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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	22-252/N000
Drug Name:	Estradiol Valerate/Dienogest (EV/DNG)
Indication(s):	Prevention of Pregnancy and Treatment of Heavy and/or Prolonged Menstrual Bleeding
Applicant:	Bayer HealthCare Pharmaceuticals Inc.
Date(s):	Date of submission: 07/06/2009
	PDUFA due date: 05/06/2010
	Review completion date: 03/04/2010
Review Priority:	Standard
Biometrics Division:	Division of Biometrics 3
Statistical Reviewer:	Xin Fang, Ph.D., Statistical Reviewer
Concurring Reviewers:	Mahboob Sobhan, Ph.D., Team Leader
Medical Division:	Division of Reproductive and Urological Drug Products
Clinical Team:	Gerald Willett, M.D., Clinical Reviewer
	Lisa Soule, M.D., Team Leader
Project Manager:	Pamela Lucarelli

Keywords: NDA review, Clinical studies, ANCOVA, Multi-center

Table of Contents

1.   EXECUTIVE SUMMARY

1.1   Conclusions and Recommendations

The data supported the efficacy of Estradiol Valerate/Dienogest (EV/DNG) in the prevention of pregnancy as demonstrated by the Pearl Index (PI) of < 2.0 in both the North American and the European studies.

From a statistical perspective, this application provided adequate data to support the efficacy of the EV/DNG as oral contraceptive.

1.2   Brief Overview of Clinical Studies

The sponsor, Bayer HealthCare Pharmaceuticals Inc, submitted two open-label uncontrolled studies (304742 and 306660) to support the efficacy of EV/DNG as oral contraceptive, and two double-blind placebo-controlled studies (308960 and 308961) to support the efficacy of EV/DNG in the treatment of DUB. The primary objective of the OC studies was to determine the contraceptive reliability of EV/DNG in terms of the Pearl Index, while the primary objective of the DUB studies was to determine the efficacy in the treatment of DUB in terms of overall success rate in 90-day efficacy period.

Study 304742 enrolled 499 healthy women of 18-35 years of age in 22 centers in US and 9 centers in Canada, while Study 306660 enrolled 1391 healthy woman of 18-50 years of age across three European countries: 17 centers in Austria, 28 centers in Germany, and 5 centers in Spain. Studies 304742 and 306660 had only one treatment group (EV/DNG). The treatment duration in Study 304742 is up to 28 cycles of 28 days each and in Study 306660 is 20 cycles of 28 days each.

Study 308960 enrolled 190 women having at least one of the three DUB symptoms (excessive, prolonged, and frequent uterine bleeding) in 33 US centers and 9 Canadian centers, while Study 308961 enrolled 231 women under the same DUB criteria in 2 Australian centers, 2 Czech Republic centers, 3 Finland centers, 5 German Centers, 3 Hungary centers, 3 The Netherlands centers, 5 Poland Centers, 3 Sweden centers, 3 UK centers and 4 Ukraine centers. The treatment duration in both Studies 308960 and 308961 is 7 cycles of 28 days each.

1.3   Statistical Issues and Findings

We noted two minor issues with regards to calculation of pregnancy rates. First, in OC studies, the sponsor estimated pregnancy rate (Pearl Index) using exposure time in days, rather than

woman's menstrual cycle. Secondly, we also noted that backup days in some patients in the North American study were not counted. Therefore, we corrected for the above two issues and concluded that the results were still consistent.

The sponsor also presented pooled (across studies) results, but in our evaluation we looked at each study result separately, since pooling across US and Europe was not appropriate due to differences in demographic characteristics. Moreover, each study must demonstrate efficacy prior to post-hoc pooled analysis.

(b) (4)

2. INTRODUCTION

2.1 Overview

The sponsor, Bayer Healthcare Pharmaceuticals, is seeking approval of Estradiol Valerate/Dienogest (EV/DNG), for the two following indications:

- Prevention of pregnancy
- Treatment of heavy and/or prolonged menstrual bleeding (HMB) in women without organic pathology who choose to use an oral contraceptive as their method of contraception.

The EV/DNG tablets were developed as a combined oral contraceptive (OC), four-phase, 28-day, sequential regimen:

Cycle days 1 to 2	3.0 mg EV
Cycle days 3 to 7	2.0 mg EV + 2.0 mg DNG
Cycle days 8 to 24	2.0 mg EV + 3.0 mg DNG
Cycle days 25 to 26	1.0 mg EV
Cycle days 27 to 28	Placebo

One component of this regimen, EV, is currently approved only as an inject product (Delestrogen®) in the US. In many other countries, EV is used as a component of hormonal therapy (HT) products and is an approved component of OC product in Finland. The other component, DNG, is an approved component of a HT product (with EV) and a monophasic OC in several European countries.

To support the safety and efficacy of EV/DNG in OC, clinical data from two pivotal studies (P306660/A35179, P304742/A39818) were submitted as well as a short period (7 cycle) study (A35644). The details of these studies were listed in Table 2.1 below. In addition, safety data from a QT study (A35653) was also submitted to rule out any abnormal QT prolongation or other cardiac abnormality.

To support the safety and efficacy of EV/DNG in the treatment of dysfunctional uterine bleeding (DUB), clinical data from two pivotal studies (P308961/A42568, P308960/A29849) were submitted.

The OC efficacy review will be based on data from Studies P306660 and P304742. The DUB efficacy review will be based on data from Studies P308960 and P308961.

			Table 2.1 Summary of Pivotal Studies		
Indica tion	Study	Study Site [1]	Study Design	Number Randomized/ Study Regimens	Duration of Treatment
OC	A35179	Austria: 17 Germany: 28 Spain: 5	Multi-center, open-label, uncontrolled, one-arm study	**Planned**: 1200 (including a subgroup of 250 for endometrial biopsy) **Analyzed**: 1377 (including a subgroup of 266 for endometrial biopsy)	20 cycles of 28 days each
	A39818	US: 22 Canada: 9	Multicenter, open-label, uncontrolled one-arm study	**Planned**: 480 **Analyzed**: 490 subjects (Full Analysis Set), 305 subjects (Per Protocol Set).	Up to 28 cycles of 28 days each.
	A35644	Germany: 19 Czech Republic: 5 France: 10	Multi-center, double-blind, double-dummy, active-controlled, randomized study	**Planned**: 800 (400 per treatment group; 200 per age stratum) **Analyzed**: 798 (399 per treatment group; 198 to 201 per age stratum) (full analysis set; FAS)	7 cycles of 28 days each
HMB	A42568	Australia: 2 Czech Republic: 2 Finland: 3 Germany: 5 Hungary: 3 The Netherlands: 3 Poland: 5 Sweden: 3 UK: 3 Ukraine: 4	Multicenter, double-blind, randomized, parallel-group, placebo-controlled study	**Planned**: 180 patients to be randomized (120 EV/DNG and 60 placebo); **Analyzed**: 231 patients randomized (149 EV/DNG and 82 placebo) ITT, 226 (145 EV/DNG and 81 placebo) SAF, 89 (55 EV/DNG and 34 placebo) PPS	196 days (7 cycles of 28 days)
	A29849	US: 33 Canada: 9	Multicenter, double-blind, randomized, parallel-group, placebo-controlled study	**Planned**: 180 patients (120 EV/DNG and 60 placebo) **Analyzed**: 190 patients intent-to-treat (120 EV/DNG and 70 placebo),185 patients safety (119 EV/DNG and 66 placebo), 68 patients per protocol (41 EV/DNG and 27 placebo)	196 days (7 cycles of 28 days each)
1: only include those sites having randomized patients					

2.2   Data Sources

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

Table 2.2 Data Sources		
Study	File	Location
A35179	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-a35179\analysis\
(P306660)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report- a35179\analysis\define.pdf
A39818	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-a39818\analysis\
(P304742)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report- a39818\analysis\define.pdf
A35644	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-a35644\analysis\
(P304004)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report- a35644\analysis\define.pdf
A42568	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-a42568analysis\
(P308961)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report- a42568\analysis\define.pdf
A29849	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-a29849\analysis\
(P308960)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report- a29849\analysis\define.pdf
ISE	Datasets	\\CDSESUB1\EVSPROD\NDA022252\0000\m5\datasets\study-report-ise\analysis\
(OC)	Definition	\\CDSESUB1\EVSPROD\ NDA022252\0000\m5\datasets\study-report-ise\analysis\define.pdf

2.3   Indication

The EV/DNG is indicated for 1) oral contraceptive, and 2) treatment of DUB.

3.   STATISTICAL EVALUATION

Our evaluation consisted of study description followed by a discussion on the results of our analysis as depicted in Sections 3.1-3.2 and in Sections 3.3-3.4, respectively.

3.1   Overview of Study P304742 and Study P306660 for OC

The efficacy evaluation for OC focused on two studies: P304742 and P306660, although the sponsor submitted data from an additional Phase-3 study: P304004. The seven-cycle treatment in Study P304004 was too short to do comparable PI calculation and therefore was not reviewed for the efficacy of OC. The primary efficacy endpoint was evaluated by Pearl Index (PI).

In addition, the PI was evaluated separately for each of the two studies although the sponsor reported the PIs in the integrated summary of efficacy. Study P306660 was conducted in Europe and Study P304742 was conducted in US and Canada. The heterogeneity of the PIs between the European and US/Canada was expected and observed. Therefore, our conclusions were based on the results from each study separately, rather than combining the studies.

3.1.1   Design and objectives

Design: Both studies were multi-center, open-label, uncontrolled. In Study P306660, the sponsor planned to enroll 1200 females with ages between 18 and 50 across 50 centers: 28 in Germany, 17 in Austria and 5 in Spain. In Study P304742, the sponsor planned to enroll 480 females with age between 18 and 35 across 31 centers: 22 in the US and 9 in Canada. The primary objectives in both studies were to investigate the contraceptive efficacy, cycle control and safety of EV/DNG.

Both studies had only one treatment group. In Study P304742, each subject was planned to receive at least 13 cycles of treatment with approximately 300 women treated up to 28 cycles, and in Study P306660 each subject was planned to receive 20 cycles of treatment. In both studies, each cycle pack of EV/DNG contained one-cycle course (28 days) of treatment. And each subject took one tablet per day in her cycle as follows:

Cycle days	No. of intake days	Content
1 – 2	2	3.0 mg estradiol valerate
3 – 7	5	2.0 mg estradiol valerate + 2.0 mg dienogest
8 – 24	17	2.0 mg estradiol valerate + 3.0 mg dienogest
25 – 26	2	1.0 mg estradiol valerate
27 – 28	2	placebo

If a subject missed the tablets in less than 12 hours, she could take the tablets as soon as she remembered. Otherwise, she should take the medication in a rule as follows:

Text Table 1: Procedures if Tablet Intake is Greater Than 12 Hours Late

Days	Content of EV/DNG	Delay of <u>more than 12 hours</u>
1 – 2	3.0 mg EV	1. Take missed tablet immediately and the following tablet as usual, 2. Use of back-up contraception until <u>day 9</u>
3 – 7	2.0 mg EV + 2.0 mg DNG	1. Take missed tablet immediately and the following tablet as usual, 2. Back-up contraception for the next <u>7 days</u>
8 – 17	2.0 mg EV + 3.0 mg DNG	
18 – 24	2.0 mg EV + 3.0 mg DNG	1. Take missed tablet and *continue tablet intake as usual (use up the blister in the given sequence)* 2. Use of back-up contraception until <u>day 9</u> of the following cycle
25 – 26	1.0 mg EV	intake of missed tablet (no further action)
27 – 28	Placebo	

DNG = dienogest; EV = estradiol valerate.

Treatment compliance was defined as the total number of days that a tablet was taken divided by the study medication duration. The study medication duration was defined as the time between the day of first drug administration as recorded in the diary and the day of last drug administration, as determined by the End of Study Medication page.

Primary Efficacy Endpoints: In both studies, number of observed pregnancies was the primary efficacy variable. The Pearl Index was the statistics for this efficacy variable. The Kaplan-Meier estimate of probability of pregnancies was the supportive statistics.

Secondary Efficacy Endpoints: Secondary efficacy parameters were bleeding pattern and cycle control parameters.

Determination of Sample Size: In Study P306660, assuming a Pearl Index of 1 and dropout rate of 30%, 881 volunteers aged 18 to 35 years were needed in order to obtain a 2-sided 95% confidence interval (CI) for the PI such that the difference between the upper limit of the confidence interval and the point estimate did not exceed 1 with a probability of at least 90%.

In Study P304742, the sample size was not based on statistical consideration. It was decided to enroll 480 subjects so that about 400 subjects could complete 13×28-day cycles of treatment.

Definition of Analysis Sets (Population): In both studies, analysis sets were (1) Full Analysis Set (FAS), and (2) Per Protocol Set (PPS). The FAS included all subjects who took at least one pill of study medication and had at least one on-treatment observation. The PPS included all subjects in FAS who did not have any major violation of the inclusion/exclusion criteria, major non-compliance with the study medication intake schedule within the first 7 cycles of treatment, and other major deviation from the protocol within the first 7 cycles of treatment.

Handling of Missing Data: The missing dates of start and end of treatment were replaced by the first and last available date of treatment, respectively. Missing day-part was replaced by 15 and missing month-and day-part was replaced by 1st of July. In case of a visit date was used to

replace the start date, 12 days were added for treatment length. If end date was replaced by a visit date, 0 days were added to calculate the treatment length for Visits 3-6, 8-11. For Visit 7 (follow-up under main protocol) or Visit 12, 24 days were subtracted for the treatment length.

The missing bleeding events were replaced using a worst case approach. A diary that ended before the end of the reference period (90 days) was still evaluated if its length was at least 60 days. The correction factor was the length of the reference period divided by the length of the diary.

Pooling of Sites: Center effect (heterogeneity) was not evaluated.

Statistical Methods: The Pearl Index was calculated by $\dfrac{\text{Number of unintended pregnancies}}{\text{Number of women years}} \times 100$.

The upper bound of PI was calculated by $I_u = \dfrac{1}{2E} \chi^2_{0.975, 2(x+1)}$, where x was the number of unintended pregnancies and E was the exposure in 100 woman years. The cumulative failure rate was estimated by Kaplan-Meier estimate.

Multiple Comparisons/Multiplicity: There was no multiplicity issue with regards to the primary efficacy evaluation.

3.1.2 Reviewer's Comments on the Design

The design of Study P306660 appeared to be adequate, except that the exposure unit (either cycle or day) was not clearly defined in the report. In this review, however, we will calculate the Pearl Index using both cycle and day.

3.2 Results: Studies P306660 and P304742

3.2.1 Subject Disposition

As shown in Table 3.2.1, there were 51.3% drop-outs in the US study compared to 21% in the European study with mostly due to adverse event and lost to follow-up. Discontinuations due to withdrawal was much higher in the US study (9.8%) compared to European study (1.4 %).

Table 3.2.1: Summary of Patient Disposition by Study and Treatment			
		Treatment Group	
Study	**Disposition**	**EV/DNG N (%)**	**Total N (%)**
304742 (US/Canada)	Randomized	499 (100%)	499 (100%)
	Completed (13 Cycles)	243 (48.7%)	243 (48.7%)
	Discontinued:	256 (51.3%)	256 (51.3%)
	Withdrawal	49 (9.8%)	49 (9.8%)
	Protocol deviation	7 (1.4%)	7 (1.4%)
	Adverse event	72 (14.4%)	72 (14.4%)
	Lost	65 (13.0%)	65 (13.0%)
	Pregnancy	6 (1.2%)	6 (1.2%)
	Death	0 (0.0%)	0 (0.0%)
	Other	58 (11.6%)	58 (11.6%)
306660 (Europe)	Randomized	1391 (100%)	1391 (100%)
	Completed (13 Cycles)	1074 (77.2%)	1074 (77.2%)
	Discontinued:	295 (21.2%)	295 (21.2%)
	Withdrawal	20 (1.4%)	20 (1.4%)
	Protocol deviation	26 (1.9%)	26 (1.9%)
	Adverse event	140 (10.1%)	140 (10.1%)
	Lost	26 (1.9%)	26 (1.9%)
	Pregnancy	11 (0.8%)	11 (0.8%)
	Death	1 (0.07%)	1 (0.07%)
	Other	71 (5.1%)	71 (5.1%)
Source: Table 6 in Study 304742 and Text Figure 2 (TF 2) in Study 306660			

3.2.2 Patient Demographics and Baseline Characteristics

Patient's age and BMI were similar in both studies. However, there were disparities between two studies in terms of ethnicity, smoking history and alcohol consumption as show in Table 3.2.2.

Table 3.2.2. Subject Demographic and Baseline Characteristics (FAS)				
Variable/Parameters	**Study 304742 (US/Canada)**		**Study 306660 (Europe)**	
	EV/DNG 18-35 years of age N=490	**Total N=490**	**EV/DNG 18-35 years of age N=998**	**Total N=1377**
Age (SD, Min-Max)	25.1 (4.42, 18-35)	25.1 (4.42, 18-35)	26.2 (4.7, 18-35)	30.3 (7.9, 18-50)
BMI* (SD)	23.3 (3.25)	23.3 (3.25)	22.434 (2.855)	22.808 (2.916)
Ethnic group — Caucasian	371 (76%)	371 (76%)	990 (99.2%)	1366 (99.2%)
Ethnic group — Black	34 (7%)	34 (7%)	2 (0.2%)	2 (0.1%)
Ethnic group — Hispanic	64 (13%)	64 (13%)	3 (0.3%)	3 (0.2%)
Ethnic group — Asian	16 (0.3%)	16 (3%)	1 (0.1%)	4 (0.3%)
Ethnic group — Other	5 (1%)	5 (1%)	2 (0.2%)	2 (0.1%)
Smoking history — Yes	166 (34%)	166 (34%)	271 (27.2%)	273 (19.8%)
Smoking history — No	324 (66%)	324 (66%)	727 (72.8%)	1104 (80.2%)
Alcohol consumption — Never	46 (9%)	46 (9%)	147 (14.7%)	187 (13.6%)
Alcohol consumption — Seldom	130 (27%)	130 (27%)	594 (59.5%)	839 (60.9%)
Alcohol consumption — Occasionally	203 (41%)	203 (41%)	234 (23.4%)	320 (23.2%)
Alcohol consumption — Regularly	75 (15%)	75 (15%)	23 (2.3%)	31 (2.3%)
Alcohol consumption — Missing	36 (7%)	36 (7%)	-	-

Source: Table 7 in Study 304742, Table 20-22, 28-29, 33-34 in Study 306660, and reviewer's analysis
*: BMI stands for Body Mass Index in kg/m^2

3.2.3 Primary Efficacy

The primary efficacy endpoint was the unadjusted PI (PI_u). Kaplan-Meier estimate of probability of no pregnancy was also presented as supportive statistics. The sponsor calculated the PI_u and its 95% upper bound using as follows:

- PI_u = (number of pregnancies $\times$ 35625) / (total exposure time in days)

- 95% upper $PI_u = 0.5 \times \chi^2_{0.975, 2(number\ of\ pregnancies+1)}$ / (total exposure days / 35625)

A total of 15 pregnancies were reported in Study 304742. Among the 15 pregnancies, three subjects (512041, 518017, and 602016) were pregnant before taking the study medication, only six subjects (501028, 505008, 509008, 511005, 519007, 602006) were considered pregnant during the study treatment, one subject (604006) was considered to have pregnant within 14 days of end of study treatment, and five subjects (509027, 519008, 604023, 605004, 609012) were considered to be pregnant after 14 days of end of study treatment. As a consequence, the sponsor considered six subjects (501028, 505008, 509008, 511005, 602006, and 604006) to be pregnant within 13 cycles. However, the Division recommended counting the pregnancies within 13 cycles and 7 days (rather than 14 days) of the last treatment. At the end, 5 subjects (501028, 505008, 509008, 511005, and 602006) were considered pregnant in Study 304742.

A total of 30 pregnancies were reported in Study 306660. Among the 30 pregnancies, 5 pregnancies (3274, 3846, 3849, 4046, and 4635) were before the start of study medication; 12 pregnancies (3001, 3481, 3568, 3574, 3675, 3928, 3942, 4010, 4084, 4164, 4455, and 4579) were after end of study medication; 13 pregnancies (3050, 3406, 3553, 3584, 4126, 4579, 3399, 3565, 3712, 3829, 4150, 4462, and 4505) were during the study treatment. In addition, Subject 4579 was pregnant once during the study and once during the follow-up period. Among the 13 pregnancies during the study treatment, Subject 3399 was pregnant around Cycle 18; Subject 3565 was pregnant around Cycle 20; Subject 4126 was older than 35; Subject 4462 was pregnant around Cycle 15. Among the 12 pregnancies after end of study treatment, Subject 3574 was pregnant within 14 days after end of study treatment. The Division considered Subject 3574 was pregnant after 7 days of the end of study treatment. Therefore, a total of 9 subjects of age 18-35 (3050, 3406, 3553, 3584, 4579, 3712, 3829, 4150, and 4505) were considered to be pregnant within 13 cycles and contributed to PI calculation.

The exposure start time and end time were the first date and the last date of bleeding diary, respectively. The backup days included all the cycles (28 day per cycle) in which subjects used condom or other hormone medications (recorded in PD dataset). These hormone medications were Petibelle, Lamuna, Levonorgestrel, Yasmin, Progestogel, NuvaRing, Estriol-Salbe, IUD (INTRAUTERINE CONTRACEPTIVES), Desmin, Loette, Meliane, Microgynon, Minisiston, Valette, Belara, Cilest, Depocon, Diane mite, Leios, Mirena, Monostep, Mercilon, Estriol-Ovula a 0.5mg Estriol, and Darstin. The sponsor's results were summarized in the row having Analysis = Sponsor in Table 3.2.3.1.a.

We found three subjects (520005, 520015 and 608006) had vasectomized partners. All four cycles of Subject 520005, all 13 cycles of Subject 520015 and 3 cycles of Subject 608006 were removed from the PI calculation. As shown in Table 3.2.3.1.a, although more backup cycles were excluded from the PI calculation, the total backup days were relatively small compared to the total exposure days. Therefore, the estimated PI was robust to these minor protocol deviations.

The PI calculated using woman cycle was summarized in Table 3.2.3.1.b.

Table 3.2.3.1.a Summary of Pearl Index Time Unit is Day							
Study	Analysis	Total time of Exposure [days]	Days with back-up contraception	Relevant exposure time [days]	Number of pregnancies	PI_u	Upper 95% CI
304742 (US/Canada)	Sponsor[1]	124995	16321	108674	6	2.0166	4.3892
	FDA[1]	124995	16797	108198	6	2.0255	4.4086
	FDA[2]	124995	16797	108198	5	1.6879	3.9389
306660 (Europe)	Sponsor[1]	323305	8278	315027	10	1.1594	2.1322
	FDA[1]	323305	8278	315027	10	1.1594	2.1322
	FDA[2]	323305	8278	315027	9	1.0435	1.9809
Source: Reviewer's analysis and Table 4 in sponsor's ISE additional tables addressing FDA requests							
1: Pregnancies are counted based on first 13 cycles and within 14 days after the last treatment.							
2: Pregnancies are counted based on first 13 cycles and within 7 days after the last treatment.							

Table 3.2.3.1.b Summary of Pearl Index Time Unit is Woman Cycle							
Study	Analysis	Total time of Exposure [cycle]	Cycles with back-up contraception	Relevant exposure time [cycle]	Number of pregnancies	PI_u	Upper 95% CI
304742 (US/Canada)	FDA	4575	606	3969	5[a]	1.6377	3.8218
306660 (Europe)	FDA	11576	302	11274	9[a]	1.0378	1.9700

Source: Reviewer's analysis
a: Pregnancies are counted based on first 13 cycles and within 7 days after the last treatment.

The Kaplan-Meier estimates were also robust under the above protocol deviation. As shown in Table 3.2.3.2, the pregnant rates were similar between the sponsor and FDA analysis.

Table 3.2.3.2 Kaplan-Meier Estimate of Pregnancy Rate During First 13 Cycles					
Study	Relevant exposure days	Sponsor's analysis[1]		FDA analysis[2]	
		Probability of no conception	Pregnant Rate (95% CI)	Probability of no conception	Pregnant Rate (95% CI)
304742 (US/Canada)	75	0.9974	0.0026 (0.0004, 0.0186)	0.9973	0.0027 (0.0004, 0.0187)
	88	0.9946	0.0054 (0.0014, 0.0215)	--	--
	102	0.9918	0.0082 (0.0027, 0.0253)	0.9945	0.0055 (0.0014, 0.0218)
	186	0.9851	0.0149 (0.0062, 0.0356)	0.9878	0.0122 (0.0046, 0.0323)
	203	0.9816	0.0184 (0.0083, 0.0406	0.9843	0.0157 (0.0065, 0.0375)
306660 (Europe)	13	0.9990	0.0010 (0.0001, 0.0071)	0.9990	0.0010 (0.0001, 0.0071)
	14	0.9980	0.0020 (0.0005, 0.0081)	0.9980	0.0020 (0.0005, 0.0081)
	69	0.9969	0.0031 (0.0010, 0.0095)	0.9969	0.0031 (0.0010, 0.0095)
	95	0.9959	0.0041 (0.0016, 0.0110)	0.9959	0.0041 (0.0016, 0.0110)
	149	0.9947	0.0053 (0.0022, 0.0126)	0.9947	0.0053 (0.0022, 0.0126)
	201	0.9936	0.0064 (0.0029, 0.0142)	0.9936	0.0064 (0.0029, 0.0142)
	207	0.9924	0.0076 (0.0036, 0.0158)	0.9924	0.0076 (0.0036, 0.0158)
	208	0.9913	0.0087 (0.0044, 0.0174)	0.9913	0.0087 (0.0044, 0.0174)
	280	0.9901	0.0099 (0.0052, 0.0190)	0.9901	0.0099 (0.0052, 0.0190)
	336	0.9887	0.0113 (0.0061, 0.0209)	--	--

Source: Reviewer's analysis
1: Pregnancies are counted based on first 13 cycles and within 14 days after the last treatment.
2: Pregnancies are counted based on first 13 cycles and within 7 days after the last treatment.

Pearl indexes based on exposure time in days, the estimated PI is 1.7 and 1.0 in North America and Europe, respectively. The estimates based on exposure time in cycles were 1.6 and 1.0 in the North America and the Europe, respectively with the upper bound of the 95% confidence intervals below 4.0. The cumulative pregnancy rates estimated by Kaplan-Meier method were also consistently similar to Pearl Indexes in both studies.

Despite minor protocol deviation and errors in database the Pearl Index was consistently similar whether exposure was counted in cycles or days. Since the PI is considered a parameter from a Poisson distribution, we strongly recommend the exposure time should be menstrual cycle even though there is a mathematical relationship between cycle and day (1 cycle = 28 days).

3.2.4   Reviewer's Comments on the Efficacy Results

The efficacy results from the North American (US/Canada Study 304742) study was supportive of the results from the European study, despite less than half (approximately 4600 women cycles) the exposure cycles that are generally required for a study to be considered pivotal. The European study (P306660) appeared to be adequate to assess the Pearl Index with more than 10,000 cycles of women exposure. The demographic characteristics were different between the two study populations. Therefore, our conclusions are based on the results from each individual study rather than the integrated summary of efficacy. The Pearl Index for the US and Canadian population is larger compared to the Pearl Index for the European population, however, the upper bounds of the 95% confidence intervals in both studies were <4.0.

3.3 Overview of Study 308960 and Study 308961 for DUB

3.3.1 Design and objectives

Design: The study design and objectives of both DUB studies were identical and conducted under separate protocol. Both studies were multi-center, double-blind, randomized, parallel-group, and placebo-controlled. The sponsor planned to enroll 180 healthy women of 18 years of age or older, who were not pregnant or nursing with a diagnosis of dysfunctional uterine bleeding (DUB) in both Studies. Study 308960 enrolled patients across 42 centers in US (33) and Canada (9), while Study 308961 enrolled patients across 33 European centers in Australia (2), Czech Republic (2), Finland (3), Germany (5), Hungary (3), The Netherlands (3), Poland (5), Sweden (3), UK (3), and Ukraine (4). The primary objective of both studies was to determine the efficacy and safety of EV/DNG in the treatment of DUB, defined as prolonged, frequent, or excessive uterine bleeding.

Both studies had duration of about 344 days including a screening (up to 28 days), a 90-day run-in phase, 196 days of treatment, and a 30-day follow-up period. Each subject was planned to receive either EV/DNG or matching placebo. The randomization ratio of EV/DNG vs. Placebo is 2:1.

Treatment compliance was defined as the total number of pills taken divided by the study duration (in days). The study duration was defined as the time of the day of first drug administration as recorded in the diary until the day of last drug administration. In case of missing diary information, the information was supplemented by appropriate CRF data to calculate compliance.

Primary Efficacy Endpoints: In both studies, the primary efficacy variable was the overall success rate, defined as the number of patients with the absence of any DUB symptoms and who had met all the relevant inclusion criteria for success during the 90-day efficacy assessment phase, divided by the number of patients having at least one qualifying DUB symptom during the run-in phase.

The DUB was defined as experiencing at least one of following symptoms within 90-day run-in phase
- Prolonged bleeding: 2 or more bleeding episodes, each lasting 8 or more days
- Frequent bleeding: greater than 5 bleeding episodes, with a overall minimum of 20 bleeding days
- Excessive bleeding: 2 or more bleeding episodes each with blood loss volume of 80 mL or more, as assessed by the alkaline hematin method

The absence of DUB symptoms was defined as:
- No bleeding episodes lasting more than 7 days, and
- No more than 4 bleeding episodes, and
- No bleeding episodes with blood loss volume of 80 mL or more, in addition

- No more than 1 bleeding episodes increase from baseline, and
- Total number of bleeding days not to exceed 24 days, and
- No increase from baseline in an individual patient's total number of bleeding days, in addition
- If patients enrolled with prolonged bleeding, the decrease between the maximum duration during run-in phase and the maximum duration during the efficacy phase should be at least 2 days
- If patients enrolled with excessive bleeding: (1) the blood loss volume associated with each episode be < 80 mL and (2) the blood loss volume associated with each bleeding episode represents a decrease of at least 50% from the average of the qualifying bleeding episodes, where the qualifying bleeding episodes are those with a blood loss volume ≥ 80 mL that occurred during the run-in phase

The bleeding episode was defined as follows:
- Bleeding for at least 2 days
- Bleeding days can be separated by no more than 1 bleeding-free day
- An episode stops with 2 consecutive bleeding-free days

Any randomized patient, who did not complete at least 90 days of treatment from the beginning or did not have sufficient data to evaluate the absence or presence of DUB symptoms, was considered a treatment failure.

Secondary Efficacy Endpoints: Secondary efficacy were**:**
- Proportion of patients cured from each individual symptom
- Proportion of patients with improvement in the patient's overall assessment scale on days 84 and 196
- Proportion of patients with improvement in the investigator's global assessment scale on days 84 and 196
- Change in blood loss volume for all patients and for patients with excessive bleeding
- Changes in number of bleeding days and bleeding episodes
- Changes in number of sanitary protection used
- Change from baseline in QoL scores on days 84 and 196
- Resource use assessment at baseline, day 84, and day 196
- Change from baseline in serum Ferritin concentrations and hemoglobin on days 84 and 196

Determination of Sample Size: Assuming overall success rate of 50% and 20%, in the EV/DNG group and the placebo group respectively, and a drop-out rate of 30%, a total of 120 patients in the EV/DNG group and 60 patients in the placebo group was estimated to provide a power of 90% to test the null hypothesis that the proportions of success in the 2 treatment groups were equal at an alpha of 0.05.

Definition of Analysis Sets (Population): Three analysis sets were defined: (1) intention-to-treat (ITT) population set, (2) Per protocol (PP) set, and (3) Safety analysis set. The ITT set included all randomized patients. The PP set included all randomized patients who met all the

inclusion/exclusion criteria, did not take any prohibited medication, had at least 75% overall study drug compliance, had no major protocol violations, and completed 7 treatment cycles. The safety set included all randomized patients who took at least one pill of study medication.

Handling of Missing Data: For the 90-day efficacy phase, it shifted backward. The starting of the 90-day period matched the first day of a treatment cycle so that the last 90 days with complete data were captured. For the missing bleeding intensity data, they were replaced with the highest bleeding intensity of the bordering days. The amount of replacement was up to 9 non-consecutive days or <10% of actual available data per phage. For missing blood loss volume data in run-in period, it was replaced by "0". In treatment period, it was replaced according to the following table.

Type	Bleeding Intensity on the respective day:	Sanitary protection completely collected on the respective day:	MBLV Data
Type 1	'none', 'spotting', ≥ 'light', or missing	'yes', 'no' or missing	> 0
Type 2	≥ light	'yes', 'no' or missing	missing / 0 (zero)
Type 3	'none', 'spotting', 'missing'	'yes', 'no' or missing	missing

Pooling of Sites: Small centers (<5 PP patients with primary efficacy post-baseline data) were pooled. All small centers were ranked from the largest to the smallest with respect to the total number of PP set patients having primary efficacy post-baseline data. These centers were pooled from the largest to the smallest until the pooled center had ≥ 5 PP set patients having post-baseline data for the primary efficacy variables at the efficacy phase in each treatment group. Any leftover centers that did not have a sufficient number of patients to form a pooled center were pooled with the last pooled center. For centers with no per protocol patients, they were ordered by center number and assigned sequentially to the pooled centers formed.

Statistical Methods: The analysis of the primary efficacy variable was based on ITT population, as well as the PP set. However, the analysis of the secondary variable was based on ITT population only.

The overall success rate (primary efficacy variable) was analyzed by the difference of proportions between EV/DNG and placebo groups. In order to claim efficacy of EV/DNG, the proportion of successful responders in the EV/DNG group must be statistically significantly greater than that in the placebo group, and the point estimate for the proportion of successful responders in the EV/DNG group must be at least 50%. The proportion of success between EV/DNG and placebo at efficacy phase would also be compared with a logistic regression model to analyze the effects of pooled center and treatment-by-center interaction.

For secondary analysis, categorical responses were analyzed by differences of proportions and their 95% CI. All continuous responses (all changes) and the laboratory value Hematocrit were analyzed using an ANCOVA model to compare the treatment groups. The model included fixed

effects of treatment, pooled center, and baseline as a covariate. However, if the assumption of homogeneity of slopes in the ANCOVA model did not hold at significance level of 0.05, the covariate was dropped from the model. The normality of the residual was assessed with the Shapiro-Wilk test. If residual analysis did not support the normality assumption at significant level of 0.05, rank ANCOVA was performed with the same terms used in the parametric model.

Multiple Comparisons/Multiplicity: There was no issue of multiplicity for the primary efficacy variable (overall success rate).

3.3.2 Reviewer's Comments on the Design

The design for both studies appeared to be adequate.

3.4 Results: Studies 308960 and 308961

3.4.4 Subject Disposition

The study completion rate in US/Canada population was 71.1% compared to 78.8% in European population. No single site was predominant in terms of subject enrollment. The study discontinuation rate was also higher in the US/Canada (26.3%) compared to the Europe (19.5%). The major reasons for discontinuation were the adverse event (8.4% vs. 7.8%), and withdrawal (7.9% vs. 6.1%) in the US/Canada population and Europe, respectively. More AE and withdrawal were observed in the EV/DNG group. The ITT population sizes in both studies were larger than the planned sample sizes.

Table 3.4.1. Disposition of Subjects: Studies 308960 and 308961						
	Study 308960			**Study 308961**		
Category	**EV/DNG** **N (%)**	**Placebo** **N (%)**	**Total** **N (%)**	**EV/DNG** **N (%)**	**Placebo** **N (%)**	**Total** **N (%)**
Total Randomized	120 (100%)	70 (100%)	190 (100%)	149 (100%)	82 (100%)	231 (100%)
Completed Study	84 (70.0%)	51 (72.9%)	135 (71.1%)	117 (78.5%)	65 (79.3%)	182 (78.8%)
Medication never administered	1 (0.8%)	4 (5.7%)	5 (2.6%)	3 (2.0%)	1 (1.2%)	4 (1.7%)
Discontinued Study	35 (29.2%)	15 (21.4%)	50 (26.3%)	29 (19.5)	16 (19.5%)	45 (19.5%)
Adverse event	12 (10%)	4 (5.7%)	16 (8.4%)	13 (8.7%)	5 (6.1%)	18 (7.8%)
Consent withdrawn	11 (9.2%)	4 (5.7%)	15 (7.9%)	10 (6.7%)	4 (4.9%)	14 (6.1%)
Lack of Efficacy	2 (1.7%)	0 (0.0%)	2 (1.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lost to follow-up/missing	5 (4.2%)	4 (5.7%)	9 (4.7%)	0 (0.0%)	1(1.2%)	1 (0.4%)
Pregnancy	0 (0.0%)	1 (1.4%)	1 (0.5%)	0 (0.0%)	1 (1.2%)	1 (0.4%)
Protocol deviation	2 (1.7%)	0 (0.0%)	2 (1.1%)	3 (2.0%)	3 (3.7%)	6 (2.6%)
Others	3 (2.5%)	2 (2.9%)	5 (2.6%)	3 (2.0%)	2 (2.4%)	5 (2.2%)
ITT Analysis Set	120	70	190	149	82	231
Per Protocol Set	41	27	68	55	34	89
Safety Set	119	66	185	145	81	226
Source: Text Figure 2 in Study 308960, Text Figure 5 in Study 308961 and reviewer analysis						

3.4.4   Patient Demographics and Baseline Characteristics

The baseline characteristics such as age, race, body mass index (BMI) were similar in the two treatment groups within each study. The US/Canada population had much less Caucasian (61.6%) then that (97.0%) in European. Smoking history and Alcohol consumption were also similar except for the non-alcohol user in Study 308960 (25% non-user in EV/DNG vs. 11.4% non-user in placebo). The difference could be due to randomness.

Table 3.4.2. Subject Demographic and Baseline Characteristics (FAS)						
Variable/Parameters	**Study 308960**			**Study 308961**		
	EV/DNG N=120	**Placebo N=70**	**Total N=190**	**EV/DNG N=149**	**Placebo N=82**	**Total N=231**
Age (SD, Min-Max)	36.9 (7.45, 20-53)	37.0 (6.67, 21-49)	36.9 (7.15, 20-53)	39.5 (6.57, 18-51)	38.5 (7.52, 19-54)	39.2 (6.92, 18-54)
BMI* (SD)	26.3 (3.56)	25.8 (3.60)	26.1 (3.57)	24.5 (3.49)	25.7 (3.01)	24.9 (3.37)
Ethnic group — Caucasian	71 (59.2%)	46 (65.7%)	117 (61.6%)	144 (96.6%	80 (97.6%)	224 (97.0%)
Ethnic group — Black	38 (31.7%)	14 (20.0%)	52 (27.4%)	1 (0.7%)	-	1 (0.4%))
Ethnic group — Hispanic	8 (6.7%)	6 (8.6%)	14 (7.4%)	-	-	-
Ethnic group — Asian	1 (0.8%)	2 (2.9%)	3 (1.6%)	2 (1.3%)	1 (1.2%)	3 (1.3%)
Ethnic group — Other	2 (1.7%)	2 (2.9%)	4 (2.1%)	2 (1.3%)	1 (1.2%)	3 (1.3%)
Smoking history — Yes	37 (30.8%)	24 (34.3%)	61 (32.1%)	42 (28.2%)	31 (37.8%)	73 (31.6%)
Smoking history — No	83 (69.2%)	46 (65.7%)	129 (67.9%)	107 (71.8%)	51 (62.2%)	158 (68.4%)
Alcohol consumption — Never	30 (25.0%)	8 (11.4%)	38 (20.0%)	25 (16.8%)	15 (18.3%)	40 (17.3%)
Alcohol consumption — Seldom	38 (31.7%)	26 (37.1%)	64 (33.7%)	62 (41.6%)	32 (39.0%)	94 (40.7%)
Alcohol consumption — Occasionally	44 (36.7%)	30 (42.9%)	74 (38.9%)	47 (31.5%)	24 (29.3%)	71 (30.7%)
Alcohol consumption — Regularly	8 (6.7%)	6 (8.6%)	14 (7.4%)	15 (10.1%)	11 (13.4%)	26 (11.3%)
Alcohol consumption — Missing	-	-	-	-	-	-
Source: Reviewer's analysis, Table 7 in Study 308960, and Text Table 9 and 15 in Study 308961 *: BMI stands for Body Mass Index in kg/m^2						

3.4.3   Primary Efficacy

The primary efficacy variable was the overall success rate. The numerator of this overall success rate was the number of patients with the absence of any DUB symptom and who had met all the relevant criteria for success during the 90-day efficacy assessment phase. The associated denominator was the number of patients having at least one qualifying DUB symptom during the run-in phase.

For randomized patients who did not complete at least 90 days of treatment from the beginning, or did not have sufficient data to evaluate the absence or presence of DUB symptoms, they were considered a treatment failure.

Treatment compliance was defined as the total number of pills taken divided by the study duration in days, which was obtained from the patients' diaries. The study duration was defined as the time in days between the first and the last study drug administration.

In order to claim efficacy of EV/DNG, the overall success rate must (1) be statistically significant higher in the EV/DNG group comparing to placebo group, and (2) have point estimate of ≥50% in the EV/DNG group.

(b) (4)

3.5   Evaluation of Safety

The safety review is referring to the medical officer's review.

4.   FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

There was no subgroup analysis performed for the OC indication. For DUB indication, subgroup analysis was part of the secondary analysis.

(b) (4)

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

No major statistical issues were noted in this submission. In the two OC studies, exposure was based on both cycle and days and the results were similar. Three subjects had vasectomized partners in the US/Canada study. However, these errors in database did not significantly change the PI estimates. Our conclusions are based on the results of each study separately, rather than based on pooled studies, as presented by the sponsor.

5.2 Conclusions and Recommendations

The data from the two OC studies demonstrated contraceptive efficacy of EV/DNG, despite that the North American study provided a wider 95% confidence interval of the PI. The Pearl Index calculated using exposure time in days was 1.7 and 1.0 in the North American and the European study, respectively. The upper bound of the 95% confidence interval was less than 4.0. The cumulative pregnancy rates by Kaplan-Meier method were also similar in both studies.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
--------------------	--------------------	--------------------	--
NDA-22252	ORIG-1	BAYER HEALTHCARE PHARMACEUTICALS INC	Qlaira

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
03/29/2010

MAHBOOB SOBHAN
03/31/2010