CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

201110Orig1s000

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/SN #:	201110/53
Drug Name:	Milprosa (progesterone) vaginal ring
Indication(s):	To support embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an assisted reproductive technology (ART) treatment in infertile women
Safety Issue(s):	Spontaneous abortion
Applicant:	Ferring Pharmaceutical Inc.
Date(s):	Stamp date: October 29, 2019 PDUFA date: April 29, 2019 Completion Date: April 8, 2020
Review Priority:	Standard
Biometrics Division:	Division of Biometrics VII
Statistical Reviewer:	Thanh Van Tran, PhD
Concurring Reviewers:	Clara Y. Kim, PhD, Team Leader Mat Soukup, PhD, Deputy Division Director
Consulting Division:	Division of Urology, Obstetrics, and Gynecology Products (DUOG)
Consult Request Team:	Regina Zopf, MD; Shelley Slaughter, MD, Team Leader
Project Manager:	Nikia Morris

Keywords: spontaneous abortion, progesterone vaginal ring, assisted reproductive technology, non-inferiority test, Bayesian statistics, single arm trial, open label

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

Milprosa (progesterone) vaginal ring (PVR), indicated to support embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an assisted reproductive technology (ART) treatment in infertile women, was resubmitted on October 29, 2019, by Ferring Pharmaceuticals, Inc; the first submission was in 2010. The resubmission aimed to address a complete response letter that requested a biocompatibility study and a safety study to establish a clinical safety bridge between the legacy PVR (studied in a phase 3 clinical trial) and the new PVR. FDA recommended the safety study to evaluate the clinical safety and tolerability of the new PVR over the entire treatment duration (up to 10 weeks post-embryo transfer) in women undergoing ART procedures. In this document, we review the results of safety trial 000293.

Trial 000293 was a multi-center, open label, single arm, safety trial in the United States of women 18-34 years old treated once weekly with PVR for up to 10 weeks. The primary outcome was any spontaneous abortion occurring on or before 12 weeks following oocyte retrieval. Spontaneous abortion was defined as two positive β-hCG tests occurring at least two days apart on or after two weeks post-oocyte retrieval, followed by observation of blighted ovum, intrauterine gestation without a fetal heartbeat, or absence of viable fetuses, as documented by transvaginal ultrasound (TVUS). The primary outcome was analyzed for the modified intention-to-treat (mITT) population, defined as all subjects treated with PVR and completed embryo transfer. To establish the clinical safety bridge, the study was designed to rule out a 15% spontaneous abortion rate.

Among 254 subjects treated with PVR (safety population), 243 subjects were eligible for the mITT population. By the end of the trial, 239 subjects had complete information to determine their status for spontaneous abortion. The remaining four subjects with incomplete information were as follows: one subject withdrew consent, one subject was lost to follow-up, two subjects exited the trial after β-hCG testing but before TVUS were completed.

The mITT population's mean age was 31 years, which ranged from 22 to 34 years, and 70% of subjects were at least 30 years old. Race was mainly comprised of 82% White, 9% Asian, and 7% Black/African American. The average duration of infertility was 34 months with standard deviation 24 months.

In the mITT population of 243 subjects, there were 18 spontaneous abortions. The percentage of spontaneous abortion was 7.4% (95% CI 4.4, 11.5). A 15% spontaneous abortion rate was ruled out with the upper bound of the CI. Post-hoc sensitivity analysis to missing data of four patients whose status for spontaneous abortion could not be determined resulted in similar conclusions to the primary analysis.

Among adverse secondary outcomes in the mITT population, biochemical abortion at 10 weeks (25 events, 10.3%; 95% CI 6.8, 14.8) was more frequent but expected clinically. In the safety population, 49% (124/254) of subjects had 312 treatment emergent adverse events (TEAEs), with 0 events leading to death, 44 events leading by PVR discontinuation in 44 (17%) subjects, and 53 adverse drug reactions, TEAEs that an investigator judged to have reasonable possibility

of causality by the PVR, in 25 (10%) subjects. Inspection of TEAEs by preferred term, frequency, and intensity showed that the results were consistent with the progesterone class and early pregnancy.

Although there were subjects with extreme changes from baseline for laboratory parameters alanine aminotransferase and aspartate aminotransferase and blood pressure, the frequencies were low (1-3 subjects).

In conclusion, this single arm, safety trial ruled out a 15% spontaneous abortion rate. There were no major safety concerns for other pregnancy outcomes, adverse events, laboratory tests, and blood pressure.

2 INTRODUCTION

2.1 Overview

Milprosa (progesterone) vaginal ring (PVR), indicated to support embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an assisted reproductive technology (ART) treatment in infertile women, was submitted in April 2010 under NDA 201110 by Ferring Pharmaceuticals, Inc. The NDA received a complete response letter (CRL); one of the main issues being the presence of (b) (4) particulates contamination in the PVR.

In 2016, the sponsor resubmitted the application addressing particulate contamination by altering the manufacturing process with the use of a new (b) (4) process. A bioequivalence study showed that the legacy PVR (studied in a phase 3 clinical trial) and the new PVR manufactured with the new (b) (4) process were similar. Upon review of the resubmission, FDA issued another CRL that included a request for the sponsor to conduct a biocompatibility study and a safety study to establish a clinical safety bridge between the legacy PVR and the new PVR. FDA recommended the safety study to evaluate the clinical safety and tolerability of the new PVR over the entire treatment duration (up to 10 weeks post-embryo transfer) in women undergoing ART procedures.

On January 23, 2017, the sponsor proposed a multi-center, open label, single arm trial titled, "A Prospective, Multi-Center, Non-Comparative Trial of the Clinical Safety of the Progesterone Vaginal Ring in Women Undergoing Assisted Reproductive Technology (ART) Procedures." After multiple exchanges between the sponsor and FDA over clinical and statistical concerns with subsequent protocol versions and associated statistical analysis plans (first version submitted October 23, 2018), FDA asked the sponsor to propose a frequentist analysis to rule out a spontaneous abortion rate of 15% with at least 80% power and a one-sided type I error of 0.025 in an advice letter, dated July 24, 2019.

The sponsor agreed to FDA's recommendation of a frequentist study design and resubmitted a revised protocol (version 6.0) and SAP (version 4.0) on August 12, 2019. A Bayesian analysis would serve as a sensitivity analysis. A resubmission of the application on October 29, 2019, included results from biocompatibility tests and trial 000293. DB7 provided comments in an information request, dated December 27, 2019, asking about data missingness and a statistical test formula.

The following review focuses on trial 000293's statistical analysis methods and study results.

2.2 Data Sources

The location of the data in SDTM and ADaM formats are:

\\cdsesub1\evsprod\NDA201110\0047\m5\datasets\000293\analysis\adam
\\cdsesub1\evsprod\NDA201110\0047\m5\datasets\000293\tabulations\sdtm

Software code in SAS and R were submitted by the sponsor.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

There were formatting issues when converting certain ADaM data sets from .xpt format to .Rdata format. The Office of Computational Science Help Desk assisted with this issue. Other than this issue, the submitted data were of sufficient quality to assess the sponsor's statistical methods, analyses, and interpretation.

3.2 Evaluation of Safety

This review is focused entirely on safety.

3.2.1 Study Design

Trial 000293 was a multi-center, open label, single arm safety trial for luteal phase support in women undergoing in vitro fertilization in the United States. Pre-menopausal females 18-34 years old at the time of consent with tubal, idiopathic, male factor, ovulatory dysfunction, or endometriosis-linked infertility were included in the trial. The trial duration was approximately one year, July 2018 to July 2019.

To establish the clinical safety bridge, the safety trial was designed to rule out a 15% spontaneous abortion rate with at least 80% power using a sample size of at least 240 subjects treated with PVR following fresh embryo transfer and assuming a true spontaneous abortion rate of 9% and a one-sided type 1 error of 0.025.

The trial procedures were as follows:

1. Ovarian stimulation
2. Oocyte retrieval
3. Initiation of treatment with PVR. Throughout the trial, PVRs were inserted once weekly for at most 10 weeks
4. Embryo transfer
5. At predetermined intervals after embryo transfer, β-hCG was measured to determine PVR discontinuation, and ultrasound exams were performed to determine pregnancy and PVR discontinuation
6. End of trial 12 weeks after oocyte retrieval

Adverse events and vital signs were monitored throughout the trial. Treatment was stopped if the subject had a failed attempt at becoming pregnant or found to be no longer pregnant. The study was designed to be similar to legacy phase 3 trial DR-PGN-302 with minor differences that the sponsor claimed should have minimal impact on the primary outcome.

There were three analysis populations:

- Modified intention-to-treat (mITT) population consisted of all subjects who had successful oocyte retrieval, received at least one dose of PVR, and had completed embryo transfer.

- Safety population consisted of all subjects treated with PVR.

- Eligible subjects population consisted of all subjects who were eligible for trial participation based on inclusion and exclusion criteria at screening and treated with oral contraceptives, leuprolide acetate, or MENOPUR.

3.2.2 Study Outcomes

The primary outcome was any spontaneous abortion occurring on or before 12 weeks following oocyte retrieval in the mITT population. Spontaneous abortion was defined as two positive β-hCG tests occurring at least two days apart on or after two weeks post-oocyte retrieval, followed by observation of blighted ovum, intrauterine gestation without a fetal heartbeat, or absence of viable fetuses, as documented by transvaginal ultrasound (TVUS).

Secondary outcomes included:

- Spontaneous abortions determined at 6 and 10 weeks post-oocyte retrieval in all subjects treated with PVR

- Biochemical abortions determined at 6 and 10 weeks post-oocyte retrieval in all subjects treated with PVR. Biochemical abortion was defined as two positive β-hCG tests (at 2 weeks and 2 weeks + 3-4 days post-oocyte retrieval) but followed by no observed gestational sac on a later TVUS, or followed by a negative β-hCG test (e.g., at visit 7, 3 weeks post-oocyte retrieval). Additionally, a biochemical abortion could be defined by an adverse event that met the definition of biochemical abortion (e.g., subjects with declining β-hCG, a clear indication of biochemical abortion).

- Ectopic and heterotopic pregnancies in all subjects treated with PVR following oocyte retrieval

- Positive β-hCG tests at 2 weeks and 2 weeks + 3-4 days post-oocyte retrieval in all subjects treated with PVR

- Clinical pregnancy (TVUS showing at least 1 intrauterine gestational sac with fetal heartbeat) at 6 and 10 weeks post-oocyte retrieval in all subjects treated with PVR

Subjects who did not have two β-hCG tests due to missing data, early withdrawal, or other reasons were classified as not having a positive β-hCG unless a positive result is observed at a later pregnancy assessment. For example, if the result of a β-hCG test was missing and a subsequent TVUS confirmed clinical pregnancy, then the β-hCG test was imputed as 'positive.'

Other safety outcomes included:

- Treatment emergent adverse events (TEAEs), defined as adverse events that occur during the treatment phase or pre-existing medical conditions that worsen in intensity during the treatment phase, for all subjects treated with PVR. Adverse drug reactions referred to TEAEs that an investigator judged to have reasonable possibility of causality by the PVR

- Vaginal bleeding/spotting, vaginal hemorrhage, pain, vaginal infection, and vaginal irritation for all subjects treated with PVR

- TEAEs associated with vaginal or cervical abrasions and lesions and TEAEs associated with vaginal adhesions for all subjects treated with PVR

- Reasons for PVR discontinuation

- Abnormal findings in clinical laboratory tests and vital signs for all subjects treated with PVR

3.2.3 Statistical Methods

The prespecified primary analysis was the estimation of the proportion of any spontaneous abortion and the Clopper-Pearson two-sided exact (binomial distribution) 95% confidence interval (CI) for the mITT population. The upper bound of the CI was compared to 15% to establish the safety clinical bridge between the legacy PVR and the new PVR.

Secondary outcomes were summarized by frequency and percent for the mITT population and safety population.

Adverse events were summarized by frequency and percent for the safety population and eligible subjects population. Laboratory test values and vital signs were summarized by mean change from baseline, categorical shifts (low, normal, high) from baseline, and frequency/percent of extreme values for the safety population and eligible subjects population. This review presents results for the safety population only.

Among the laboratory tests and vital signs, alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatinine (CREAT), blood glucose (GLUC), hematocrit (HCT), hemoglobin (HGB), and blood pressure were considered of greater clinical relevance and are

presented in this review. The reviewer investigated extreme laboratory values and changes from baseline to Visit 9 and from baseline to the End of Trial Visit. The reviewer analyzed blood pressure changes of at least 20 mmHg from baseline to any visit during the treatment phase (Visit 4 to End of Trial Visit) and blood pressures exceeding cutoffs 130/80 and 160/110, per the Division of Urology, Obstetrics, and Gynecology Products's request.

A sensitivity analysis of the primary outcome was conducted with a Bayesian hierarchical model (see Appendix section 6.3 Bayesian Hierarchical Model) that borrowed information from legacy phase 3 trial DR-PGN-302, which had a 10% spontaneous abortion rate, a secondary endpoint.

No preplanned analyses were specified to assess missing data. The reviewer conducted a post-hoc sensitivity analysis of the primary outcome to assess the impact of missing β-hCG and TVUS from subjects whose spontaneous abortion status could not be determined by the end of the trial.

3.2.4 Subject Disposition, Demographic and Baseline Characteristics

3.2.4.1 Subject Disposition

Starting with 352 screened subjects, 274 subjects became eligible for the eligible subjects population, 254 were treated with PVR and became eligible for the safety population, and 243 had embryo transfer and became eligible for the mITT population. By the end of the trial, 239 subjects had complete information to determine their status for spontaneous abortion. The remaining four subjects with incomplete information were as follows: one subject withdrew consent, one subject was lost to follow-up, and two subjects exited the trial after β-hCG testing but before TVUS were completed.

3.2.4.2 Demographic and Baseline Characteristics

Baseline characteristics collected at the screening period included demographics, medical and gynecological history, physical examination results, height, weight, and body mass index (BMI), and concomitant medication. Demographics for the mITT population and safety population are presented in Table 1. For the mITT population, the mean age was 31 years, which ranged from 22 to 34 years, and 70% of subjects were at least 30 years old. Race was mainly comprised of 82% White, 9% Asian, and 7% Black/African American.

Select baseline characteristics are presented in Table 2. In the mITT population, the average duration of infertility was 34 months, 73% had BMI in the range of 18.5 and 29.9 kg/m^2, 57% drank alcohol 3 months prior to baseline, and 87% did not have a history of smoking.

The safety population was similar to the mITT population in demographics and select baseline characteristics.

Table 1 Demographics

	mITT Population (N=243)	Safety Population (N=254)
Age (years)		
Mean (standard deviation)	30.8 (2.7)	30.8 (2.7)
Min; Max	22; 34	22; 34
Age group, n (%)		
< 25	6 (2.5)	6 (2.4)
25-30	68 (28.0)	71 (28.0)
>=30	169 (69.5)	177 (69.7)
Race, n (%)		
American Indian/Alaska Native	3 (1.2)	3 (1.2)
Asian	21 (8.6)	22 (8.7)
Black/African American	17 (7.0)	17 (6.7)
Native Hawaiian or other Pacific Islander	2 (0.8)	2 (0.8)
White	200 (82.3)	210 (82.7)
Ethnicity, n (%)		
Hispanic/Latina	28 (11.5)	30 (11.8)
Not Hispanic/Latina	215 (88.5)	224 (88.2)

% = n/N
Source: Reviewer's analysis

Table 2 Select baseline characteristics

	mITT Population (N=243)	Safety Population (N=254)
Duration of infertility (months)		
Mean (standard deviation)	34.1 (23.8)	34.7 (24.2)
Min; Max	0; 144	0; 144
Body mass index (kg/m^2), n (%)		
< 18.5	7 (2.9)	7 (2.8)
18.5 - 24.9	106 (43.6)	111 (43.7)
25.0 - 29.9	72 (29.6)	74 (29.1)
30.0 - 34.9	37 (15.2)	40 (15.7)
35.0 - 38.0	21 (8.6)	22 (8.7)
Alcohol in the last 3 months, n (%)		
No	104 (42.8)	108 (42.5)
Yes	139 (57.2)	146 (57.5)
History of smoking, n (%)		
No	212 (87.2)	222 (87.4)
Yes	31 (12.8)	32 (12.6)

% = n/N
Source: Reviewer's analysis

3.2.5 Exposure

For the mITT population, the average number of PVRs used was 6 PVRs (standard deviation (SD) 4 PVRs), and the average duration of treatment was 42 days (SD 27 days) (Table 3). Forty-one percent (99/243) of subjects had 10 weeks of treatment, and among those who had less than 10 weeks, 63% (90/142) had 2-3 weeks of treatment, which was at the time of β-hCG testing to determine pregnancy status. During the trial, subjects discontinued treatment if they were determined not pregnant (negative β-hCG test) or found to have lost a pregnancy (biochemical abortion, ectopic pregnancy, heterotopic pregnancy, spontaneous abortion).

Exposure in the safety population was similar to the mITT population, except there were 11 more subjects who had < 2 weeks of treatment.

Table 3 Exposure to progesterone vaginal ring (PVR)

	mITT Population (N=243)	Safety Population (N=254)
Number of PVRs used		
Mean (standard deviation)	6.1 (3.7)	5.8 (3.8)
Min; Max	2; 12	1; 12
Duration of PVR treatment (days)		
Mean (standard deviation)	41.6 (26.8)	40.0 (27.2)
Min; Max	12; 78	4; 78
Duration of PVR treatment (weeks), n (%)		
< 1 week	0 (0)	10 (3.9)
1-2 weeks	14 (5.8)	15 (5.9)
2-3 weeks	90 (37.0)	90 (35.4)
3-4 weeks	15 (6.2)	15 (5.9)
4-5 weeks	8 (3.3)	8 (3.1)
5-6 weeks	4 (1.6)	4 (1.6)
6-7 weeks	3 (1.2)	3 (1.2)
7-8 weeks	1 (0.4)	1 (0.4)
8-9 weeks	1 (0.4)	1 (0.4)
9-10 weeks	6 (2.5)	6 (2.4)
10-11 weeks	99 (40.7)	99 (39.0)
11-12 weeks	2 (0.8)	2 (0.8)

% = n/N
Source: Reviewer's analysis

3.2.6 Results and Conclusions

3.2.6.1 Primary Outcome

The primary outcome was any spontaneous abortion occurring on or before 12 weeks in the mITT population. There were 18 events out of 243 subjects in the mITT population. The percentage of spontaneous abortion was 7.4% (95% CI 4.4, 11.5). A 15% spontaneous abortion rate was ruled out with the upper bound of the CI.

The number of spontaneous abortions ranged from 0 to 4 and the number of subjects ranged from 7 to 29 by study site (Appendix Table 1). The proportion of spontaneous abortion was somewhat consistent across study sites (Appendix Figure 1).

3.2.6.2 Secondary Outcomes

Table 4 presents the estimates for secondary outcomes, which were consistent with the sponsor's results. Estimates in the mITT population were slightly greater with less precision compared to the safety population because the mITT population had 11 subjects less. The percent of biochemical abortion at 6 or 10 weeks, 10.3% (CI 6.8, 14.8), was higher than other adverse secondary outcomes, but this result was expected clinically. There was one ectopic pregnancy and zero heterotopic pregnancy.

Table 4 Number of events and proportion estimates (95% exact confidence interval (CI)) for the modified intention-to-treat (mITT) population and safety population by outcome

Outcomes	mITT Population Number of events (N=243)	mITT Population Percentage (95% CI)	Safety Population Number of events (N=254)	Safety Population Percentage (95% CI)
Spontaneous abortion 12 weeks (primary)	**18**	**7.4 (4.4, 11.5)**	**-**	**-**
Spontaneous abortion 10 weeks	18	7.4 (4.4, 11.5)	18	7.1 (4.3, 11.0)
Spontaneous abortion 6 weeks	14	5.8 (3.2, 9.5)	14	5.5 (3.0, 9.1)
Biochemical abortion 10 weeks[1]	25	10.3 (6.8, 14.8)	25	9.8 (6.5, 14.2)
Biochemical abortion 6 weeks[1]	25	10.3 (6.8, 14.8)	25	9.8 (6.5, 14.2)
Ectopic pregnancy	1	0.4 (0.0, 2.3)	1	0.4 (0.0, 2.2)
Heterotopic pregnancy	0	0.0 (0.0, 1.5)	0	0.0 (0.0, 1.4)
Positive β-hCG	148	60.9 (54.5, 67.1)	148	58.3 (51.9, 64.4)
Clinical pregnancy 10 weeks	105	43.2 (36.9, 49.7)	105	41.3 (35.2, 47.7)
Clinical pregnancy 6 weeks	109	44.9 (38.5, 51.3)	109	42.9 (36.7, 49.2)

[1]Results included two subjects who completed β-hCG testing, but left before TVUS, and had β-hCG serum levels that qualified the pregnancies as biochemical abortions.
Source: Reviewer's analysis

3.2.6.3 Sensitivity Analyses

Bayesian Analysis

The reviewer's implementation of the sponsor's Bayesian hierarchical model (see Appendix section 6.3 Bayesian Hierarchical Model for model details) resulted in a posterior mean of 8.3% and 95% credible interval (5.3, 11.3) for spontaneous abortion, similar to the sponsor's estimates. The effective number of borrowed subjects was approximately 81, or 25% (81/(243+81)) of the effective sample size. The posterior estimates should be interpreted with care given that the proposed study sample size (N = 240) corresponded to an inflated type I error of 0.08, which was more than 3 times the error probability of ruling out 15% for a frequentist non-inferiority test (0.025). More appropriate for borrowing from many studies, the proposed Bayesian hierarchical model cannot reliably estimate between-study variability when borrowing from only one study. Additionally, posterior distribution inference, power, and type I error were sensitive to the specification of the between trial variation hyperprior distribution.

Missing Data

Classifying the 4 subjects who had incomplete information on their status for spontaneous abortion (see section 3.2.4.1 Subject Disposition) as having spontaneous abortion increased the number of events from 18 to 22, which corresponded to a rate of 9.1% (95% CI 5.8, 13.4%). The interpretation of the sensitivity analysis remained the same, i.e., the safety trial ruled out a spontaneous abortion rate of 15%.

3.2.6.4 Adverse Events

In the safety population, 49% (124/254) of subjects had 312 treatment emergent adverse events (TEAEs), with 0 events leading to death, 44 events leading to PVR discontinuation in 44 (17%) subjects, and 53 adverse drug reactions in 25 (10%) subjects. Inspection of TEAEs by preferred

term, frequency, and intensity (Appendix Table 2 through Appendix Table 5) showed that the results were consistent with the progesterone class and early pregnancy.

Table 5 Treatment emergent adverse event (TEAE) summary in the safety population

	Number of subjects with events (%) (N=254)	Number of events
TEAE	124 (48.8)	312
TEAE leading to death	0 (0.0)	0
Serious TEAE	4 (1.6)	4
TEAE leading to progesterone vaginal ring (PVR) discontinuation	44 (17.3)	44
TEAE of special interest[a]	5 (2.0)	10
TEAE adverse drug reaction[b]	25 (9.8)	53

[a] vaginal bleeding/spotting, pain/irritation, hemorrhage, infection, adhesions, and vaginal or cervical abrasions and lesions
[b] TEAE were investigator-judged to have reasonable possibility of causality by the PVR.
Note: A subject can have more than one event type.
Source: Reviewer's analysis

3.2.6.5 Laboratory Tests and Vital Signs

In addition to the sponsor's change from baseline analyses, the reviewer's analyses of laboratory tests (Appendix Table 6) showed:

- 3 subjects with ALT values ≥ 2 times the upper limit of the normal reference range (ULN) at Visit 9
- 1 subject with AST ≥ 2 times the ULN at Visit 9
- 2 subjects with ALT ≥ 2 times the ULN at the End of Trial Visit
- 1 subject with AST ≥ 2 times the ULN at the End of Trial Visit

In addition to the sponsor's change from baseline analyses, the reviewer's analyses of vital signs (Appendix Table 7) showed:

- 1 subject with systolic blood pressure ≥ 160 mmHg and an increase ≥ 20 mmHg from baseline to any visit during the treatment phase (Visit 4 to End of Trial Visit)
- 1 subject with diastolic blood pressure ≥ 110 mmHg and an increase ≥ 20 mmHg from baseline to any visit during the treatment phase (Visit 4 to End of Trial Visit)

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The sponsor did not provide an analysis by subgroup populations. Given that this is a single arm trial that did not have a control, on-treatment subgroup differences cannot be separated from potential baseline subgroup differences; precluding meaningful conclusions.

5 SUMMARY AND CONCLUSIONS

Trial 000293 was a multi-center, open label, single arm, safety trial of women 18-34 years old treated once weekly with Milprosa progesterone vaginal ring (PVR) for up to 10 weeks. The primary outcome was any spontaneous abortion occurring on or before 12 weeks following oocyte retrieval in the modified intention-to-treat (mITT) population, defined as all subjects treated with PVR and completed embryo transfer. The trial was designed to rule out a 15% spontaneous abortion rate.

Among 254 subjects treated with PVR (safety population), 243 subjects were eligible for the mITT population. By the end of the trial, 239 subjects had complete information to determine their status for spontaneous abortion. The remaining four subjects with incomplete information were as follows: one subject withdrew consent, one subject was lost to follow-up, and two subjects exited the trial after β-hCG testing but before TVUS were completed.

The mITT population's mean age was 31 years, which ranged from 22 to 34 years, and 70% of subjects were least 30 years old. Race was mainly comprised of 82% White, 9% Asian, and 7% Black/African American. The average duration of infertility was 34 months with standard deviation 24 months.

In the mITT population of 243 subjects, there were 18 spontaneous abortions. The percentage of spontaneous abortion was 7.4% (95% CI 4.4, 11.5). A 15% spontaneous abortion rate was ruled out with the upper bound of the CI. Post-hoc sensitivity analysis to missing data of four patients whose status for spontaneous abortion could not be determined resulted in similar conclusions to the primary analysis.

Estimates for the adverse secondary outcomes in the mITT population were as follows: biochemical abortion at 10 weeks (25 events, 10.3%; 95% CI 6.8, 14.8), ectopic pregnancy (1 event, 0.4%; 95% CI 0.0, 2.3), and heterotopic pregnancy (0 event, 0.0%; 95% CI 0.0, 1.5).

In the safety population, 49% (124/254) of subjects had 312 treatment emergent adverse events (TEAEs), with 0 events leading to death, 44 events leading to PVR discontinuation in 44 (17%) subjects, and 53 adverse drug reactions, TEAEs that an investigator judged to have reasonable possibility of causality by the PVR, in 25 (10%) subjects. Inspection of TEAEs by preferred term, frequency, and intensity showed that the results were consistent with the progesterone class and early pregnancy.

Laboratory tests and vital signs of clinical relevance included alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatinine, blood glucose, hematocrit, hemoglobin, and blood

pressure. Although there were subjects with extreme changes from baseline for ALT, AST, and blood pressure, the frequencies were low (1-3 subjects).

Meaningful conclusions for demographic subgroups cannot be drawn because on-treatment subgroup differences cannot be separated from potential baseline subgroup differences in a single arm trial.

In conclusion, this single arm, safety trial ruled out a 15% spontaneous abortion rate. There were no major safety concerns for other pregnancy outcomes, adverse events, laboratory tests, and blood pressure.

6 APPENDIX

6.1 Tables

Appendix Table 1 Number of spontaneous abortions and percent by site for the modified intention-to-treat (mITT) population

Site ID	Number of spontaneous abortions (n)	N	Percent (n/N)
USA02	2	20	10.0
USA03	1	12	8.3
USA04	1	25	4.0
USA05	1	7	14.3
USA06	2	19	10.5
USA07	0	16	0.0
USA08	1	15	6.7
USA09	1	22	4.5
USA10	4	29	13.8
USA12	1	19	5.3
USA13	2	13	15.4
USA14	1	14	7.1
USA15	1	23	4.3
USA16	0	9	0.0

Source: Reviewer's analysis

Appendix Table 2 Treatment emergent adverse event (TEAE) with proportion ≥ 5% by body system or organ class and dictionary-derived term

Body system organ class/ dictionary-derived term	Number of subjects with events (%) (N=254)	Number of events
Frequent TEAE	119 (46.9)	262
Pregnancy, puerperium and perinatal conditions	61 (24.0)	75
Biochemical abortion	25 (9.8)	25
Spontaneous abortion	18 (7.1)	18
Gastrointestinal disorders	46 (18.1)	73
Nausea	22 (8.7)	22
Reproductive system and breast disorders	34 (13.4)	56
Infections and infestations	25 (9.8)	30
Nervous system disorders	20 (7.9)	28
Headache	13 (5.1)	17

Note: A subject can have more than one event type.
Source: Reviewer's analysis

Appendix Table 3 Treatment emergent adverse event (TEAE) leading to progesterone vaginal ring (PVR) discontinuation by body system or organ class and dictionary-derived term

Body system or organ class/ dictionary-derived term	Number of subjects with events (%) (N=254)	Number of events
TEAE leading to PVR discontinuation	44 (17.3)	44
Pregnancy, puerperium and perinatal conditions	43 (16.9)	43
Biochemical abortion	25 (9.8)	25
Ectopic pregnancy	1 (0.4)	1
Spontaneous abortion	17 (6.7)*	17
Psychiatric disorders	1 (0.4)	1
Suicidal ideation	1 (0.4)	1

* There were 18 spontaneous abortions whereby 1 subject had multiple gestations continuing in the trial because she had both a spontaneous abortion and an ongoing pregnancy.
Note: A subject can have more than one event type.
Source: Reviewer's analysis

Appendix Table 4 Treatment emergent adverse event (TEAE) by intensity

	Number of subjects with events (%) (N=254)	Number of events
Any TEAE	124 (48.8)	312
Mild	100 (39.4)	228
Moderate	46 (18.1)	75
Severe	9 (3.5)	9

Note: A subject can have more than one event type.
Source: Reviewer's analysis

Appendix Table 5 Treatment emergent adverse event (TEAE) of special interest

Body system or organ class/ dictionary-derived term	Number of subjects with events (%) (N=254)	Number of events
TEAE of special interest	5 (2.0)	10
Reproductive system and breast disorders	5 (2.0)	10
Vaginal hemorrhage	4 (1.6)	6
Vulvovaginal pain	2 (0.8)	3
Cervix lesion	1 (0.4)	1

Note: A subject can have more than one event type.
Source: Reviewer's analysis

Appendix Table 6 Laboratory shift from baseline (Visit 4) to post-baseline visit (Visit 9 or End of Trial Visit) for safety population (N=254)

Visit	Lab	Shift	N1	n	Percent (n/N1)
9	ALT (U/L)	Normal --> Normal	233	218	93.6
9	ALT (U/L)	Normal --> High	233	15	6.4
9	ALT (U/L)	Low --> Normal	3	2	66.7
9	ALT (U/L)	Low --> Low	3	1	33.3
9	ALT (U/L)	High --> Normal	12	10	83.3
9	ALT (U/L)	High --> High	12	2	16.7
9	ALT (U/L)	≥ 2x ULN	248	3	1.2
9	ALT (U/L)	≥ 3x ULN	248	1	0.4
9	ALT (U/L)	≥ 5x ULN	248	0	0
9	AST (U/L)	Normal --> Normal	242	234	96.7
9	AST (U/L)	Normal --> Low	242	4	1.7
9	AST (U/L)	Normal --> High	242	4	1.7
9	AST (U/L)	High --> Normal	6	6	100
9	AST (U/L)	≥ 2x ULN	248	1	0.4
9	AST (U/L)	≥ 3x ULN	248	0	0

9	AST (U/L)	≥ 5x ULN	248	0	0
9	CREAT (mg/dL)	Normal --> Normal	247	226	91.5
9	CREAT (mg/dL)	Normal --> Low	247	21	8.5
9	CREAT (mg/dL)	Low --> Normal	1	1	100
9	CREAT (mg/dL)	≥ 2x ULN	248	0	0
9	CREAT (mg/dL)	≥ 3x ULN	248	0	0
9	CREAT (mg/dL)	≥ 5x ULN	248	0	0
9	GLUC (mg/dL)	Normal --> Normal	235	229	97.4
9	GLUC (mg/dL)	Normal --> Low	235	5	2.1
9	GLUC (mg/dL)	Normal --> High	235	1	0.4
9	GLUC (mg/dL)	Low --> Normal	7	6	85.7
9	GLUC (mg/dL)	Low --> Low	7	1	14.3
9	GLUC (mg/dL)	High --> Normal	4	3	75
9	GLUC (mg/dL)	High --> High	4	1	25
9	GLUC (mg/dL)	≥ 2x ULN	246	0	0
9	GLUC (mg/dL)	≥ 3x ULN	246	0	0
9	GLUC (mg/dL)	≥ 5x ULN	246	0	0
9	HCT (%)	Normal --> Normal	220	194	88.2
9	HCT (%)	Normal --> Low	220	21	9.5
9	HCT (%)	Normal --> High	220	5	2.3
9	HCT (%)	Low --> Normal	22	17	77.3
9	HCT (%)	Low --> Low	22	5	22.7
9	HCT (%)	High --> Normal	6	6	100
9	HCT (%)	≥ 2x ULN	248	0	0
9	HCT (%)	≥ 3x ULN	248	0	0
9	HCT (%)	≥ 5x ULN	248	0	0
9	HGB (g/dL)	Normal --> Normal	223	196	87.9
9	HGB (g/dL)	Normal --> Low	223	27	12.1
9	HGB (g/dL)	Low --> Normal	23	12	52.2
9	HGB (g/dL)	Low --> Low	23	10	43.5
9	HGB (g/dL)	Low --> High	23	1	4.3
9	HGB (g/dL)	High --> Normal	2	2	100
9	HGB (g/dL)	≥ 2x ULN	248	0	0
9	HGB (g/dL)	≥ 3x ULN	248	0	0
9	HGB (g/dL)	≥ 5x ULN	248	0	0
End of Trial	ALT (U/L)	Normal --> Normal	229	216	94.3
End of Trial	ALT (U/L)	Normal --> High	229	13	5.7
End of Trial	ALT (U/L)	Low --> Normal	3	2	66.7
End of Trial	ALT (U/L)	Low --> Low	3	1	33.3
End of Trial	ALT (U/L)	High --> Normal	12	10	83.3
End of Trial	ALT (U/L)	High --> High	12	2	16.7
End of Trial	ALT (U/L)	≥ 2x ULN	244	2	0.8
End of Trial	ALT (U/L)	≥ 3x ULN	244	0	0
End of Trial	ALT (U/L)	≥ 5x ULN	244	0	0
End of Trial	AST (U/L)	Normal --> Normal	238	230	96.6
End of Trial	AST (U/L)	Normal --> Low	238	3	1.3
End of Trial	AST (U/L)	Normal --> High	238	5	2.1
End of Trial	AST (U/L)	High --> Normal	6	6	100
End of Trial	AST (U/L)	≥ 2x ULN	244	1	0.4
End of Trial	AST (U/L)	≥ 3x ULN	244	1	0.4
End of Trial	AST (U/L)	≥ 5x ULN	244	0	0
End of Trial	CREAT (mg/dL)	Normal --> Normal	243	213	87.7
End of Trial	CREAT (mg/dL)	Normal --> Low	243	29	11.9
End of Trial	CREAT (mg/dL)	Normal --> High	243	1	0.4
End of Trial	CREAT (mg/dL)	Low --> Normal	1	1	100

End of Trial	CREAT (mg/dL)	≥ 2x ULN	244	0	0
End of Trial	CREAT (mg/dL)	≥ 3x ULN	244	0	0
End of Trial	CREAT (mg/dL)	≥ 5x ULN	244	0	0
End of Trial	GLUC (mg/dL)	Normal --> Normal	232	221	95.3
End of Trial	GLUC (mg/dL)	Normal --> Low	232	10	4.3
End of Trial	GLUC (mg/dL)	Normal --> High	232	1	0.4
End of Trial	GLUC (mg/dL)	Low --> Normal	8	6	75
End of Trial	GLUC (mg/dL)	Low --> Low	8	2	25
End of Trial	GLUC (mg/dL)	High --> Normal	4	4	100
End of Trial	GLUC (mg/dL)	≥ 2x ULN	244	0	0
End of Trial	GLUC (mg/dL)	≥ 3x ULN	244	0	0
End of Trial	GLUC (mg/dL)	≥ 5x ULN	244	0	0
End of Trial	HCT (%)	Normal --> Normal	214	178	83.2
End of Trial	HCT (%)	Normal --> Low	214	34	15.9
End of Trial	HCT (%)	Normal --> High	214	2	0.9
End of Trial	HCT (%)	Low --> Normal	22	14	63.6
End of Trial	HCT (%)	Low --> Low	22	8	36.4
End of Trial	HCT (%)	High --> Normal	6	6	100
End of Trial	HCT (%)	≥ 2x ULN	242	0	0
End of Trial	HCT (%)	≥ 3x ULN	242	0	0
End of Trial	HCT (%)	≥ 5x ULN	242	0	0
End of Trial	HGB (g/dL)	Normal --> Normal	217	180	82.9
End of Trial	HGB (g/dL)	Normal --> Low	217	37	17.1
End of Trial	HGB (g/dL)	Low --> Normal	23	13	56.5
End of Trial	HGB (g/dL)	Low --> Low	23	10	43.5
End of Trial	HGB (g/dL)	High --> Normal	2	2	100
End of Trial	HGB (g/dL)	≥ 2x ULN	242	0	0
End of Trial	HGB (g/dL)	≥ 3x ULN	242	0	0
End of Trial	HGB (g/dL)	≥ 5x ULN	242	0	0

Except for 2x, 3x, 5x ULN, shifts with zero subjects (n=0) are not presented.
Normal indicated values within and including the normal range.
Low indicated values less than the lower limit of the normal range.
High indicated values greater than the upper limit of the normal range (ULN).
n=number of subjects with indicated shift.
N1=number of subjects with a baseline value and a post-baseline value. For shifts 2x, 3x, 5x ULN, N1 was the number of subjects with a post-baseline value that was at least 2x, 3x, 5x ULN.
Source: Reviewer's analysis

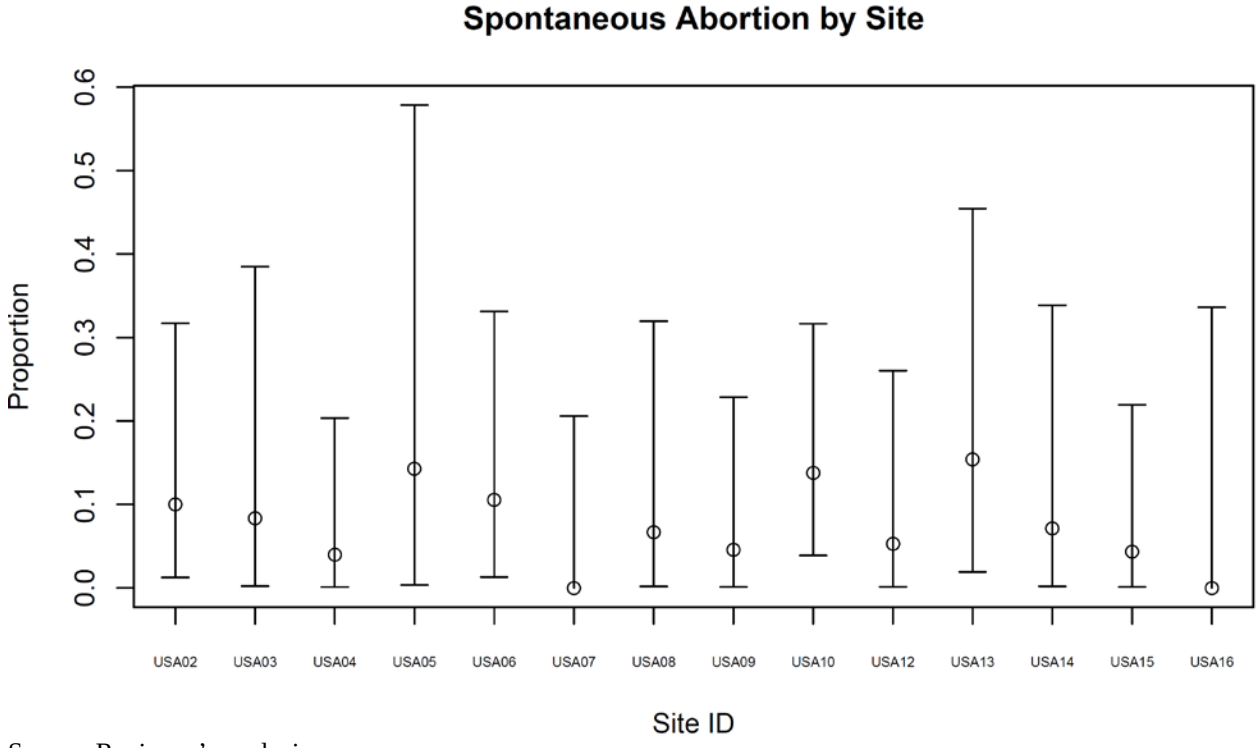
Appendix Table 7 Blood pressure analysis for measurements taken at any time during the treatment phase, Visit 4 to End of Trial

	Safety Population (N=254)	
	n	Percent (n/N)
Systolic blood pressure (mmHg)		
Increase of ≥ 20 from baseline	47	18.5
Increase of ≥ 20 from baseline and systolic blood pressure ≥ 130	28	11
Increase of ≥ 20 from baseline and systolic blood pressure ≥ 160	1	0.4
Decrease of ≥ 20 from baseline	21	8.3
Diastolic blood pressure (mmHg)		
Increase of ≥ 20 from baseline	22	8.7
Increase of ≥ 20 from baseline and diastolic blood pressure ≥ 80	22	8.7
Increase of ≥ 20 from baseline and diastolic blood pressure ≥ 110	1	0.4
Decrease of ≥ 20 from baseline	12	4.7

Source: Reviewer's analysis

6.2 Figures

Appendix Figure 1 Spontaneous abortion proportion estimates and the Clopper-Pearson two-sided exact (binomial distribution) 95% confidence intervals by site



Source: Reviewer's analysis

6.3 Bayesian Hierarchical Model

The proposed Bayesian hierarchical model is as follows:

(b) (4)

(b) (4)

A proposed sample size of 240 subjects corresponded to 0.93 power and a type I error of 0.08. Power is the probability of concluding that the 95% credible interval is below 15%, assuming a true spontaneous abortion rate of 9%. Type I error is the probability of concluding that the 95% credible interval is below 15%, assuming a true spontaneous abortion rate of 15%. The type I error of 0.08 was more than 3 times the error probability of ruling out 15% for a frequentist non-inferiority test (0.025).

Effective sample size (ESS) is a metric for quantifying the sample size of the posterior distribution that the Bayesian analysis is effectively operating when it combines prior information with trial data. The sponsor provided a formula based on generalizing the ESS of the beta posterior distribution. The generalization is given because there are no known conjugate priors for the proposed hierarchical model. Note that if a conjugate Beta(α, β) prior is used for rate p, the prior effective sample size is $\alpha + \beta$.

(b) (4) the sponsor gave the following derivations:

(b) (4)

The left-hand side is the ESS for a beta distribution and the right-hand side can be calculated from the posterior distribution.

From the ESS, the sponsor provided a metric called the effective number of borrowed subjects (ENBS) for quantifying the prior information, i.e., ESS = trial sample size + ENBS. Greater ENBS corresponds to greater prior influence.

(b) (4)

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/

THANH V TRAN
04/08/2020 11:42:23 AM

CLARA KIM
04/08/2020 11:46:25 AM

MATTHEW J SOUKUP
04/08/2020 12:09:02 PM

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

STATISTICAL ANALYSIS PLAN

NDA/SN #:	201110/47
Drug Name:	Milprosa (progesterone) vaginal ring
Indication(s):	Progesterone supplementation for up to 10 weeks post-embryo transfer as part of assisted reproductive technology treatment in infertile women
Safety Issue(s):	Spontaneous abortion and other adverse pregnancy outcomes
Applicant:	Ferring Pharmaceuticals, Inc.
Date(s):	Document Stamp Date: March 12, 2019 Date Completed: August 29, 2019
Biometrics Division:	Division of Biometrics VII
Statistical Reviewer:	Thanh Van Tran, PhD
Concurring Reviewers:	Clara Y. Kim, PhD, Team Leader
Consulting Division:	Division of Bone, Reproductive, and Urologic Products
Consult Request Team:	Shelly Slaughter, MD, Team Leader
Project Manager:	Nikia D. Morris

Keywords: progesterone vaginal ring, effective sample size, Bayesian statistics, spontaneous abortion, hierarchical model, dynamic borrowing, type I error, power

1 INTRODUCTION

Milprosa (progesterone) vaginal ring (PVR), indicated for progesterone supplementation as part of assisted reproductive technology (ART) treatment in infertile women, was submitted in April 2010 under NDA 201110 by Ferring Pharmaceuticals, Inc. The NDA received a complete response letter (CRL); one of the main issues being the presence of (b) (4) particulates contamination in the PVR.

In 2016, the sponsor resubmitted the application addressing the particulate contamination issue by altering the manufacturing process with the use of a new (b) (4) process. A bioequivalence study showed that the old legacy PVR and the new PVR manufactured with the new (b) (4) process were similar. Upon review of the resubmission, FDA issued another CRL that included a request for the sponsor to conduct a biocompatibility study and a safety study to establish a clinical safety bridge between the legacy PVR (studied in the phase 3 clinical trials) and the new PVR. FDA recommended the safety study to evaluate the clinical safety and tolerability of Milprosa over the entire treatment duration (up to 10 weeks post-embryo transfer) in women undergoing ART procedures.

The sponsor proposed an open label, single arm trial using a Bayesian hierarchical model that borrows information from pivotal phase 3 trial DR-PGN-302, which investigated clinical pregnancy rate at 8 and 12 weeks of pregnancy. DR-PGN-302 showed subjects treated with once weekly legacy PVR had a pregnancy rate that is statistically non-inferior to the active comparator, daily 8% progesterone vaginal gel, in women 18-34 years old. Spontaneous abortion, a secondary endpoint in the trial, was estimated at 55/549 (10%).

In an advice letter in February 2018, FDA recommended the sponsor to power the clinical safety study so that the upper bound of the 95% confidence interval (CI) for the primary outcome spontaneous abortion does not exceed 15%. In March 2018, the sponsor submitted study protocol (version 4.0) for the phase 3b trial 000293 titled, "A Prospective, Multi-Center, Non-Comparative Trial of the Clinical Safety of the Progesterone Vaginal Ring in Women Undergoing Assisted Reproductive Technology (ART) Procedures." The protocol review by DB7, dated June 19, 2018, focused on the Bayesian method used for the primary analysis and sample size calculation. Trial 000293 started on July 26, 2018, is ongoing, and is expected to finish in August 2019.

In October 2018, the sponsor submitted a corresponding statistical analysis plan (SAP) (version 2.0) as part of the background package for a Type C Guidance Meeting Written Response Only. The SAP repeats protocol version 4.0 description of the primary analysis and contains details about additional analyses. DB7's SAP review comments were incorporated into the Final Written Response dated December 6, 2018. Our comments asked for more information and justification of the Bayesian design and the comparability of trial 000293 and trial DR-PGN-302. In March 2019, the sponsor responded to DB7's SAP comments and provided a revised SAP, version 3.0. In this review, we provide a summary of the sponsor's proposed Bayesian hierarchical model, discuss statistical issues, and conclude with a FDA's recommendation for the sponsor to conduct a frequentist analysis.

The location of the SAP is \\cdsesub1\evsprod\nda201110\0045.

2 MODEL SUMMARY

The proposed Bayesian hierarchical model is as follows:

For trial 000293, the sponsor proposed a sample size of 215 subjects, associated with 81% power, assuming a true spontaneous rate of 10%. Power is the probability of concluding that the 95% CI is below 15%, assuming a true spontaneous rate of 10%. The type I error was estimated at 9%, i.e., the probability of 95% CI below 15%, assuming the true spontaneous rate is 15%.

As a point of reference, a sample size of 363 achieves 80% power to rule out 15% when the true rate is 10% at a one-sided significance level of 0.025 using a frequentist analysis.

3 STATISTICAL CONCERNS

3.1 Study Exchangeability

In a Bayesian hierarchical model, the trials are assumed exchangeable, i.e., trial outcomes do not depend on any reordering of the trials. The following differences between the trials challenge the exchangeability assumption:

- The 10% rate of spontaneous abortion is calculated from a randomized, active-controlled trial. However, trial 000293 is a single arm, open label trial.

- The study population for trial DR-PGN-302 consists of women 18-42 years old, but trial 000293 consists of women 18-34 years old.

- Inclusion and exclusion criteria are slightly different.

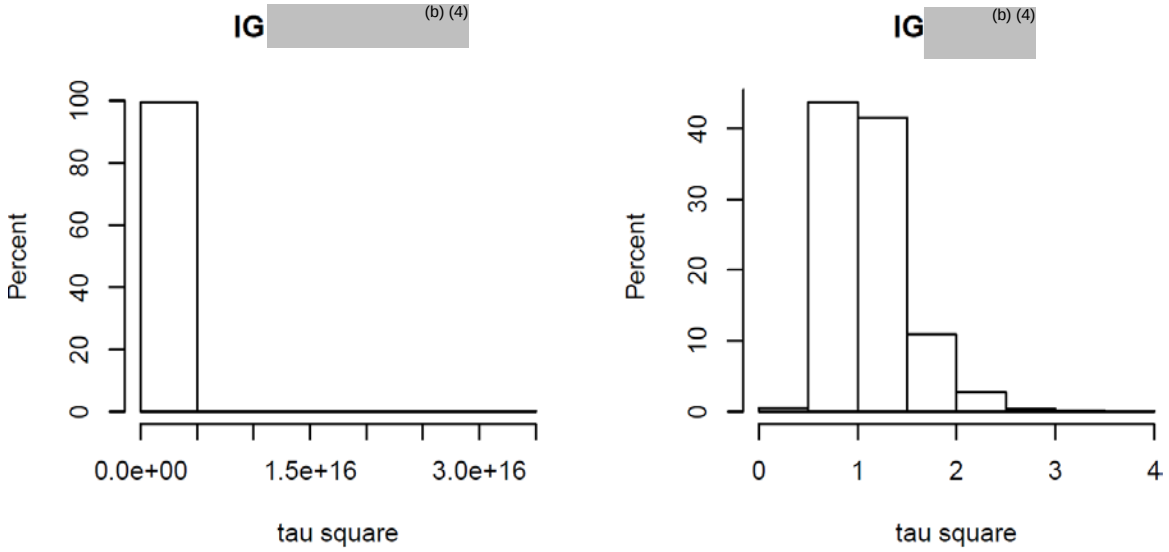
3.2 Model Specification

The proposed Bayesian hierarchical model more reliably estimates between-study variability when borrowing information from many studies. However, the proposed model borrows from only one study, resulting in inference, power, and type I error that are sensitive to the τ^2 prior specification. In this section, we present power, type I error, and borrowing operating characteristics of the proposed model.

3.2.1 Power and Type I Error

The proposed inverse gamma distribution IG((b) (4)) has substantial mass on very small and very large τ^2 values. The percentiles associated with IG((b) (4)) are: 25th percentile = 0.08, 50th percentile = 2.1, and 75th percentile = 529 (Figure 1). A specification of IG that has a smaller range is IG((b) (4)) with the 25th percentile = 0.8, 50th percentile = 1.0, and 75th percentile = 1.63.

Figure 1. Inverse gamma (IG) density plots.



Source: Reviewer's analysis

We considered other prior specifications of μ and τ^2 and assessed their impact on power and type I error. With the sponsor's prior specifications N((b) (4)) and IG((b) (4)), power increases from 0.81 to 0.91with increasing sample size when the true rate is 10% (Table 1). The type I error marginally decreases from 0.09 to 0.07 with increasing sample size when the true rate is 15%.

When the specification for τ^2 is changed to IG((b) (4)), the sponsor's proposed sample size of 215 is associated with a power of 0.59 assuming the true rate is 10% and a type I error of 0.03 assuming the true rate is 15%. Increasing the sample size to 320 increases the power to 0.80 while keeping the type 1 error at 0.03. Power and type I error appear negligibly affected by a more diffuse specification of μ with $\tau^2 \sim$ IG((b) (4)).

The sponsor's sample size proposal of 215 subjects corresponds to a 9% type I error that is more than three times the frequentist analysis type I error of 0.025 of a non-inferiority test. The reason for the sponsor's high type 1 error is the current trial is borrowing information from a previous trial with a 10% rate that is lower and far away from the risk margin of 15%. Risk margins greater than 11% correspond to type I error inflation that exceed 0.025(Appendix Figure 4). As expected, power increases with greater risk margins above 10% (Appendix Figure 5). At risk margins greater than 10%, borrowing from the previous trial decreases the estimated spontaneous abortion rate, making it easier to declare that there is no safety signal, a benefit for power, but a cost to type I error.

Table 1. The probability of the 95% credible interval being less than 15% under various true rates of spontaneous abortion, sample sizes, and prior specifications.

N(mean, var)	IG(shape, rate)	Sample Size	Spontaneous Abortion Rates				
			9%	10%	11%	12%	15%
Frequentist		363		0.80			0.025
(b) (4)		215	0.92	0.81	0.65	0.47	0.09
		300	0.97	0.89	0.73	0.53	0.08
		350	0.98	0.91	0.77	0.56	0.07
		215	0.78	0.59	0.41	0.25	0.03
		300	0.90	0.75	0.55	0.33	0.03
		320	0.93	0.80	0.59	0.37	0.03
		350	0.95	0.83	0.62	0.39	0.03
		215	0.78	0.59	0.41	0.25	0.03
		300	0.90	0.75	0.55	0.33	0.03
		320	0.93	0.80	0.59	0.37	0.03
		350	0.95	0.83	0.62	0.39	0.03

N(mean,var) normal distribution with parameters mean and variance
IG(shape, rate) inverse gamma distribution with parameters shape and rate
Results are based on 5000 simulated trials per parameter specification using FACTS software
Source: Reviewer's analysis

3.2.2 Borrowing Characteristics

Effective sample size (ESS) is a metric for quantifying the sample size of the posterior distribution that the Bayesian analysis is effectively operating when it combines prior information with study data. The sponsor provided a formula based on generalizing the ESS of

the beta posterior distribution. The generalization is given because there are no known conjugate priors for the proposed hierarchical model. Note that if a conjugate $\text{Beta}(\alpha, \beta)$ prior is used for rate *p*, the prior effective sample size is $\alpha + \beta$.

[redacted], the sponsor gave the following derivations:

[redacted]

The left-hand side is the ESS for a beta distribution and the right-hand side can be calculated from the posterior distribution.

From the ESS, the sponsor provided a metric called the effective number of borrowed subjects (ENBS) for quantifying the prior information, i.e., ESS = study sample size + ENBS. Greater ENBS corresponds to greater prior influence. To verify the sponsor's calculations, we computed the ENBS using the above equation. Selected ENBS calculations, frequentist estimates, and posterior distribution estimates are shown in Table 2 for study sample sizes 215 and 320. Figure 2 and Figure 3 show ENBS as a function of observed events, i.e., counts of spontaneous abortion, for study sample sizes 215 and 320.

As point of reference, 23 is the maximum number of events to rule out a 15% spontaneous abortion rate using a frequentist non-inferiority test. We have following observations regarding the borrowing characteristics of the sponsor's proposed hierarchical model:

- In the proposed hierarchical model, 25 is the maximum number of events to rule out a 15% spontaneous abortion rate, allowing 2 more events compared to the frequentist analysis (Table 2).

- Although there are differences in the ENBS calculations by the sponsor and by the reviewer for 22, 25, and 32 events, the borrowing percentages are similar (Table 2; 37 vs. 36, 32 vs. 33, 7 vs. 5).

[redacted]

- Under IG([redacted]), more borrowing (greater ENBS) occurred when the percentage of observed events is similar to the prior data (10%), resulting in posterior estimates closer to 10% (Table 2, Figure 2)

- Bayesian posterior estimates are more different from the frequentist estimates when the percentage of observed events is farther away from 10% because the Bayesian method borrows information from trial DR-PGN-320.

- Smaller borrowing percentages are associated with IG([redacted]) compared to IG([redacted]) when the percentages of observed events are close to 10% (Table 2, Figure 2, Figure 3).

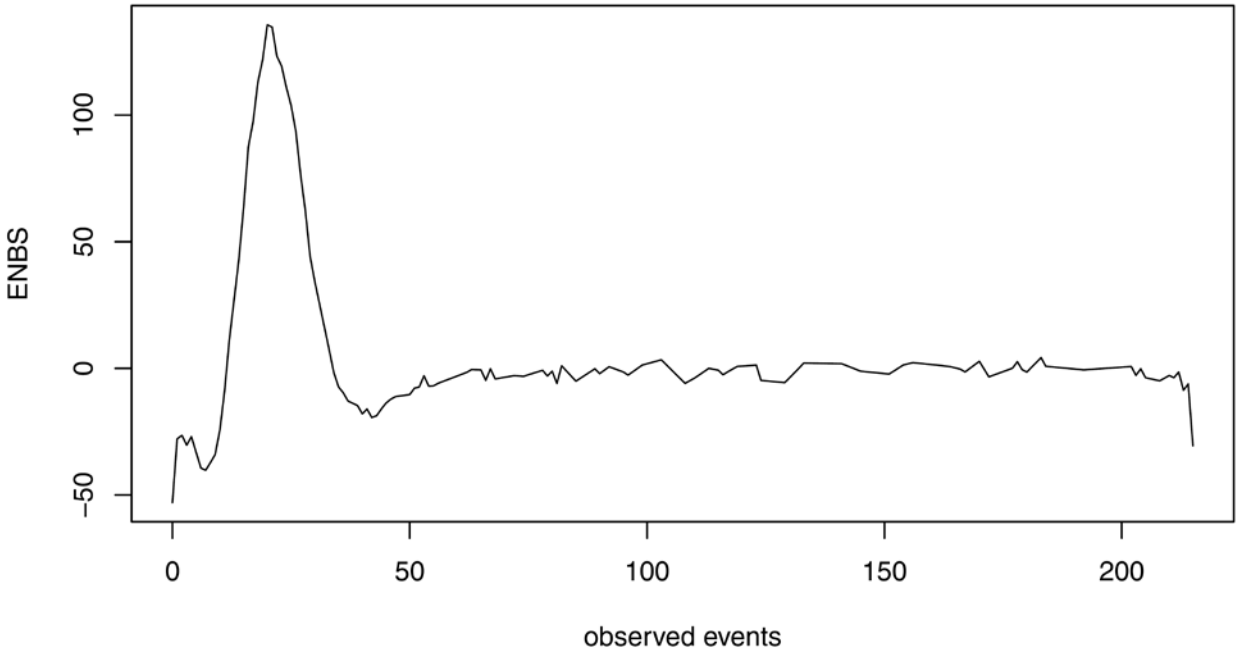
In summary, posterior distribution inference, power, type I error, and borrowing percentages are sensitive to the specification of the τ^2 hyperprior distribution.

Table 2. Frequentist and Bayesian spontaneous abortion rate estimates for different observed data and for sample sizes 215 and 320 under different inverse gamma (IG) specifications. Also shown are effective number of borrowed subjects (ENBS) from phase 3 trial DR-PGN-302. $\mu \sim$ [redacted] **.**

IG(shape, rate)	Observed Events/Sample Size	Frequentist Estimated Rate (%) (95% confidence interval)[a]	Bayesian Estimated Rate (%) (95% credible interval)[b]	Source[c]	ENBS	ENBS/ESS (%)[d]
IG([redacted])	10/215	4.65 (1.84,7.47)	5.71 (2.79,9.24)	Reviewer	-24	-
	15/215	6.98 (3.57,10.38)	7.98 (4.82,11.18)	Reviewer	64	23
	22/215	10.23 (6.18,14.28)	10.19 (7.23,13.69)	Sponsor	127	37
			10.16 (7.17,13.68)	Reviewer	123	36
	23/215	10.70 (6.57,14.83)	10.46 (7.48,14.05)	Reviewer	119	36
	25/215	11.63 (7.34,15.91)	11.06 (7.95,14.98)	Sponsor	100	32
			11.06 (8.01,14.93)	Reviewer	104	33
	32/215	14.88 (10.13,19.64)	12.62 (9.83,18.57)	Sponsor	15	7
			13.63 (9.78,18.56)	Reviewer	11	5
	36/215	16.74 (11.75,21.74)	15.42 (10.98,20.64)	Reviewer	-10	-
	40/215	18.60 (13.40,23.81)	17.48 (12.57,23.01)	Reviewer	-18	-
				Reviewer		
IG([redacted])	10/320	3.12 (1.22,5.03)	3.36 (1.73, 5.55)	Reviewer	11	5
	15/320	4.69 (2.37,7.00)	4.85 (2.85,7.40)	Reviewer	16	7
	22/320	6.88 (4.10,9.65)	6.95 (4.44,10.00)	Reviewer	4	2
	25/320	7.81 (4.87,10.75)	7.88 (5.19,11.06)	Reviewer	0	0
	32/320	10.00 (6.71,13.29)	10.00 (7.06, 13.43)	Reviewer	11	5
	36/320	11.25 (7.79,14.71)	11.24 (8.03,14.87)	Reviewer	13	6
	40/320	12.50 (8.88,16.12)	12.45 (9.07,16.23)	Reviewer	7	3

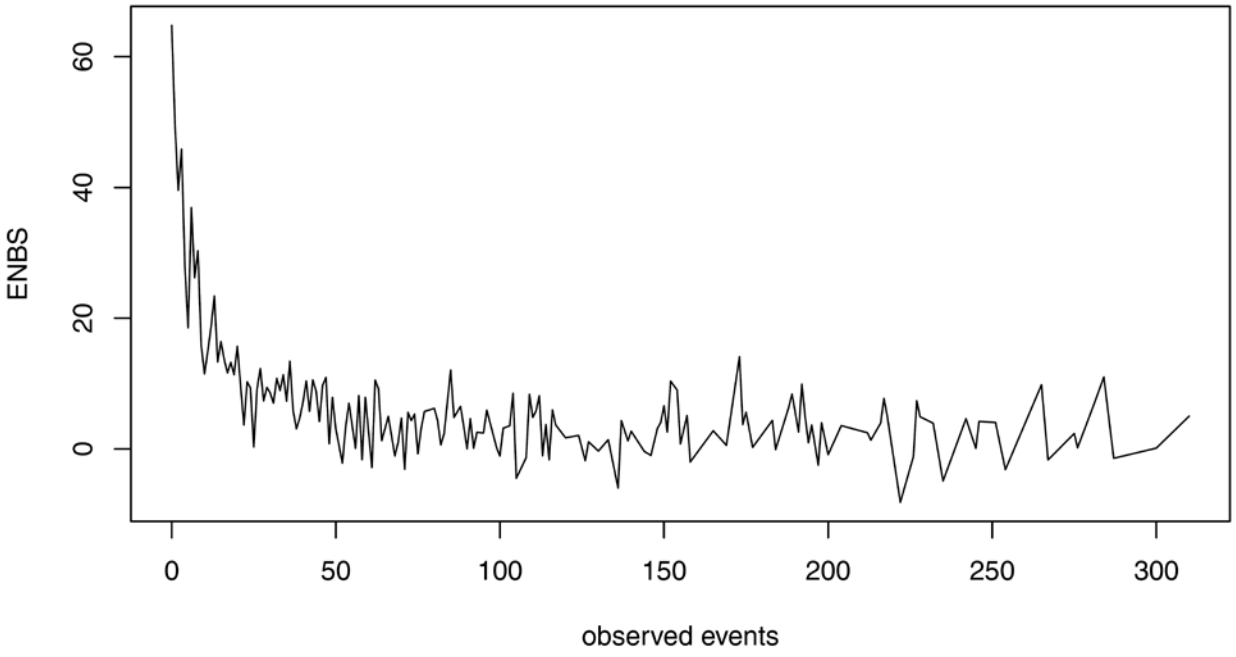
[a] Wald confidence interval
[b] 95% credible interval: (2.5th percentile, 97.5th percentile) of the posterior distribution
[c] Reviewer: calculations using the sponsor's model specifications; Sponsor: sponsor's reported estimates
[d] Effective sample size (ESS) = sample size + ENBS
- Cannot be calculated with negative ENBS

Figure 2 Effective number of borrowed subjects (ENBS) as function of number of observed events, sample size = 215, $\mu \sim N($ (b) (4) $), \tau^2 \sim IG($ (b) (4) $)$ **.**



Source: Reviewer's analysis.

Figure 3 Effective number of borrowed subjects (ENBS) as a function of number of observed events, sample size= 320, $\mu \sim N($ (b) (4) $), \tau^2 \sim IG($ (b) (4) $)$ **.**



Source: Reviewer's analysis.

4 SUMMARY

In conclusion, the type I error of 9% is too high; more than three times the frequentist analysis type I error of 0.025 of a non-inferiority test. In a Bayesian analysis, the type I error is expected to be greater than the corresponding frequentist type I error, because unlike the frequentist analysis, which does not borrow information from the previous trial, the Bayesian analysis borrows information to varying degrees. It's always possible to obtain a rate in the current trial that is different from 10%, and therefore, incur a type I error inflation. In the sponsor's proposal, the current trial borrows information from a previous trial that has a rate of 10% that is lower and far away from the risk margin of 15%, making it easier to declare no safety signal, i.e., a higher type I error.

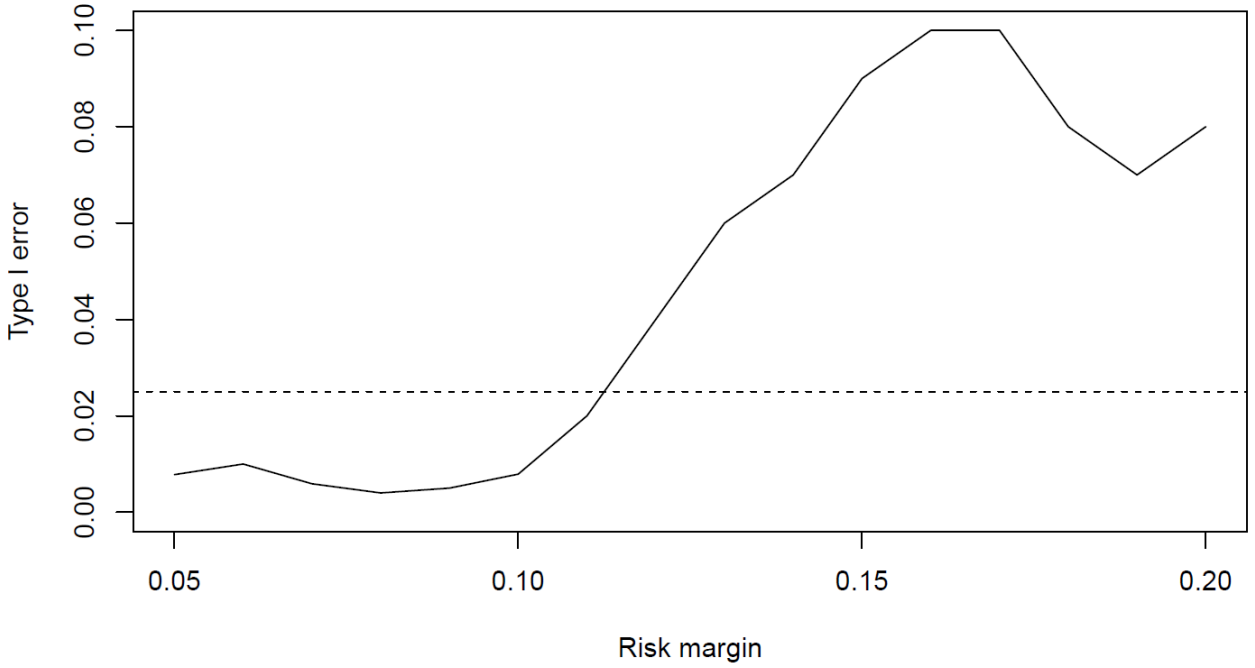
More appropriate for borrowing from many studies, the proposed Bayesian hierarchical model cannot reliably estimate between-study variability when borrowing from only one study. Additionally, posterior distribution inference, power, and type I error are sensitive to the specification of the τ^2 hyperprior distribution.

Lastly, in a Bayesian analysis, conclusions are stated in terms of posterior probabilities, e.g., $p(\pi_1 < 15\%)$. The sponsor's decision rule using the upper bound of the 95% credible interval is equivalent to a decision rule that concludes no safety risk when the posterior probability $p(\pi_1 < 15\%) = 0.975$. Whether this is an appropriate decision rule in a Bayesian framework is subject to discussion and negotiation between the sponsor and FDA.

In an internal meeting on July 3, 2019, DBRUP expressed concern that Milprosa might be a different product because of the new manufacturing process, and therefore, a standalone trial without borrowing is preferred. Trial 000293 (single arm, open label) was also not of the same design as trial DR-PGN-302 (randomized, active-controlled trial). Given clinical and statistical concerns, an advice letter, dated July 24, 2019, rejected the Bayesian design and asked the sponsor to propose a new frequentist analysis to rule out a spontaneous abortion rate of 15% with at least 80% power and a one-sided type I error of 0.025.

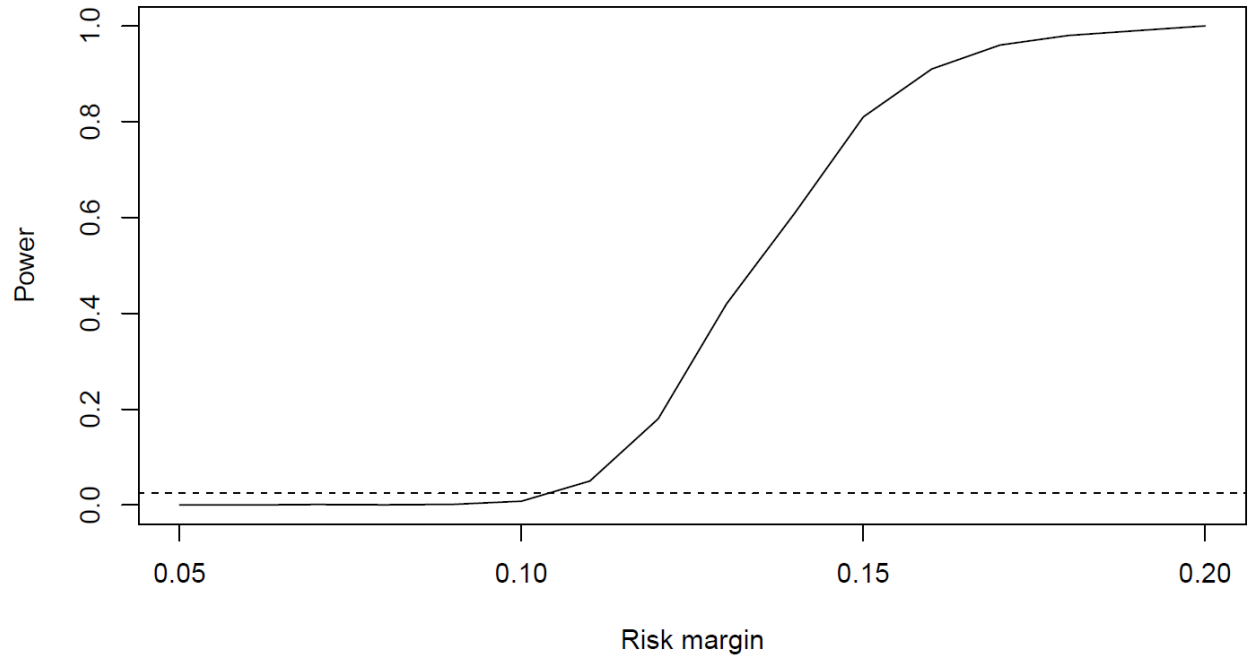
5 APPENDIX

Figure 4 Type I error as a function of risk margin, sample size = 215, $\mu \sim N($ [redacted] $), \tau^2 \sim IG($ [redacted] $)$.



Source: Reviewer's analysis

Figure 5 Power as a function of risk margin, sample size = 215, $\mu \sim N($ [redacted] $), \tau^2 \sim IG($ [redacted] $)$.



Source: Reviewer's analysis

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/

THANH V TRAN
08/29/2019 02:55:53 PM

CLARA KIM
08/29/2019 03:12:50 PM

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 #:	201110
Drug Name:	MILPROSA (Progesterone Vaginal Ring)
Indication(s):	Assisted Reproductive Technology
Applicant:	Ferring Pharmaceuticals Inc.
Date(s):	Submission Date: February 25, 2016; PDUFA Date: August 23, 2016
Review Priority:	Resubmission / Class 2
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Weiya Zhang, Ph.D.
Biometrics Team Leader:	Mahboob Sobhan, Ph.D.
Medical Division:	Division of Bone, Reproductive, and Urologic Products
Clinical Team:	Rhonda Hearns-Stewart, M.D., Clinical Reviewer Shelley Slaughter, M.D., Clinical Team Leader
Project Manager:	Nikia Morris

Keywords: NDA review, Clinical studies, Memorandum

Memorandum

This NDA is a resubmission following a Complete Response letter dated February 28, 2011 to address CMC issues in the label.

Efficacy results in the label will remain the same, and, therefore no statistical review is needed for this resubmission.

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

/s/

WEIYA ZHANG
06/22/2016

MAHBOOB SOBHAN
06/27/2016

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	201-110/N0000
Drug Name:	DR-2011 (Progesterone Vaginal Ring)
Indication(s):	To support embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an ART treatment program for infertile women
Applicant:	Teva Womens Health Research, Inc
Date(s):	Date of submission: 04/30/2010
	PDUFA due date: 02/28/2011
	Review completion date: 02/18/2010
Review Priority:	Standard
Biometrics Division:	Division of Biometrics 3
Statistical Reviewer:	Xin Fang, Ph.D., Primary Reviewer
Concurring Reviewers:	Mahboob Sobhan, Ph.D. Team Leader
Medical Division:	Division of Reproductive and Urologic Products
Clinical Team:	Christos Mastroyannis, M.D., Medical Reviewer
	Shelley R Slaughter, M.D., Medical Team Leader
Project Manager:	Celia Peacock

Keywords: Clinical studies, NDA review, Non-inferiority, Subgroup analysis

Table of Contents

LIST OF TABLES

1. EXECUTIVE SUMMARY

Based on the data from one study submitted in this application, the results of FDA statistical analysis showed that progesterone vaginal ring (DR-2011) was similar to Crinone 8% vaginal gel in younger women aged 18-34. At week 12, the clinical pregnancy rates were 48.2% and 46.1% in DR-2011 and Corinne treated women, respectively. The difference in pregnancy rate was 2.1% with the lower bound of the 95% confidence interval of -3.7%, which was within the pre-specified margin of -10%, a success criterion pre-specified in the protocol to demonstrate efficacy.

The treatment difference in 179 women aged 35-42 was -4.3%, numerically favoring Crinone, with the lower bound of the 95% confidence interval of -18.5%, thus exceeding the margin of -10%. In other words, in this older cohort, DR-2011 would retain only 53% of the Crinone effect, assuming a true rate of 40% for Crinone versus 0% for placebo. The wider confidence interval seen was due to limited number of subjects randomized in this age group. Similar results were also seen at week 8.

There are no statistical issues in establishing the non-inferiority of DR-2011 to Crinone for younger women aged 18-34. But, the study was not designed to show the non-inferiority of DR-2011 to Crinone in the older women aged 35-42. A study with a larger sample size would be required to exclude a clinically meaningful lower bound in this group

From a statistical perspective, this application provided adequate data to support the non-inferiority of DR-2011 to Crinone in supporting embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an assisted reproductive technology treatment program for younger infertile women aged 18-34. However, no definitive efficacy conclusion can be made for the limited number of women aged 35-42 studied in this trial.

2. INTRODUCTION

2.1 Overview

The applicant, Teva Womens Health Research, Inc, is seeking approval of DR-2011 progesterone vaginal ring (VR), in the treatment of in vitro fertilization through the luteal phase supplementation. DR-2011 has been developed as a flexible silicone (b) (4) progesterone VR administered weekly. Safety and efficacy data has been reported from a single clinical trial (DR-PGN-302) to support DR-2011 in the treatment of embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an assisted reproductive technology treatment (ART) program for infertile women. Study DR-PGN-302 was a randomized, single-blinded, multi-center, parallel-group, and active-controlled non-inferiority trial. Eligible women after screening pretreatment and egg retrieval were randomized in a ratio of 1:1 to one of the two treatment arms:

- DR-2011 (progesterone vaginal ring)
- Crinone® 8% (progesterone vaginal gel), 90 mg/day.

Women received 14 days of progesterone supplementation in the Embryo Transfer Cycle (ETC) and up to additional 8 weeks of progesterone if they became pregnant. The randomization was stratified by age (ages 18-34 or 35-42). The investigators and ultrasonographer were blinded to the treatment. However, the study subjects were unblinded. The study was conduced in 22 sites within the United States. The treatment duration was approximately up to 10 weeks.

The protocol specified primary endpoint was the clinical pregnancy (CP) rate defined as the percentage of subjects with visualization of a gestational sac with fetal heart motion present on ultrasound at 8 and 12 weeks of pregnancy. The clinical pregnancy rates at weeks 8 and 12 for each treatment group were reported per egg retrieval. The secondary efficacy endpoints included (1) live birth rate, (2) cycle cancellation rate, (3) rate of spontaneous abortion, (4) rate of biochemical pregnancy, and (5) rate of ectopic pregnancy.

The primary objective of the Study DR-PGN-302 was to establish that the clinical pregnancy rate in women using DR-2011 was not inferior to Crinone®. A non-inferiority margin of -10% was pre-specified. A total of 1300 women, 650 per treatment group, were planned to be enrolled. At the end of the study, 1297 women were randomized, treated and analyzed as shown in Table 2.1

Table 2.1: Summary of Clinical Study

Study	Study Site	Study Design	Number Randomized/ Study Regimens	Duration of Treatment
DR-PGN-302	US (22)	Multi-center, Single-blind, Active-controlled, Non-inferiority	**Planned**: DR-2011: 550 aged 18-34 100 aged 35-42 Crinone®: 550 aged 18-34 100 aged 35-42 **Analyzed**: DR-2011: 558 aged 18-34 88 aged 35-42 Crinone®: 560 aged 18-34 91 aged 35-42	Up to 10 weeks

2.2 Data Sources

The study report and additional information were submitted electronically. The data quality of the submission was within the acceptable limits. Analysis datasets and associated definition files were listed in Table 2.2.

Table 2.2: Data Sources

Study	File	Location
DR-PGN-302	Datasets	\\CDSESUB1\EVSPROD\NDA201110\0000\m5\datasets\dr-pgn-302\tabulations\legacy
	Definition	\\CDSESUB1\EVSPROD\NDA201110\0000\m5\datasets\dr-pgn-302\tabulations\legacy\define.pdf

2.3 Indication

DR-2011 is indicated for supporting embryo implantation and early pregnancy (up to 10 weeks post-embryo transfer) by supplementation of corpus luteal function as part of an ART treatment program for infertile women.

3. STATISTICAL EVALUATION

3.1 Overview of Study DR-PGN-302

3.1.1 Design and Objectives

Design: Study DR-PGN-302 was a randomized, single-blinded, active-controlled, parallel group, multi-center non-inferiority Phase-3 clinical trial. The primary objective of the study was to establish that the efficacy of DR-2011 was not inferior to Crinone® in terms of clinical pregnancy rate with a non-inferiority margin of -10%. The study was conducted across 22 sites in the United States. There were two treatment groups in the study: DR-2011 (progesterone VR) once weekly and Crinone® 8% (progesterone vaginal gel) once daily.

One day after egg retrieval, eligible subjects were randomized in a 1:1 ratio to receive either DR-2011 once weekly or Crinone® daily. The randomization was stratified by age (ages 18-34 or ages 35-42). Subjects were not blinded and were instructed to keep their treatment blinded to the investigator and the ultrasonographer.

Approximately 3 days after egg retrieval, subjects got embryo transfer. Two weeks later, subjects returned to their study sites to assess serum β-HCG. Subjects with β-HCG <5 mIU were withdrawn from the study as treatment failures. Subjects with β-HCG >5 mIU continued their progesterone treatment. Three days later, the serum β-HCG level was assessed again. About 3 weeks after egg retrieval, the serum β-HCG level was assessed to determine whether a subject was no longer pregnant. Subjects who were determined pregnant continued on their progesterone treatment and got their transvaginal ultrasound (TVU) 3 days later to determine the presence of an intrauterine gestation sac. Subjects who lacked an intrauterine gestational sac were withdrawn

from the study as treatment failures. All remaining subjects continued their treatment for about additional 2 weeks. Altogether, the treatment period was up to 10 weeks.

Primary Efficacy Endpoint: The primary efficacy endpoint was the **c**linical pregnancy rate at Weeks 8 and 12 defined as the proportion of subjects with visualization of a gestational sac with fetal heart motion present on ultrasound.

Secondary Efficacy Endpoints: The secondary endpoints were as follows:

- Live birth rate
- Cycle cancellation rate
- Rate of spontaneous abortion
- Rate of biochemical pregnancy
- Rate of ectopic pregnancy

The biochemical pregnancy was defined as (1) β-HCG level at Visit 6 reached at least 5 mIU; and (2) The β-HCG ratio of any two consecutive visits (blood draws) occurring at least two days apart with a ratio less than 1.0 between the later visit and the earlier visit; and (3) No gestational sac was ever seen on ultrasound; and (4) The pregnancy was not identified as ectopic.

Determination of Sample Size: A total of 1100 women aged 18-34 (550 per treatment group) would provide at least 90% power based on the following assumptions:

- A one-sided non-inferiority test of the clinical pregnancy rate at a significance level of 2.5%,
- An estimated clinical pregnancy rate of 50% for both Crinone-treated and DR-2011-treated subjects, and
- A non-inferiority margin (maximum allowable difference in clinical pregnancy rate between the two treatment groups) of 10%.

In addition, about 200 women aged 35-42 were going to be enrolled in this study for a total of 1300 women.

Definition of Analysis Sets (Population): The primary efficacy analysis set was the modified intent-to-treat population (MITT), which included all subjects having successful egg retrieval and having at least one dose of progesterone. The per-protocol (PP) set included the MITT subjects whose clinical pregnancy status could be determined by TVU at 8 or 12 weeks of pregnancy and did not have any major protocol deviations. The ovarian stimulation (OS) population included the MITT subjects who initiated ovarian stimulation using gonadotropins prior to progesterone therapy. However, analyses utilizing the OS were not undertaken. The eligible subject (ES) population included all subjects who were deemed eligible for study participation based on inclusion and exclusion criteria at screening and initiated ovarian stimulation using gonadotropins. The ES population was used to evaluate the overall cycle cancellation rate. The safety population included all subjects who received at least one dose of progesterone.

Handling of Missing Data: Subjects who terminated the study prior to 8 weeks of pregnancy (6 weeks after egg retrieval) were considered treatment failures at both 8 and 12 weeks.

Statistical Methods: The clinical pregnancy rates at Weeks 8 and 12 would be compared between the DR-2011 group and the Crinone group at one-sided alpha of 0.025. The 95% confidence interval (CI) for the difference in pregnancy rate was calculated using the normal approximation method. A non-inferiority margin of (b)(4)% was pre-specified in the statistical analysis plan although the study was powered with a margin of -10%. DR-2011 would be declared non-inferior to Crinone if the lower bound of the 95% CI for the difference in pregnancy rate was greater than (b)(4)% based on MITT population. The two co-primary endpoints of the CP rate (Weeks 8 and 12) would also be evaluated in subgroups classified by two age groups: 18-34 and 35-42, embryo transfer day, ovarian reserve as measured by (day 2-4 serum FSH, BMI, the type of insemination, and infertility diagnosis. However, no formal statistical comparisons would be made between the two treatment groups within these subgroups.

Multiple Comparisons/Multiplicity: Efficacy at weeks 8 and 12 was compared at one-sided alpha of 0.025.

3.1.2 Reviewer's Comments on the Design

The study design was adequate for testing the non-inferiority hypothesis based on a margin of -10% in women aged 18-34. A total of 179 women aged 15-42 randomized in this study were inadequate to test the same hypothesis. The sponsor erroneously indicated margin of (b)(4)% to declare success, while the study was designed using a margin of -10%. Our evaluation and statistical conclusion will be based on -10% margins.

3.2 Results: Study DR-PGN-302

3.2.1 Subject Disposition

Table 3.2.1 shows the subject disposition by pre-defined analysis datasets. At 22 US sites, a total of 1299 subjects were randomized. The MITT population consisted of 1118 subjects aged 18-34, which was well over planned 1100 subjects for younger women subgroup. The per protocol population consisted of 1210 subjects as total. The safety population was the same as the MITT population except that actual treatments were used in the safety population. The discontinuation rate and adverse event (AE) rates were similar between the two treatment groups. A total of 54.3% subjects discontinued from the study. The major reason for discontinuation was not being pregnant (45.0%) and AE (6.2%).

We found that seven subjects out of 1299 subjects did not have their times of last dose recorded. Two of these seven subjects (b) (6) (Crinone) and (b) (6) (DR-2011) were identified never took the study medications and were excluded from the MITT dataset. The other five subjects were included in the MITT set. Three (2 in Crinone and 1 in DR-2011) of these five subjects were not pregnant. Two (Crinone) of these five subjects were pregnant at Week 8 only. All subjects in the MITT set were analyzed using the randomized treatment although there were three subjects ever

provided wrong study medication. These minor mistakes did not appear to affect the non-inferiority evaluation.

Four study sites enrolled more than 100 subjects (40% of the total): 109 in Site 7, 114 in Site 17, 175 in Site 19, and 124 in Site 20. Efficacy in these four sites was similar to that in combined other sites.

Table 3.2.1: Disposition of Subjects in Study DR-PGN-302

Category	DR-2011 N (%)	Crinone N (%)	Total
Subjects with Egg Retrieval	-	-	1299
Randomized Subjects	647 (100%)	652 (100%)	1299 (100%)
Completed Study	300 (46.4%)	294 (45.1%)	594 (45.7%)
Discontinued Study	347 (53.6%)	358 (54.9%)	705 (54.3%)
Adverse Events	44 (6.8%)	36 (5.5%)	80 (6.2)
Did Not Meet Protocol Requirements	3 (0.5%)	3 (0.5%)	6 (0.5%)
Non-compliance of Study Drug	4 (0.6%)	2 (0.3%)	6 (0.5%)
Investigator Discretion	1 (0.2%)	2 (0.3%)	3 (0.2%)
Patient not pregnant or no longer Pregnant	284 (44.0%)	301 (46.2%)	585 (45.0%)
Lost to Follow-up	4 (0.6%)	7 (1.1%)	11 (0.8%)
Voluntary Withdrawal	2 (0.3%)	2 (0.3%)	4 (0.3%)
Other	5 (0.8%)	5 (0.8%)	10 (0.8%)
MITT Analysis Set	646	651	1297
Per Protocol Set	602	608	1210
Safety Set	646	651	1297

Source: Reviewer's analysis

3.2.2 Patient Demographics and Baseline Characteristics

The baseline characteristics such as age, body mass index (BMI), race and prior pregnancy are similar between the two treatment groups as shown in Table 3.2.2.

Table 3.2.2: Subject Demographic and Baseline Characteristics (MITT)

Demographic Variable	DR-2011 (N=646)	Crinone (N=651)	Total (N=1297)
Mean age (SD)	31.7 (3.73)	31.6 (3.86)	31.7 (3.80)
Body Mass Index (BMI) (SD)	25.3 (4.64)	25.7 (5.01)	25.5 (4.83)
Race [n (%)]:			
Caucasian/White	519 (80.3%)	507 (77.9%)	1026 (79.1%)
Black/African American	54 (8.4%)	48 (7.4%)	102 (7.9%)
Hispanic/Latino	34 (5.3%)	51 (7.8%)	85 (6.6%)
Asian	34 (5.3%)	37 (5.7%)	71 (5.5%)
Other	5 (0.8%)	8 (1.2%)	13 (1.0%)
Prior Pregnancy:			
Yes	312 (48.3%)	335 (51.5%)	647 (49.9%)

Sources: Reviewer's analysis

3.2.3 Primary Efficacy

The primary efficacy variable was the clinical pregnancy rate at both Week 8 and Week 12 using MITT population. Table 3.2.3 shows the results at both weeks for all subjects and for two stratified age subgroups: 18-34 and 35-42. At both weeks 8 and 12, the differences in pregnancy rates between DR-2011 and Crinone were 0.8% and 1.3%, respectively. The lower bound of the 95% CI for these rate differences (DR-2011 – Crinone) was -4.6%, and -4.1%, within the pre-specified margin of -10%. Therefore, DR-2011 was non-inferior to Crinone in all women. Similar conclusion can also be drawn for women aged 18-34.

However, for women aged 35-42, the treatment differences (DR-2011 – Crinone) were -2.0% and -4.3% at Week 8 and Week 12, respectively. The lower bound of the 95% confidence interval at both weeks exceeded the pre-specified margin of -10%, indicating DR-2011 was inferior to Crinone. This also means in the older women DR-2011 would retain only 53% of the Crinone effect, assuming a true rate of 40% for Crinone versus 0% for placebo.

Table 3.2.3: Clinical Pregnancy Rate per Retrieval by Age and Week: MITT Population

Age	Week	DR-2011		Crinone		Treatment Diff. (95% C.I.) (DR-2011 – Crinone)
		n/N	Percent	n/N	Percent	
18-34	8	275/558	49.3%	269/560	48.0%	1.2% (-4.6%, 7.1%)
	12	269/558	48.2%	258/560	46.1%	2.1% (-3.7%, 8.0%)
35-42	8	35/88	39.8%	38/91	41.8%	-2.0% (-16.4%, 12.4%)
	12	31/88	35.2%	36/91	39.6%	-4.3% (-18.5%, 9.8%)
All	8	310/646	48.0%	307/651	47.2%	0.8% (-4.6%, 6.3%)
	12	300/646	46.4%	294/651	45.2%	1.3% (-4.1%, 6.7%)

Sources: Reviewer's analysis

3.2.4 Secondary Efficacy

Table 3.2.4 shows live birth rates based on the MITT population. The difference in rates for the entire population was 1.9% with the 95% CI of (-3.5%, 7.3%) between DR-2011 and Crinone. Similar results were found in the younger women aged 18-34. For the older women aged 35-42, the point estimate was reduced to -2.2% with the 95% CI of (-16.2%, 11.8%). Results of the live birth rate supported the results of the clinical pregnancy rate.

Table 3.2.4: Live Birth Rate per Retrieval: MITT Population

Age	DR-2011		Crinone		Treatment Diff. (95% C.I.) (DR-2011 – Crinone)
	n/N	Percent	n/N	Percent	
18-34	262/558	47.0%	249/560	44.5%	2.5% (-3.3%, 8.3%)
35-42	30/88	34.1%	33/91	36.3%	-2.2% (-16.2%, 11.8%)
All	292/646	45.2%	282/651	43.3%	1.9% (-3.5%, 7.3%)
Sources: Reviewer's analysis					

3.3 Reviewer's Comments on the Efficacy Results

Results of our analysis showed that DR-2011 was similar to Crinone in supporting embryo implantation and early pregnancy in women aged 18-42. However, same conclusion can not be drawn for the older women aged 35-42, based on the limited number of women studied in this age group.

3.4 Evaluation of Safety

Evaluation of safety can be found in the Medical Officer's review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Age, Race

Since age was a stratification factor during the randomization, the results of age subgroup were discussed in the primary efficacy results Section (3.2.3).

In African-American women, the point estimate of the rate difference was -22.2% with the 95% CI of (-40.7%, -3.7%) at Week 8 and was -16.2% with the 95% CI of (-34.8%, 2.4%) at Week 12. This result suggested DR-2011 may not be non-inferior to Crinone for African-American women. Similar results were found in younger African-American women aged 18-34 (Table A.2). However, the sample size in this subgroup was not adequate to draw any meaningful conclusion.

4.2 Other Special/Subgroup Populations

Tables A.3-A.6 shows the clinical pregnancy rates by subgroups of BMI, insemination, infertility duration, and infertility diagnosis. Results in these subgroups are for exploratory purposes only. Due to small numbers in these subgroups, no meaningful conclusion was made.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The study was adequately designed to test the non-inferiority hypothesis in women aged 18-34. The sponsor indicated that a stratified (Age 18-34 and 35-42) randomization was used in this study, but we did find that only 13.8% of the women were randomized for older age group. Appropriate stratified randomization would have ensured more balanced number of women in each age strata. Therefore, testing the non-inferiority hypothesis in older women was less than adequate.

5.2 Conclusions and Recommendations

The data from one active-controlled trial supports the non-inferiority of DR-2011 to Crinone for women aged 18-34 in the treatment of human embryo implantation and early pregnancy up to 10 weeks post-embryo transfer at the non-inferiority margin of -10% (DR-2011 – Crinone). For older women aged 35-42, however, the lower bound of the 95% CI was -18.5%, much lower than -10%, although the point estimate was -4.3%. This was due to very small number of women randomized in this group. Therefore, the non-inferiority of DR-2011 to Crinone in older women aged 35-42 has not been established.

From a statistical perspective, the efficacy of DR-2011 was similar to Crinone across all women studied in this trial, including the younger women aged 18-34. We could not draw similar conclusion for older women aged 35-42. We recommended that the label should also include a statement that efficacy in older women aged 35-42 has not been established.

Appendix: Tables

Table A.1: Pregnancy Rate by Race: MITT Population

Race	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
African-American	8	15/54	27.8%	24/48	50.0%	-22.2% (-40.7%, -3.7%)
	12	16/54	29.6%	22/48	45.8%	-16.2% (-34.8%, 2.4%)
Asian	8	16/34	47.1%	14/37	37.8%	9.2% (-13.7%, 32.1%)
	12	15/34	44.1%	14/37	37.8%	6.3% (-16.6%, 29.1%)
Caucasian	8	255/519	49.1%	243/507	47.9%	1.2% (-4.9%, 7.3%)
	12	246/519	47.4%	232/507	45.8%	1.6% (-4.5%, 7.7%)
Hispanic	8	22/34	64.7%	24/51	47.1%	17.6% (-3.5%, 38.8%)
	12	21/34	61.8%	24/51	47.1%	14.7% (-6.6%, 36.0%)

Sources: Reviewer's analysis

Table A.2: Pregnancy Rate by Race in Women Aged 18-34: MITT Population

Race	Age	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
			n/N	Percent	n/N	Percent	
African-American	18-34	8	15/49	30.6%	19/41	46.3%	-15.7% (-35.7%, 4.3%)
		12	16/49	32.7%	18/41	43.9%	-11.2% (-31.3%, 8.8%)
Asian	18-34	8	15/28	53.6%	14/31	45.2%	8.4% (-17.0%, 33.9%)
		12	14/28	50.0%	14/31	45.2%	4.8% (-20.7%, 30.3%)
Caucasian	18-34	8	211/432	48.8%	223/446	50.0%	1.2% (-5.5%, 7.8%)
		12	201/432	46.5%	218/446	48.9%	2.4% (-4.3%, 9.0%)
Hispanic	18-34	8	20/30	66.7%	23/48	47.9%	18.8% (-3.3%, 40.8%)
		12	19/30	63.3%	23/48	47.9%	15.4% (-6.9%, 37.7%)

Sources: Reviewer's analysis

Table A.3: Pregnancy Rate by Body Mass Index (BMI): MITT Population

BMI	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
<18.5	8	5/11	45.5%	9/18	50%	-4.5% (-42.0%, 32.9%)
	12	5/11	45.5%	9/18	50%	-4.5% (-42.0%, 32.9%)
18.5<=BMI<25	8	171/343	49.9%	161/331	48.6%	1.2% (-6.3%, 8.8%)
	12	165/343	48.1%	155/331	46.8%	1.3% (-6.3%, 8.8%)
25 <=BMI<30	8	90/186	48.4%	78/168	46.4%	2.0% (-8.5%, 12.4%)
	12	89/186	47.8%	76/168	45.2%	2.6% (-7.8%, 13.0%)
30 <=BMI<34	8	29/65	44.6%	35/80	43.8%	0.9% (-15.4%, 17.1%)
	12	27/65	41.5%	34/80	42.5%	-1.0% (-17.1%, 15.2%)
34 <=BMI<38+	8	14/39	35.9%	24/54	44.4%	-8.5% (-28.6%, 11.5%)
	12	13/39	33.3%	20/54	37.0%	-3.7% (-23.3%, 15.9%)
Sources: Reviewer's analysis						

Table A.4: Pregnancy Rate by Insemination: MITT Population

Type of Insemination	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
Conventional	8	118/215	54.9%	95/203	46.8%	8.1% (-1.5%, 17.6%)
IVF	12	119/215	55.3%	90/203	44.3%	11.0% (1.5%, 20.5%)
ICSI	8	189/414	45.7%	204/423	48.2%	-2.6% (-9.3%, 4.2%)
	12	178/414	43.0%	196/423	46.3%	-3.3% (-10.1%, 3.4%)
Both	8	3/12	25.0%	8/19	42.1%	-17.1% (-50.2%, 16.0%)
	12	3/12	25.0%	8/19	42.1%	-17.1% (-50.2%, 16.0%)
Sources: Reviewer's analysis						

Table A.5: Pregnancy Rate by Infertility Duration: MITT Population

Duration of Infertility	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
<1 year	8	18/32	56.3%	21/42	50.0%	6.3% (-16.6%, 29.1%)
	12	17/32	53.1%	20/42	47.6%	5.5% (-17.5%, 28.5%)
1<=years<2	8	72/138	52.2%	83/184	45.1%	7.1% (-3.9%, 18.1%)
	12	68/138	49.3%	80/184	43.5%	5.8% (-5.2%, 16.8%)
2<=years<3	8	91/170	53.5%	69/158	43.7%	9.9% (-0.9%, 20.6%)
	12	88/170	51.8%	66/158	41.8%	10.0% (-0.8%, 20.7%)
>= 3 years	8	129/306	42.2%	134/267	50.2%	-8.0% (-16.2%, 0.1%)
	12	127/306	41.5%	128/267	47.9%	-6.4% (-14.6%, 1.7%)
Sources: Reviewer's analysis						

Table A.6: Pregnancy Rate by Infertility Diagnosis: MITT Population

Infertility Diagnosis	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
Tubal Factor	8	69/164	42.1%	82/169	48.5%	-6.4% (-17.1%, 4.2%)
	12	69/164	42.1%	78/169	46.2%	-4.1% (-14.7%, 6.6%)
Ovulatory Dysfunction	8	103/193	53.4%	95/184	51.6%	1.7% (-8.3%, 11.8%)
	12	100/193	51.8%	93/184	50.5%	1.3% (-8.8%, 11.4%)
Endomtriosis	8	72/146	49.3%	73/150	48.7%	0.6% (-10.7%, 12.0%)
	12	69/146	47.3%	71/150	47.3%	-0.1% (-11.4%, 11.3%)
Male Factor	8	123/247	49.8%	128/259	49.4%	0.4% (-8.3%, 9.1%)
	12	117/247	47.4%	123/259	47.5%	-0.1% (-8.8%, 8.6%)
Idiopathic Factor	8	73/160	45.6%	57/138	41.3%	4.3% (-7.0%, 15.6%)
	12	69/160	43.1%	55/138	39.9%	3.3% (-7.9%, 14.5%)
Other Factor	8	30/69	43.5%	24/74	32.4%	11.0% (-4.8%, 26.9%)
	12	29/69	42.0%	23/74	31.1%	10.9% (-4.8%, 26.7%)

Sources: Reviewer's analysis

Table A.7: Pregnancy Rate by Two Older age groups: MITT Population

Age	Week	DR-2011		Crinone		Difference (DR-2011 – Crinone) (95% Confidence Interval)
		n/N	Percent	n/N	Percent	
35-37	8	21/50	42.0%	22/50	44.0%	-2.0% (-21.4%, 17.4%)
	12	20/50	40.0%	21/50	42.0%	-2.0% (-21.3%, 17.3%)
38-42	8	14/38	36.8%	16/41	39.0%	-2.2% (-23.6%, 19.2%)
	12	11/38	28.9%	15/41	36.6%	-7.6% (-28.3%, 13.0%)

Sources: Reviewer's analysis

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

/s/

XIN FANG
02/18/2011

MAHBOOB SOBHAN
02/18/2011