety purposes. Table 1 presents a brief summary of the two studies.

Table 1: Brief summary of the phase 3 studies for AG200-15

Study Number (No. of Sites / Country) Dates of Study Conduct	Subject Population	Treatments	Sample Size	Duration of Treatment	Design[1]
ATI-CL12 (102 / U.S.) 08-05-10 to 11-03-11	Heterosexually active, 17-40 year-old females who were at risk of pregnancy	OC20 AG200-15	375 1129	Cycles 1-6 Cycles 1-13	OL, R, AC, MC
ATI-CL13 (21 / U.S.) 10-28-10 to 06-24-11	Heterosexually active, 17-40 year-old females who were at risk of pregnancy with BMI<32kg/m^2	OC20 AG200-15	206 201	Cycles 1-6 Cycles 1-6	

[1] OL = Open Label, R = Randomized, AC = Active Controlled, MC = Multicenter

2.2 Data Sources

The study reports and the data sets were submitted electronically to the Electronic Document Room. The SAS data sets were complete and well documented.

The study reports are located at:

\\Cdsesub1\evsprod\NDA204017\0000\m5\53-clin-stud-rep\535-rep-effic-safety-stud

The datasets and programs for study ATI-CL12 are located at:

\\Cdsesub1\evsprod\NDA204017\0000\m5\datasets\ati-cl12

The datasets and programs for study ATI-CL13 are located at:

\\Cdsesub1\evsprod\NDA204017\0000\m5\datasets\ati-cl13

3 STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

Study ATI-CL12 was an open-label, randomized, comparative, parallel group, multicenter, phase 3 study to evaluate the contraceptive efficacy and safety of AG200-15 compared to a low-dose oral contraceptive (OC) containing 20 µg EE and 100 µg LNG in a 21-day regimen (consecutive administration of three 7-day patches or 21 days of active pill-taking) followed by 7 days off treatment (i.e., no patch is applied or no active pills are taken).

Sexually active women, aged 17 to 40 years, were randomized in a 3:1 ratio to either AG200-15 or OC. Those randomized to AG200-15 were treated for 13 28-day cycles (1 year). Those randomized to OC were treated for six 28-day cycles and then were switched to AG200-15 for an additional seven cycles of treatment. Randomization was stratified by BMI (<30, 30 to <35, ≥35 kg/m^2) and contraceptive use status (new user, current user) with a centralized stratified randomization scheme. Enrollment in each treatment group was to have up to 33% of subjects having a BMI ≥ 30 kg/m^2 and with 50% of these subjects having a BMI ≥ 35 kg/m^2.

The primary objectives of study ATI-CL12 were to evaluate the contraceptive efficacy and safety of AG200-15 patch compared to a low-dose OC. Contraceptive efficacy was evaluated using pregnancy rates that were estimated using all in-treatment pregnancies, defined as those with an estimated date of conception from the date of first patch application through day 14 after the last patch removal. Although the primary study objectives were comparative in nature, the Division considered demonstration of the efficacy based only on the AG200-15 patch's ability to prevent pregnancies.

Study ATI-CL13 was an open-label, randomized, active-controlled, parallel group, multicenter, 6-cycle phase 3 study of the contraceptive efficacy and safety of AG200-15 compared to an OC containing 150 µg LNG and 30 µg EE. Only subjects with a BMI <32 kg/m^2 were recruited. Although the design of study ATI-CL13 was similar to study ATI-CL12, it was a smaller study whose purpose was to compare the adverse event profile with that of a low-dose OC.

The remainder of this review focuses on study ATI-CL12 for demonstration of efficacy based only on the AG200-15 treatment arm to prevent pregnancies.

3.1.2 Statistical Methodologies

The efficacy was demonstrated by calculating the Pearl Index and cumulative pregnancy rate for all in-treatment pregnancies in women aged 17 to 35 years with BMI <32 kg/m^2. The Pearl Index was the number of on-therapy pregnancies times 1300 divided by the number of 28-day on-therapy cycles.

Efficacy was evaluated using the following two intent-to-treat populations (ITT1 and ITT2) and two method failure populations (MF1 and MF2). The ITT2 population which included women aged 17 to 35 years with a BMI < 32 kg/m^2 was considered the primary analysis dataset.

Intent-to-treat (ITT) Population:

A. ITT1: Population consisted of all subjects with complete/incomplete on-therapy cycles where all in-treatment pregnancies were included in the calculation of pregnancy rates.

B. ITT2: Population consisted of all subjects aged 17-35 years with a BMI <32 kg/m^2 with complete/incomplete on-therapy cycles with intercourse and no other use of birth control methods. All in-treatment pregnancies were included in the calculation of pregnancy rates.

Method Failure (MF) Population Efficacy Datasets:

A. MF1: Population consisted of all subjects with complete/incomplete on-therapy cycles with intercourse excluding two cohorts: (1) cycles with use of other birth control methods; (2) cycles during which the subject missed more than one day of active drug-taking and immediately following cycles. In-treatment pregnancies with an estimated date of conception attributed to the cycles from the first and second cohorts were not included in the calculation of pregnancy rates.

B. MF2: Population consisted of all subjects with complete/incomplete on-therapy cycles with intercourse excluding two cohorts: (1) cycles with other birth control method used for reasons other than missed days of drug taking unless pregnancy occurs; and (2) cycles during which the subject missed one or more days of active drug taking and did not adhere to the procedures recommended for the missed days of drug taking as well as immediately following cycles. In-treatment pregnancies with an estimated date of conception attributed to the cycles from the first and second cohorts were not included in the calculation of the pregnancy rates.

The Applicant identified high number of laboratory verified non-compliant subjects (10%-14%), defined as subjects with undetectable concentrations for both EE and LNG at one or more study visits, and decided to amend the above datasets to exclude these subjects. These datasets were not pre-specified in the statistical analysis plan and not agreed to by the Division; therefore, their results are not presented in this review.

3.1.3 Patient Disposition, Demographic and Baseline Characteristics

As presented in Table 2, a total of 1,129 women were randomized to the AG200-15 group for 13 cycles of treatment and 375 women were randomized to the oral contraceptive group for six cycles of treatment and were then switched to AG200-15 for an additional seven cycles of treatment. Of the 375 subjects in the OC group, 249 subjects were switched to AG200-15 for an additional seven cycles of treatment.

Discontinuation information is also presented in Table 2. For the AG200-15 treatment group, 436 (38.6%) subjects discontinued the study before the end of cycle 6 and an additional 208 (18.4%) subjects discontinued the study after cycle 6 and before the end of cycle 13. The primary reasons for

study discontinuation in the AG200-15 treatment group are "lost to follow-up" (20.3%), "subject decision" (16.2%) and "adverse event" (10.9%).

Table 2: Subject Disposition: Study ATI-CL12

Category	OC		AG200-15					
			AG200-15		OC/AG200-15		AG200-15 Total	
Safety Population	375		1,129		249		1,378	
ITT1	330	88.0%	998	88.4%	228	91.6%	1,226	89.0%
Completed the Study	249	66.4%	485	43.0%	150	60.2%	635	46.1%
Discontinued the Study up to Cycle 6	126	33.6%	436	38.6%				
Discontinued the Study up to Cycle 13			644	57.0%	99	39.8%	743	53.9%
Reason for Discontinued								
Lost to Follow-up	61	16.3%	229	20.3%	37	14.9%	266	19.3%
Subject's Decision	30	8.0%	183	16.2%	28	11.2%	211	15.3%
Adverse event	16	4.3%	123	10.9%	16	6.4%	139	10.1%
Non-Compliance	9	2.4%	42	3.7%	6	2.4%	48	3.5%
Pregnancy	4	1.1%	34	3.0%	8	3.2%	42	3.0%
Protocol Violation	4	1.1%	11	1.0%	0	0.0%	11	0.8%
Investigator's Decision	1	0.3%	13	1.2%	2	0.8%	15	1.1%
Others	1	0.3%	9	0.8%	2	0.8%	11	0.8%

* Percentage based on the total number of treated subjects within each corresponding treatment group.
(Source: Reviewer's Analysis)

Overall in the ITT1 population, subjects had a mean age of 26.0 years (range of 17 to 40 years), 55.1% were new starts, and the majority (72.9%) was Caucasian. The AG200-15 group included 21.2% black subjects. In the ITT2 population (Table 3), 92% (863/940) of the AG200-15 treatment group subjects were less than 35 years of age. In addition, 72% (675/940) of the ITT2 population were less than 35 years of age and had a BMI < 32 kg/m^2 included in the primary analysis for those in the AG200-15 treatment group. ITT2 with 17 to 35 year-old women with a BMI< 32 kg/m^2 is considered to be the primary analysis dataset.

Table 3: Analysis Populations: Study ATI-CL12

Population	AG200-15		OC/AG200-15		AG200-15 Total	
ITT1	998		228		1,226	
ITT2	940	94.2%	218	95.6%	1,158	94.5%
ITT2 with Age 17-35	863	86.5%	197	86.4%	1,060	86.5%
ITT2 with Age 17-35 and BMI < 32 kg/m^2	**675**	**67.6%**	**152**	**66.7%**	**827**	**67.5%**
ITT2 with Age 17-35 and BMI < 30 kg/m^2	621	62.2%	146	64.0%	767	62.6%

* Percentage based on the total number of treated subjects within each corresponding treatment group.
(Source: Reviewer's Analysis)

3.1.4 Results and Conclusions

The primary efficacy of pregnancy was evaluated by calculating the failure rates by Pearl Index and life table method. Table 4 shows both the FDA and the Applicant's analysis results based on ITT2 population. For primary analysis, the Applicant reported 30 pregnancies with a Pearl Index of 6.34 (95% Confidence interval: 4.13 to 8.27), while FDA analysis counted 35 pregnancies with a Pearl Index of 7.50 (95% C.I.: 5.02 to 9.97). The Pearl Index excluding switchers from OC to AG200-15 after cycle 6, is 7.14 (95% C.I.: 4.55 to 9.74) as shown in Table 5.

Table 4: Pregnancy Rates for All AG200-15 Treated Subjects: ITT2 Population, Study ATI-CL12

	Population	N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
Applicant	Age 17-35 years	1,060	38	7,685	6.43	(4.39, 8.47)
	Age 17-35 with BMI < 32 kg/m^2	**827**	**30**	**6,070**	**6.43**	**(4.13, 8.27)**
	Age 17-35 with BMI < 30 kg/m^2	767	27	5,645	6.22	(3.88, 8.56)
	All Subjects	1,158	40	8,446	6.16	(4.25, 8.06)
Reviewer	Age 17-35 years	1,060	45	7,685	7.61	(5.39, 9.83)
	Age 17-35 with BMI < 32 kg/m^2	**827**	**35**	**6,070**	**7.50**	**(5.02, 9.97)**
	Age 17-35 with BMI < 30 kg/m^2	767	32	5,645	7.37	(4.82, 9.92)
	All Subjects	1,158	48	8,446	7.39	(5.30, 9.47)

(Source: Reviewer's Analyses & Adapted from Clinical Study ATI-CL12 Report; Table 9, page 83)

Table 5: FDA Sensitivity Analysis Excluding Switchers: ITT2 Population, Study ATI-CL12

	Population	N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
Using Applicant Reported Pregnancies	Age 17-35 years	863	32	6,673	6.23	(4.08, 8.39)
	Age 17-35 with BMI < 32 kg/m^2	**675**	**25**	**5,278**	**6.16**	**(3.75, 8.57)**
	Age 17-35 with BMI < 30 kg/m^2	621	22	4,885	5.85	(3.41, 8.30)
	All Subjects	940	33	7,320	5.86	(3.87, 7.86)
Using Agency Indentified Pregnancies	Age 17-35 years	863	37	6,673	7.21	(4.89, 9.52)
	Age 17-35 with BMI < 32 kg/m^2	**675**	**29**	**5,278**	**7.14**	**(4.55, 9.74)**
	Age 17-35 with BMI < 30 kg/m^2	621	26	4,885	6.92	(4.27, 9.57)
	All Subjects	940	39	7,320	6.93	(4.76, 9.09)

(Source: Reviewer's Analyses)

The life table analysis included only the subjects who received AG200-15, not the switchers (OC to AG200-15 after cycle 6). Table 6 shows cumulative pregnancy rate of 6.12 (95% C.I.: 3.37 to 8.88) by the Applicant, while FDA reports a rate of 7.30 (95% C.I.: 4.31 to 10.29) with an additional four pregnancies.

Table 6: Life Table Analysis: ITT2 Population, Age 17-35 with BMI< 32 kg/m^2, Study ATI-CL12

		Number of On-Treatment Pregnancies	Cumulative Pregnancy Rate	95% Confidence Interval
Applicant	Cycles 1-6	12	2.12	(0.85, 3.39)
	Cycles 1-13	**25**	**5.25**	**(2.87, 7.63)**
Reviewer	Cycles 1-6	12	2.12	(0.85, 3.39)
	Cycles 1-13	**29**	**6.27**	**(3.68, 8.86)**

(Source: Reviewer's Analyses & Adapted from Clinical Study ATI-CL12 Report; Table 14, page 107)

Both Pearl index and life table analyses using either all AG200-15 treated subjects or excluding switchers consistently demonstrated that AG200-15 had high pregnancy rate raging from 7.14 to 7.50 with the upper bound of 95% CI for the point estimate ranging from 9.74 to 10.29. FDA analysis also confirmed the Applicant's results using the MF1 and MF2 analysis populations (Table 8, Appendix). The point estimates of these exploratory analyses were lower than the primary analysis population due to the fact that additional pregnancies that FDA identified during the review were not included in these analyses. Nevertheless, the upper bound of the 95% CI for the point estimate was 7.63 in women aged 17 to 35 with BMI of < 32 kg/m^2.

3.2 Evaluation of Safety

Safety information can be found in the clinical reviewer's report.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region

The efficacy results were based on women aged 17 to 35 only so the subgroup analyses by gender and age were not necessary for this indication. The phase 3 studies were conducted in the US so no by-region analysis was necessary.

Subgroup analyses by race were presented in Table 7. Pregnancy rate in white (not Hispanic) subjects were lower than that of other races.

Table 7: Pregnancy Rates by Race: ITT2 Population, Study ATI-CL12

Population		N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
Sponsor	White (not Hispanic)	512	12	3,749	4.16	(1.81, 6.51)
	White (Hispanic)	124	7	924	9.85	(2.58, 17.12)
	Black	137	8	960	10.83	(3.36, 18.31)
	Other	54	3	437	8.92	(0.00, 18.99)
Reviewer	White (not Hispanic)	512	13	3,749	4.51	(2.06, 6.95)
	White (Hispanic)	124	8	924	11.26	(3.49, 19.02)
	Black	137	11	960	14.90	(6.14, 23.65)
	Other	54	3	437	8.92	(0.00, 18.99)

(Source: Reviewer's Analyses & Adapted from Clinical Study ATI-CL12 Report; Table 10, page 87)

4.2 BMI Subgroup Populations

Efficacy subgroup analysis by BMI was evaluated and these results are presented in Section 3.1.4.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

There were no statistical issues in this submission. However, the Clinical Reviewer identified eight additional pregnancies in study ATI-CL12. Therefore, these additional pregnancies were included in the efficacy evaluation.

Efficacy was evaluated by the pregnancy rate based on the Pearl Index in women aged 17 to 35 years with a BMI <32 kg/m^2 and excluding cycles with no intercourse and where other birth control methods were used. In study ATI-CL12, the Pearl Index for the AG200-15 patch was 7.50% (95% Confidence Interval: 5.02% to 9.97%). Although the Applicant amended the datasets to exclude laboratory verified non-compliant subjects, these datasets were not pre-specified in the statistical analysis plan and conclusions cannot be based on these *post-hoc* datasets.

5.2 Conclusions and Recommendations

From a statistical perspective, the Applicant reported efficacy results based on pre-specified endpoint and statistical methods. Both the Pearl Index and life table method consistently showed much higher pregnancy rate in this population than seen in support of other contraceptives. However, the use of AG200-15 in the prevention of pregnancy with large Pearl Index is the clinical Division's decision.

6 APPENDIX

Table 8: Pregnancy Rates for All AG200-15 Treated Subjects: Study ATI-CL12

	Population	N	# of On-Treatment Pregnancies	Number of Cycles	Pearl Index	95% Confidence Interval
MF1	Age 17-35 years	1,029	30	7,240	5.39	(3.46, 7.31)
	Age 17-35 with BMI < 32 kg/m^2	**873**	**24**	**5,722**	**5.45**	**(3.28, 7.63)**
	Age 17-35 with BMI < 30 kg/m^2	805	21	5,316	5.14	(2.94, 7.33)
	All Subjects	1,126	31	7,954	5.07	(3.29, 6.85)
MF2	Age 17-35 years	1,021	27	7,176	4.89	(3.05, 6.73)
	Age 17-35 with BMI < 32 kg/m^2	**869**	**21**	**5,663**	**4.82**	**(2.76, 6.88)**
	Age 17-35 with BMI < 30 kg/m^2	804	19	5,259	4.70	(2.59, 6.80)
	All Subjects	1,119	28	7,874	4.62	(2.91, 6.33)

(Source: Reviewer's Analyses & Adapted from Clinical Study ATI-CL12 Report; Table 12, page 94)

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

KATE L DWYER
01/22/2013

MAHBOOB SOBHAN
01/22/2013