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

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	21-945
Drug Name:	Makena (formerly Gestiva; 17 α-hydroxyprogesterone; caproate injection)
Indication(s):	Reduction of risk of preterm birth in pregnant women with a history of at least one spontaneous preterm birth
Applicant:	Hologic Corporation
Date(s):	Date stamp: 7/13/2010
	PDUFA date: 1/13/2011
	3-month extension: 4/13/2011
Review Priority:	6-month – Complete Response to Approvable Letter
Biometrics Division:	DB 3
Statistical Reviewer:	Lisa A. Kammerman, Ph.D.
Concurring Reviewers:	Stephen E. Wilson, Dr.P.H.
Medical Division:	Reproductive and Urological Products
Clinical Team:	Barbara Wesley, MD; Lisa Soule, MD
Project Manager:	Charlene Williamson

Keywords: clinical studies, NDA review, standard of evidence.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 5

2. INTRODUCTION 8

2.1 Overview of Complete Response 8

2.2 Data Sources 8

3. STATISTICAL EVALUATION 9

3.1 Approval under Subpart H, 21 CFR 314.510 using a single study 9

3.2 Evaluation of efficacy 10
3.2.1 Time-to-delivery by race 10
3.2.2 Deliveries <37 weeks gestation 18
3.2.3 Late pre-term deliveries 22

3.3 Evaluation of Safety 24

3.4 Comments on labeling and promotional materials 29
3.4.1 Preterm deliveries 29
3.4.2 Neonatal mortality/morbidity index 31
3.4.3 Racial subgroups 32
3.4.4 Gestational age at randomization 32

3.5 Comments on ongoing confirmatory study, 17P-ES-003 32

3.6 Draft protocol for a follow-up study (Study 17P-FU-004) of children born to mothers who received 17P or placebo in the ongoing confirmatory study required for Subpart H approval 33
3.6.1 Summary of protocol 33
3.6.2 Statistical comments 35

4. CONCLUSIONS AND RECOMMENDATIONS 37

TABLES

Table 1. Racial distribution, by center. Entries are percentages of entire study enrollment 14
Table 2. Distribution of blacks, by center. Entries are percentages of each center's enrollment. 15
Table 3. Distribution of blacks by gestational age at randomization 17
Table 4. Applicant's Analysis: Delivery $<37^0$ Weeks Gestation 19
Table 5. Delivery $<37^0$ Weeks, $<35^0$ Weeks, $<32^0$ Weeks, $<28^0$ Weeks Gestation, ITT population. The estimates of rates of delivery do not account for duration of drug exposure. 20
Table 6. Proportion of subjects who delivered $<37^0$ Weeks, by gestational age at randomization 21
Table 7. Proportion of subjects who delivered $<37^0$ Weeks, by gestational age at randomization and race 22
Table 8. Percentage (number) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group. 23
Table 9. Percentage (# delivered) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group and race 23
Table 10. Percentage (# delivered) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group and gestational age at randomization. 24
Table 11 Distribution of Miscarriages, Stillbirths and Neonatal Deaths, by Gestational Age at Randomization 24
Table 12 Distribution of Fetal and Neonatal Deaths, by Center and Gestational Age at Randomization 25
Table 13. Crude rates of fetal and neonatal losses, by race. Rates are not adjusted for duration of exposure to study treatment 25
Table 14. Estimated Rates of Fetal Losses (miscarriages and stillbirths) and Neonatal Deaths, accounting for time on study drug, by race. 26
Table 15. Listing of Miscarriages, Stillbirths and Neonatal Deats, sorted by Treatment, Race and Gestational age at delivery 31

FIGURES

Figure 1. Time-to-delivery as a function of gestational age 11
Figure 2. Time-to-delivery as a function of gestational age, by race. 12
Figure 3. Time-to-delivery from date of randomization, by race 13
Figure 4. Time-to-delivery from date of randomization, by gestational age at randomization 16
Figure 5. Time to delivery from date of randomization, by race and gestational age at the time of randomization. 18
Figure 6. From time of randomization to fetal and neonatal deaths, by treatment group.27
Figure 7. From time of randomization to fetal and neonatal deaths, by treatment group and race. 28
Figure 8. From time of randomization to fetal and neonatal deaths, by treatment group and gestational age at the time of randomization..... 29

ATTACHMENTS

Attachment 1. Statistical Review of NDA 21-945, Complete Response, dated 1/23/2009
Attachment 2. Statistical Review of NDA 21-945, dated 10/19/2006

1. EXECUTIVE SUMMARY

From a statistical perspective, the information and data submitted by the Applicant do not provide convincing evidence regarding the effectiveness of 17 α-hydroxyprogesterone, caproate injection (17P) for the prevention of preterm deliveries among women with a history of at least one spontaneous preterm delivery.

The Applicant is seeking approval based on the results from only one adequate and well-controlled study, which has been submitted for review. The study, submitted with the original NDA, had several features that do not allow the study to stand on its own to establish the efficacy of 17P on the surrogate endpoint of preterm deliveries, as described in the guidance document, "Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products."[1]

In my previous review of the study (see Attachment 2), I focused on the endpoints of delivery <35 weeks, delivery <32 weeks and time-to-delivery. My reasons for concluding that a single study was not sufficient to support the effectiveness of 17P in preventing preterm deliveries were:

- Optimal time to start study drug was not identified.
  - 17P appeared most effective when started at 18 weeks of gestation or earlier; did not appear effective when started at 20 weeks of gestation or later.
  - Rate of fetal and neonatal deaths is most pronounced among births to women who started 17P at 18 weeks gestation or earlier (10%).
- Apparent confounding of study site and gestational age at randomization.
  - One center accounted for 44% of subjects enrolled at 18 weeks of gestation or earlier.
  - Some centers had a deficit of subjects enrolled at 18 weeks of gestation or earlier.
- Fetal and neonatal deaths among women treated with 17P occur earlier than among women treated with placebo.
- One center accounted for a relatively large proportion of all subjects enrolled.

However, recognizing an important public health need for the commercialization of this drug product, the medical division is currently recommending approval under Subpart H, based on a statistically significant treatment effect for the surrogate endpoint of deliveries prior to 37 weeks gestational age. This endpoint is a departure from the earlier review cycles that focused on the surrogate endpoints of deliveries prior to 32 weeks and deliveries prior to 35 weeks. My previous reviews did not sufficiently address the results at 37 weeks at the depth required to establish the efficacy of 17P based on a single study.

[1] Available at www fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf

In addition, my reviews did not explore whether the results from these endpoints were consistent among racial subgroups.

In this review of the second Complete Response, I have done additional analyses to address whether the data are sufficient to support approval if the endpoint of deliveries <37 weeks gestation is used as the surrogate endpoint. I have also done additional analyses exploring the effect of race on the efficacy results. However, the results from these analyses do not support the efficacy of 17P based on a single study.

My conclusion that the results from these additional analyses do not support the efficacy of 17P based on a single study are:

- The treatment effect at 37 weeks does not appear to be consistent among groups defined by gestational age at randomization. This finding may be confounded with race and study center.
- Lack of consistency of efficacy results among subgroups defined by race.
    - For subjects who were black, the benefit of 17P compared with Placebo appears to emerge at around 24 weeks.
    - For subjects who were non-blacks, a treatment benefit does not emerge until 35 weeks gestation.
- Lack of consistency of safety results at Week 24 among subgroups defined by race.
    - Among subjects who were black, the estimated rate of fetal and neonatal losses was 6% for subjects, regardless of treatment assignment.
    - Among subjects who were non-black, subjects randomized to Placebo did not have any fetal or neonatal losses compared with an estimated rate of 9% among those randomized to 17P.
- The doubling of the treatment effect from <35 weeks to <37 weeks is likely due to the increased number of deliveries among non-black subjects randomized to Placebo.

These exploratory analyses were necessary because of the reliance on a single study to support the approval of 17P. In some cases, the observed treatment effects may have been based on small numbers of subjects. However, the overall objective was to look at consistency among various endpoints and across various subgroups to determine whether the results could be extrapolated to a larger population in the absence of a second study, and these are the only data we have.

I recommend that the final label (1) include only those data on which approval will be based and (2) describe the limitations of the results. Because the approval of 17P will be based on the surrogate endpoint of deliveries <37 weeks gestation, I recommend that the label include efficacy data for this endpoint only and exclude efficacy information pertaining to deliveries <35 weeks and to deliveries <32 weeks. As noted in my previous reviews, the data for these two endpoints coming from a single study are insufficient to support the efficacy of 17P. Moreover, I recommend adding text indicating 17P has not been shown to be effective in reducing the risk of deliveries at earlier time points. These

recommendations are consistent with the guidance document, "Guidance for Industry: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content and Format".

The guidance document also recommends the inclusion of summary statements about the results of required explorations. As such, I recommend including a statement that examination of racial subgroups suggests a larger treatment effect in African-American women, a higher rate of early losses among women who are not African-American, and the apparent absence of an effect when treatment is started after 20 weeks gestation.

I recommend excluding from the label the results from the neonatal mortality/morbidity index. First, including the index seems to contradict the statement contained in the indications and usage section: "There are no controlled trials demonstrating a direct clinical benefit, such as improvement in neonatal mortality and morbidity."

In addition, the index is limited to live births only and, therefore, counts miscarriages and stillbirths as successes regardless of their timing. For example, a neonatal death at 20.1 weeks gestation was counted a failure while a stillbirth at 20.1 weeks gestation was counted a success. If an index is needed, I recommend replacing the neonatal index with an index that includes all cause mortality, not just deaths occurring among live births. By not including all deaths, the finding reported in the draft labeling overstates the efficacy of 17P.

Finally, including results in labeling with point estimates favoring a drug product even with the disclaimer "not statistically significant" is not advisable. Conceivably, the lower observed rate in 17P versus Placebo could be promoted as a benefit for 17P despite the phrase "not statistically significant".

The ongoing confirmatory study, 17P-ES-003, will be used to confirm the clinical benefit of 17P in order to fulfill the Subpart H requirements. As currently designed the primary endpoint is a surrogate endpoint, deliveries <35 weeks. The use of this surrogate endpoint, instead of clinical endpoint, will not be sufficient for confirming the clinical benefit of 17P.

Any clinical endpoint that will be used for the basis of approval needs to account for all subjects enrolled in the study. The proposed co-primary endpoint, the neonatal composite index, assesses live births only; miscarriages and stillbirths are excluded. An analysis set that excludes subjects based on post-randomization events violates the intention-to-treat principle. The primary analysis for the purpose of approval needs to account for all subjects and their births. An analysis limited to live births could be a secondary analysis.

2. INTRODUCTION

2.1 Overview of Complete Response

This submission is a second response to the Complete Response Letter for NDA 21-945 (dated 10/20/06); the first response was submitted on 4/5/2008. The current submission contains draft labeling and references a draft protocol for a follow-up study of children aged 23 to 25 months, whose mothers received 17P[2] or vehicle in the ongoing confirmatory study required for Subpart H approval. The medical division requested that I review this protocol, which was submitted to IND 68,108 on 6/29/2009.

In my previous reviews of the original submission and the first Complete Response, I concluded the evidence coming from the single study submitted was insufficient to support the effectiveness of 17P (Attachments 1 and 2). In those review cycles, the recommended surrogate endpoints for approval were deliveries prior to 35 weeks gestation and prior to 32 weeks gestation. As I discussed in those reviews, the evidence from the single clinical trial was not sufficient to support the efficacy of 17P based on these endpoints and based on time to delivery.

However, the medical division is now recommending approval under Subpart H, using deliveries prior to 37 weeks of gestation as the primary surrogate endpoint. The medical division now believes that deliveries prior to 37 weeks of gestation is an appropriate surrogate endpoint for Subpart H approval. According to the medical division, research articles published in the medical literature over the last several years support the clinical benefit of delaying so-called "late pre-term" deliveries. Late pre-term deliveries are defined as occurring, approximately, between 35 and 37 weeks gestation.

As a result of the medical division's decision to use deliveries <37 weeks gestation as the basis for approval, I have done additional analyses to explore whether the level of evidence from a single study, using deliveries <37 weeks gestation, is sufficient to support the effectiveness of 17P. This review reports the results of those analyses. I also include the results of analyses that compare the efficacy of 17P among blacks with the efficacy of 17P among non-blacks.

2.2 Data Sources

Applicant's "Response to Information Request from 14 January 2011," dated 1/25/2011.

"Request for Consultation" from Carrie Newcomer, PharmD (DDMAC); dated 1/10/2011

[2] The planned marketed drug product is 17 α-Hydroxyprogesterone; Caproate Injection, 250 mg/mL. This drug product is abbreviated as 17P throughout the NDA, Complete Response and this review.

Proposed Updates to Protocol 17P-ES-003, submitted on 1/7/11 by email from Robb Hesley, Hologic, Inc.

Applicant's "Draft Responses to Information Request," included in email dated 11/18/2010; see Memorandum to File signed by Ms. Williamson and dated 12/22/2010.

Applicant's Complete Response, dated 7/13/2010 (paper submission)

Protocol 17P-FU-004, submission dated 6/29/2009, submitted to IND 68,108

Statistical Review of NDA 21-945, Complete Response, dated 1/23/2009 (Attachment 1)

Approvable Letter for NDA 21-945, dated 10/20/2006

Statistical Review of NDA 21-945, dated 10/19/2006 (Attachment 2)

Transcripts from Reproductive Health Drugs Advisory Committee Meeting held on 8/29/2006 (http://www.fda.gov/ohrms/dockets/ac/cder06.html#rhdac)

3. STATISTICAL EVALUATION

3.1 Approval under Subpart H, 21 CFR 314.510 using a single study

From my perspective, the information submitted to date does not fulfill the requirements for approval under Subpart H, 21 CFR 314.510. Under Subpart H, a drug product may be approved if adequate and well-controlled clinical trials establish the drug product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit.

Only one adequate and well-controlled study has been submitted for review. The study, submitted with the original NDA, had several features that does not allow the study to stand on its own to establish the efficacy of 17P on the surrogate endpoint of preterm deliveries, as described in the guidance document, "Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products."[3]

In my previous review of the study (see Attachment 2), I focused on the endpoints of delivery <35 weeks, delivery <32 weeks and time-to-delivery. My reasons for concluding that a single study was not sufficient to support the effectiveness of 17P in preventing preterm deliveries are summarized in the Executive Summary of this review of the second Complete Response.

In this review of the second Complete Response, I have done additional analyses to address whether the data are sufficient to support approval if the endpoint of deliveries

[3] Available at www fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf

<37 weeks gestation is used as the surrogate endpoint. I have also done additional analyses exploring the effect of race on the efficacy results. My conclusion that the results from these additional analyses do not support the effectiveness of 17P based on a single study are also summarized in the Executive Summary.

In addition to the issues surrounding the level of evidence provided by a single study, the use of Subpart H as a pathway for approval does not seem appropriate for 17P. Unlike studies of HIV and cancer where the difference in time between the outcome of a surrogate endpoint and a clinical endpoint can be years, in this situation, the time between the clinical outcome of interest (i.e., mortality and neonatal morbidity) and the surrogate outcome (<37 weeks) is literally weeks. The fact that a confirmatory study is currently ongoing does not translate into a lesser standard of evidence needed to conclude efficacy based on the evaluation of an endpoint from a single study. The data from the single study submitted for approval, for the reasons summarized above, is insufficient to support the efficacy of 17P.

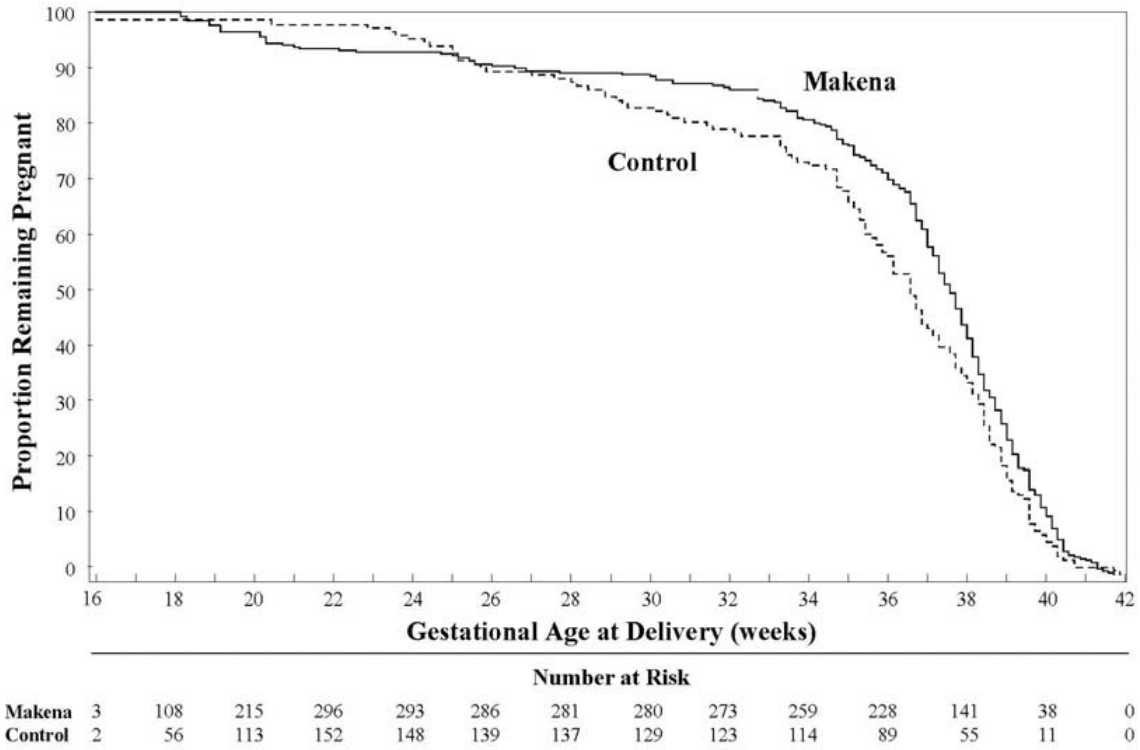
3.2 Evaluation of efficacy

3.2.1 Time-to-delivery by race

This section of my review supplements the time-to-delivery results presented in my original review of Study 17P-CT002 by exploring the consistency of the effect of treatment on time-to-delivery among subgroups defined by race.

For the entire study population, the Kaplan-Meier curves for time-to-delivery as a function of gestational age at the time of randomization are shown in Figure 1. These curves account for staggered entry into the study. Of interest is the crossing of the curves. The first birth that was not classified as a fetal loss or neonatal death occurred shortly after was 24 weeks gestation. Deliveries prior to 24 weeks were miscarriages, stillbirths or neonatal deaths. As can be seen, women randomized to 17P had a higher rate of losses than did women randomized to placebo.

Figure 1. Time-to-delivery as a function of gestational age



Source: Draft labeling

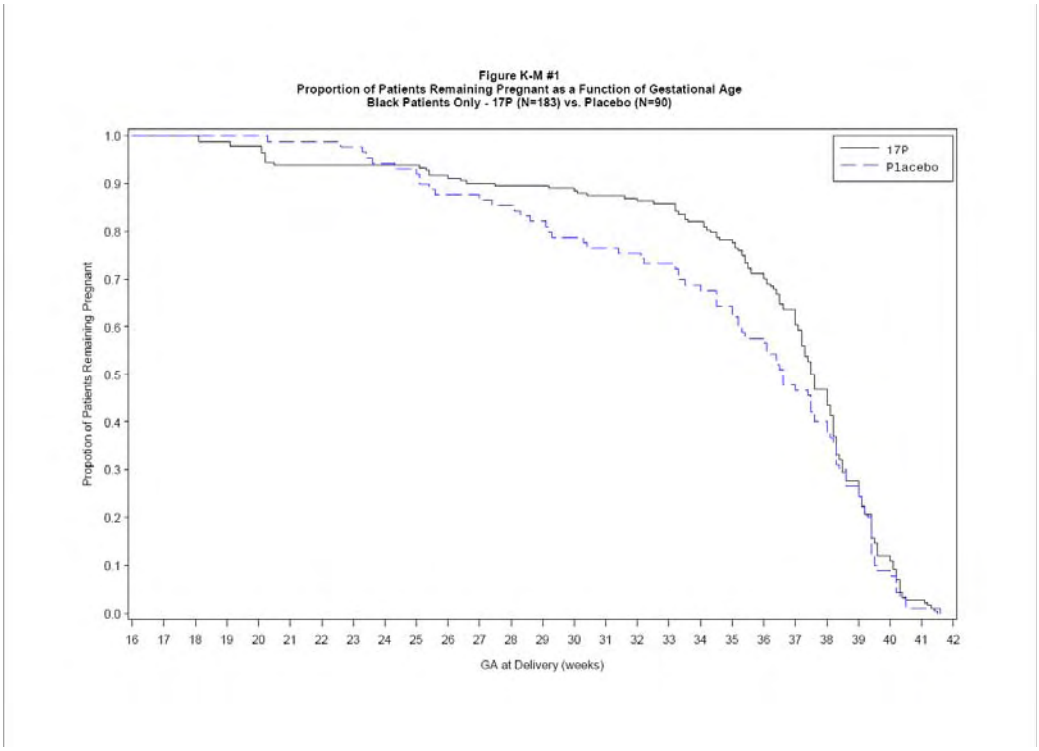
When time-to-delivery is examined by racial subgroups, differences between the subgroups in the shapes of the time-to-delivery curves are apparent, pointing up potential differences in treatment effects for subjects who were black as compared with those who were not black (Figure 2 and Figure 3).

Although both racial subgroups mimic the pattern seen in the overall population, the differences between subjects who were black and non-black is notable. Overall, subjects who were black appeared to receive a greater benefit from 17P than did subjects who were non-black. Initially, in both subgroups, subjects randomized to 17P experienced higher rates of early deliveries than those randomized to Placebo. Among subjects who were black, the treatment difference began favoring 17P at approximately Week 24. By contrast, the higher rates of deliveries among those randomized to 17P persisted until approximately Week 33; after Week 35, the rate of deliveries favored 17P.

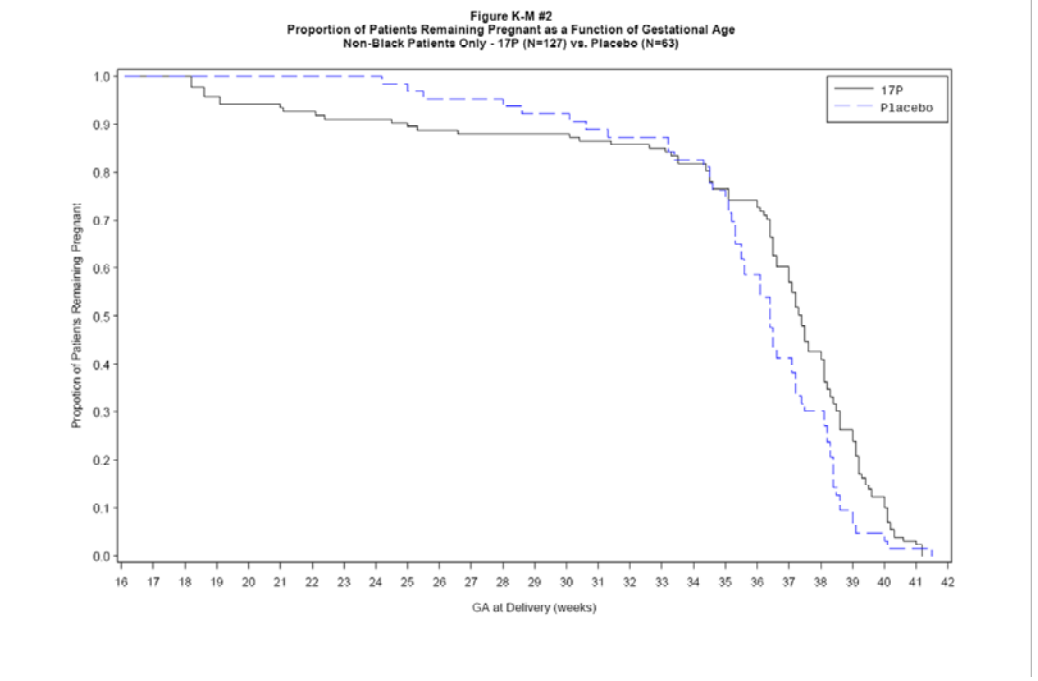
These patterns are also seen when examining time-to-delivery using date of randomization as the baseline. The crossing point occurs earlier among subjects who were black than among those who were non-black (Figure 3).

Figure 2. Time-to-delivery as a function of gestational age, by race.

Black

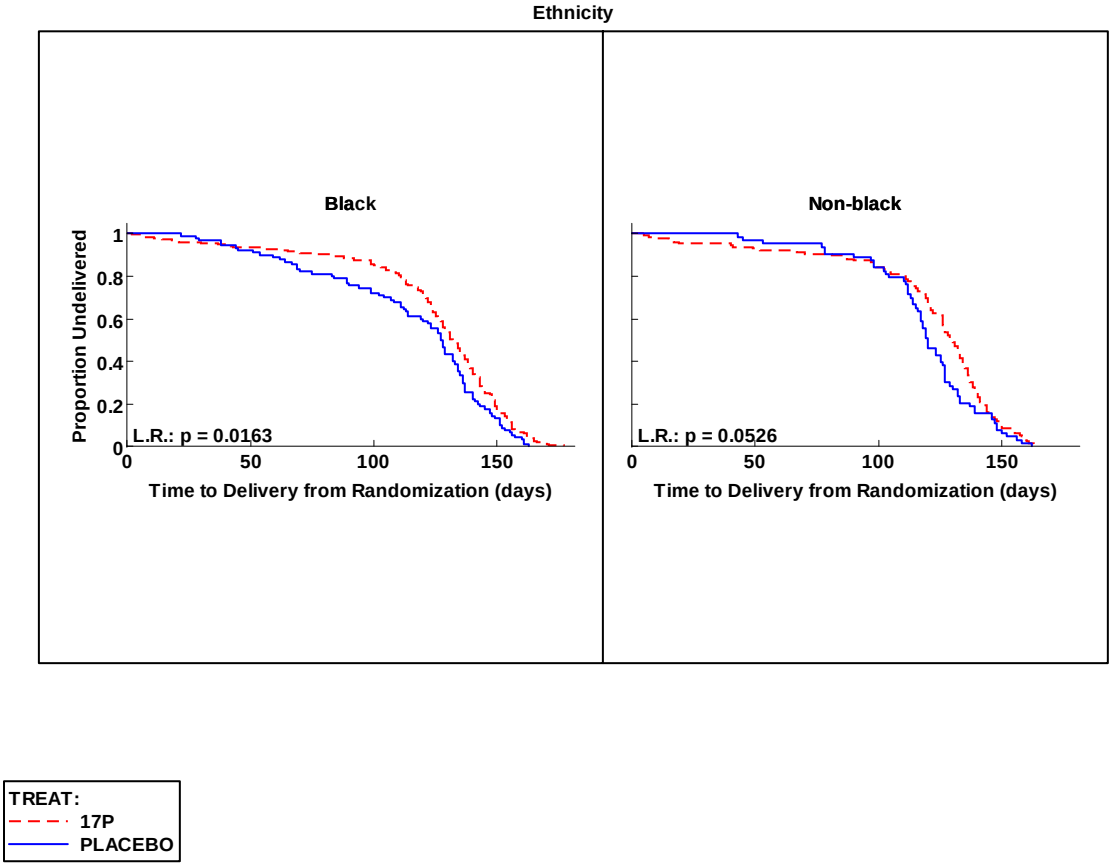


Non-black



Source: Applicant's "Response to Information Request from 14 January 2011," submission dated 1/25/2011.

Figure 3. Time-to-delivery from date of randomization, by race



In the following paragraphs, I attempt to explore possible reasons for these differences between the racial subgroups.

Potentially, some of the differences may be confounded with study center. While almost 60% of subjects enrolled in the study were black, the enrollment of black subjects was not uniform across study sites (Table 1). The largest center, University of Alabama (Center 8), accounted for 27% of all subjects enrolled in the study but represented 43% of all subjects who were black. Almost all subjects (93%) enrolled at the University of Alabama were black (Table 2). In addition, almost a quarter of all subjects who were non-black were enrolled at Center 20 (University of Utah), which accounted for only 9% of all subjects enrolled in the study.

Table 1. Racial distribution, by center. Entries are percentages of entire study enrollment

Center #	Number of subjects enrolled at center	% of all subjects enrolled in study (N=463	% of all subjects who were black (N=273)	% of all subjects who were non-black (N=190)
2	36	7.8	7.0	8.9
4	45	9.7	15.4	1.6
8	126	27.2	42.9	4.7
9	24	5.2	8.4	0.5
11	13	2.8	3.7	1.6
13	22	4.8	1.8	8.9
14	7	1.5	2.2	0.5
15	28	6.0	5.9	6.3
17	11	2.4	2.6	2.1
18	39	8.4	2.6	16.8
19	13	2.8	0.0	6.8
20	43	9.3	0.0	22.6
21	24	5.2	4.8	5.8
22	5	1.1	0.0	2.6
23	11	2.4	0.4	5.3
25	6	1.3	1.1	1.6
26	4	0.9	0.7	1.1
27	4	0.9	0.4	1.6
28	2	0.4	0.4	0.5
All subjects	*463*	*100%*	*59%*	*41%*

Source: Statistical Reviewer

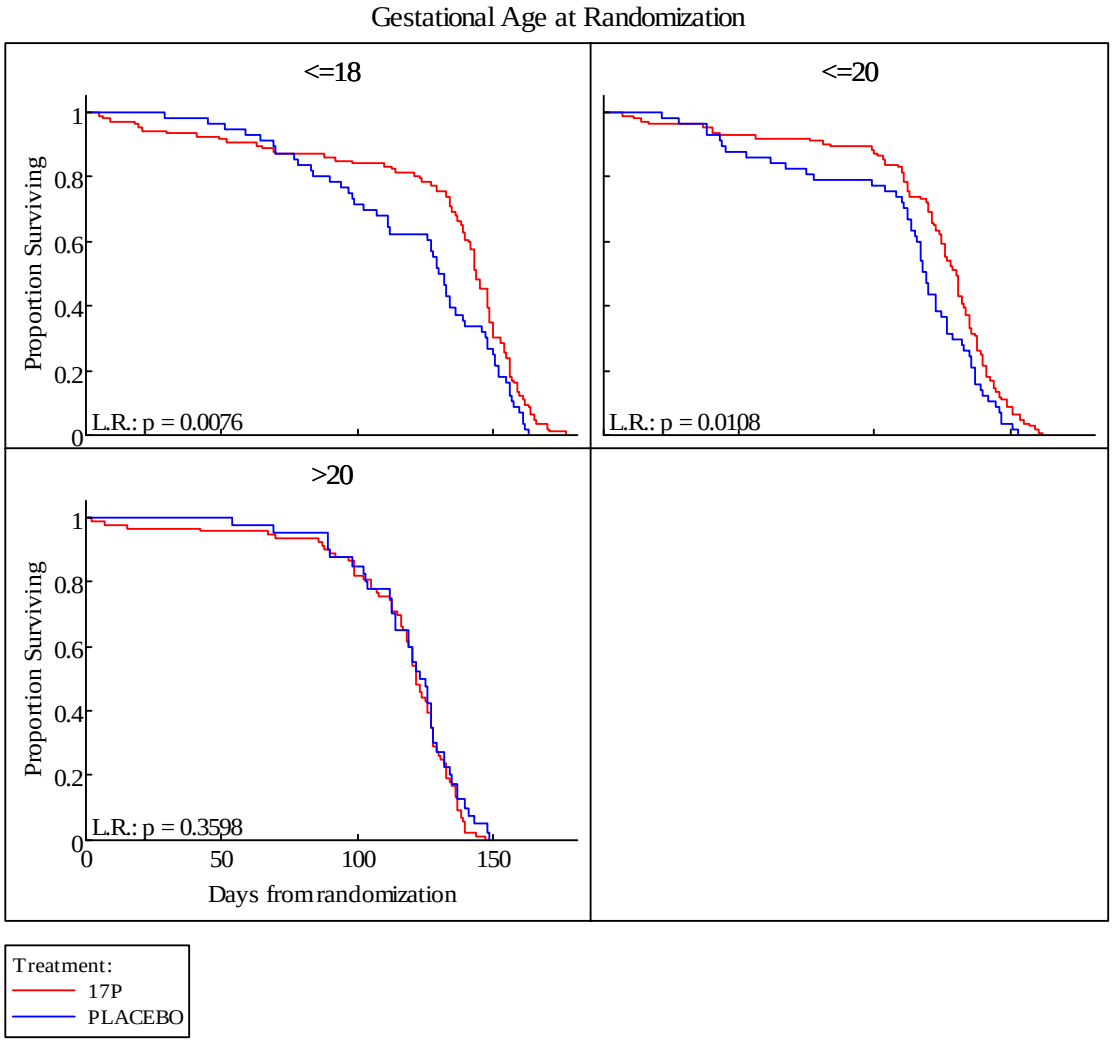
Table 2. Distribution of blacks, by center. Entries are percentages of each center's enrollment.

Center #	Number of subjects enrolled at center	% of subjects at center who were black
2	36	52.8
4	45	93.3
8	126	92.9
9	24	95.8
11	13	76.9
13	22	22.7
14	7	85.7
15	28	57.1
17	11	63.6
18	39	17.9
19	13	0.0
20	43	0.0
21	24	54.2
22	5	0.0
23	11	9.1
25	6	50.0
26	4	50.0
27	4	25.0
28	2	50.0
All subjects	*463*	*59%*

Source: Statistical Reviewer

In addition, the results may be confounded with gestational age at the time of randomization. My review of Study 17P-CT002 noted that time-to-delivery appeared to depend on when study drug was started (Figure 4). In these exploratory analyses, there di not appear to be any effect among women randomized after 20 weeks of gestation.

Figure 4. Time-to-delivery from date of randomization, by gestational age at randomization



Source: Statistical Review of NDA 21-945, dated 10/19/2006

Compared with subjects who were nonblack, subjects who were black tended to enroll at earlier gestational ages; see Table 3. Among those randomized at 18 weeks of gestation or earlier, 68% were black. The proportion of blacks decreased to 50% among those randomized after 20 weeks.

Table 3. Distribution of blacks by gestational age at randomization

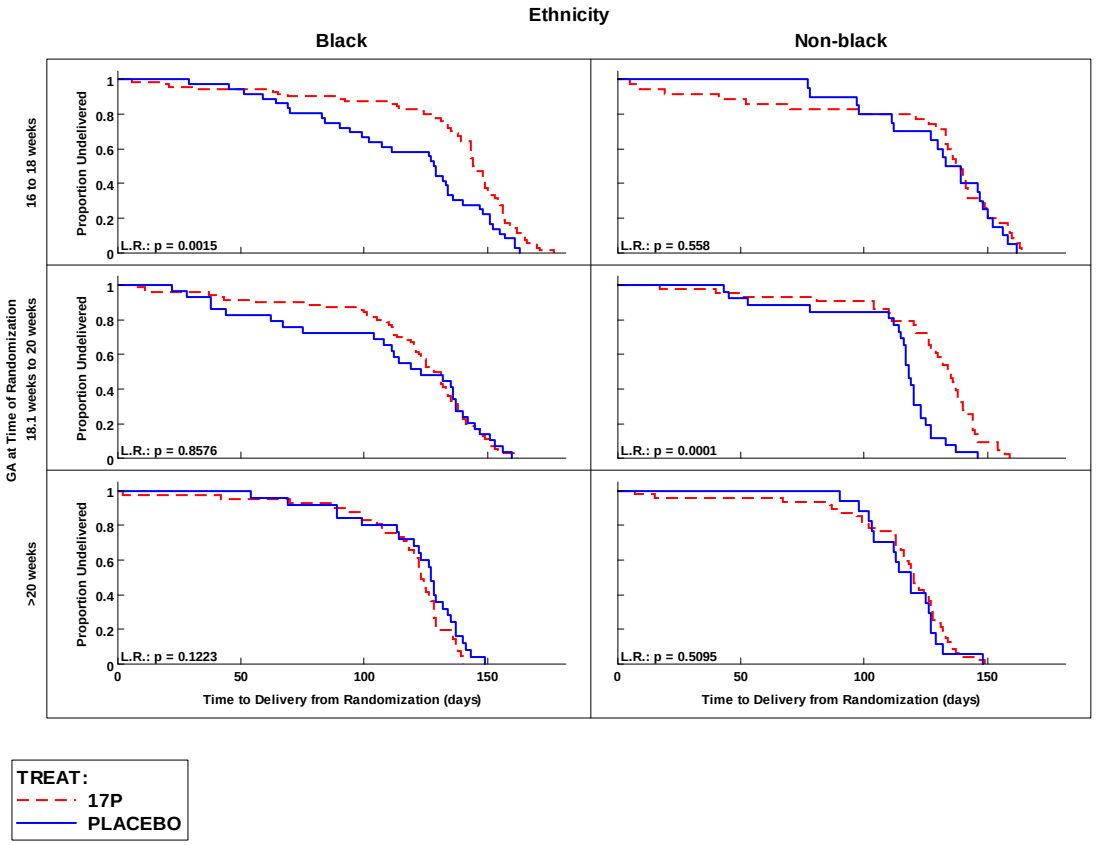
Gestational Age at Randomization	Total number of subjects	% Black	% Non-black
≤18 weeks	164	68%	32%
>18 and ≤ 20 weeks	170	58%	42%
>20 weeks	129	50%	50%
All subjects	*463*	*59%*	*41%*

Source: Statistical Reviewer

Potentially, results observed for women enrolled prior to 18 weeks of gestational age might be confounded with race, the University of Alabama or both. Notably, in addition to enrolling 43% of all study subjects who were black, the University of Alabama accounted for 44% (72/164) of all study subjects who enrolled prior to 18 weeks gestational age compared with 18% (54/199) of all study subjects who enrolled after 18 weeks of gestational age.

When time-to-delivery is examined by gestational age at the time of randomization, among subjects who are black, the treatment effect is most pronounced among those who were randomized prior to 18 weeks gestation; see Figure 5. Among subjects who are non-black, the treatment effect is most pronounced among subjects who were randomized between 18 weeks gestation and 20 weeks gestation. I discuss this finding later in the discussion on late pre-term deliveries in Section 3.2.3. These results are potentially confounded with the University of Alabama and, possibly, other centers that tended to enroll subjects at later gestational ages.

Figure 5. Time to delivery from date of randomization, by race and gestational age at the time of randomization.



Note: These figures exclude four subjects who were losses-to-follow-up,
Source: Statistical Reviewer

3.2.2 Deliveries <37 weeks gestation

The medical division is recommending approval under Subpart H, using deliveries prior to 37 weeks of gestation as the primary surrogate endpoint. I did not focus my attention on this endpoint in my earlier reviews because, during those previous review cycles, the surrogate endpoints recommended for approval were deliveries prior to 35 weeks gestation and prior to 32 weeks gestation. As I discussed in those reviews, the evidence from the single clinical trial was not sufficient to support the efficacy of 17P based on these endpoints and based on time-to-delivery.

The medical division now believes that a delivery prior to 37 weeks of gestation is an appropriate surrogate endpoint for Subpart H approval. According to the medical division, research articles published in the medical literature over the last several years support the clinical benefit of delaying so-called “late pre-term” deliveries. Late pre-term deliveries are defined as occurring, approximately, between 35 and 37 weeks gestation.

The protocol for Study 17P-CT002 specifies the analyses of the endpoint, *Delivery <37^0 Weeks Gestation (yes/no)*, would be based upon the total cohort of patients randomized, regardless of whether subjects took any study medication or not. This was the primary endpoint for the study when it was designed and conducted. The statistical analysis plan (SAP) further specifies that missing outcomes would be classified as a treatment failures (i.e., delivery < 37^0 weeks gestation). This affects four subjects who were losses to follow-up. Each of these subjects came from Center 18 and all 4 of these subjects were randomized to 17P.

Because of the interim analyses and the decision to stop the study early, the final analyses use a nominal p-value of 0.0345 (Z-score = 2.1232) to preserve the overall Type I error of 0.05. This nominal p-value is based on the 463 women who were randomized and who had outcome data. (The second interim analysis used outcome data from 351 women.)

In Study 17P-CT002 the results were:

Table 4. Applicant's Analysis: Delivery <37^0 Weeks Gestation

Data Source	17P		Placebo		Nominal P-value[a]	Treatment difference and its 96.6% Confidence Interval[b]
	N	n (%)	N	n (%)		
ITT population (all data)	310	115 (37.1)	153	84 (54.9)	0.0003	-17.8% [-28%, -7%]
All available data	306	111 (36.3)	153	84 (54.9)	0.0000	-18.6% [-29%, -8%]
Per-protocol population	271	99 (36.5)	134	75 (56.0)	0.0002	-19.5% [-30%, -8%]

[a] Chi-square test. To account for the interim analyses, the nominal p-values need to be compared to 0.0345.

[b] I calculated these confidence intervals, which are adjusted for the 2 interim analyses and the final analysis. To preserve the overall Type I error rate of 0.05, the confidence intervals use the final p-value boundary of 0.0345.

Although the statistical significance for the overall result for deliveries <37 weeks gestational age appears persuasive, an examination of the robustness of the result is important in determining whether the results from this single study are sufficient to support the effectiveness for 17P. Given that approval under Subpart H will based on this

single study, I believe it is important to understand the reasons for the increase in treatment effect between 35 weeks and 37 weeks.

In this section of my review, I explore whether a single study is sufficient for demonstrating the efficacy of 17P with deliveries prior to 37 weeks gestation as the endpoint of interest. In the next section, I also explore whether the deliveries that occurred between 35 and 37 weeks – "late pre-term deliveries" – can support the efficacy of 17P.

Of potential interest here is whey the point estimates for the treatment difference at 32 weeks and 35 weeks are consistent with each other (<32 weeks: -7.7%; <35 weeks: -9.4%), and then appear to approximately double between 35 weeks and 37 weeks (<37 weeks: -17.8%) (Table 5).

Table 5. Delivery <37^0 Weeks, <35^0 Weeks, <32^0 Weeks, <28^0 Weeks Gestation, ITT population. The estimates of rates of delivery do not account for duration of drug exposure.

Data Source	17P[a] (N=310) %	Placebo (N=153) %	Treatment difference and its 95% Confidence Interval, adjusted for interim analyses[b]
<37^0 weeks	37.1	54.9	-17.8% [-28%, -7%]
<35^0 weeks	21.3	30.7	-9.4% [-18.7%, -0.2%]
<32^0 weeks	11.9	19.6	-7.7% [-16.1.%, -0.3%]

[a] Four 17P-treated patients were losses-to-follow-up. They are counted as deliveries at their gestational ages at time of last contact (18.6, 22.0, 34.4 and 36.6 weeks).
[b] To preserve the overall Type I error rate of 0.05, the adjusted confidence intervals (equivalent to a 96.6% confidence interval) use the final p-value boundary of 0.0345. [4]

As discussed in Section 3.2.1 and my original review of Study 17P-CT002, it appears that the efficacy of 17P might depend on the gestational age at which subjects were randomized. This review supplements these time-to-delivery results described in my earlier reviews by exploring the effect, if any, of the timing of the start of treatment on the percentage of subjects who delivered <37 weeks gestation (Table 6).

Table 6 shows approximately 43% of all subjects, without regard to treatment, delivered <37 weeks gestation. Of the subjects who were randomized after 20 weeks of gestation, approximately 33% delivered at <37 weeks compared with around 50% who were

[4] The FDA presentation at the Advisory Committee meeting reported [-15.5%, 0.1%] as the confidence interval for the treatment effect for preterm deliveries <32 weeks. The Applicant provided this interval. Upon further review, I determined the interval should be [-16.1%, -0.3%].

randomized earlier. This suggests that gestational age at the time of randomization, independently of treatment assignment, could be an important predictor of outcome.

The observed treatment effects by gestational age at randomization for delivery <37 weeks are consistent with my previous review, in which I examined time-to-delivery. Among women randomized after 20 weeks of gestation, the treatment effect is close to zero, whereas the observed treatment effect is around 20% for those randomized at earlier ages.

This finding suggests that either 17P may not be effective, as assessed by deliveries <37 weeks gestational age, if started relatively late or, potentially, there was selection bias. That is, women who enrolled at later gestational ages may not have been at the same risk for preterm deliveries as those who enrolled at earlier gestational ages.

Among women randomized to 17P the proportion of those who delivered <37 weeks was fairly consistent among the three randomization categories (Table 6). By contrast, among women randomized to Placebo, the rate of delivery decreased from around 60% among those randomized prior to 20 weeks gestation to 35% among those randomized after 20 weeks gestation.

Table 6. Proportion of subjects who delivered <37^0 Weeks, by gestational age at randomization

	All Subjects		17P		Placebo		
Gestational Age at Randomization	N	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Treatment Difference
≤18 weeks	164	45.1	108	37.0	56	60.7	-23.7
>18 and ≤ 20 wks	170	48.8	113	41.6	57	63.2	-21.6
>20 weeks	129	32.6	89	31.5	40	35.0	-3.5
All subjects	*463*	*43.0*	*310*	*37.1*	*153*	*54.9*	*-17.8*

Source: Statistical Reviewer

The observation that the treatment effect is more pronounced when treatment is started at earlier gestational ages also might be confounded with race and study center, including the University of Alabama. As I discussed in Section 3.2.1, subjects who enrolled at earlier gestational ages tended to be black, while those who enrolled later tended to be non-black. Table 7 shows the treatment effects among subjects who were black mimic those seen for the entire study population. The treatment effect is most pronounced among black subjects randomized at earlier gestational ages and is essentially zero among those randomized after 20 weeks. Among subjects who were non-black, the only noticeable treatment effect occurred among subjects who were randomized between 18 and 20 weeks.

Table 7. Proportion of subjects who delivered <37^0 Weeks, by gestational age at randomization and race

		All Subjects		17P		Placebo		
Race	Gestational Age at Randomization	N	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Treatment Difference
Black	≤18 weeks	111	42.3	73	31.5	38	63.2	-31.7
	>18 and ≤ 20 weeks	98	49.0	68	45.6	30	56.7	-11.1
	>20 weeks	64	28.1	42	28.6	22	27.3	1.3
	All subjects	*273*	*41.4*	*183*	*36.1*	*90*	*52.2*	*-16.1*
Non-black	≤18 weeks	53	50.9	35	48.6	18	55.6	-7.0
	>18 and ≤ 20 weeks	72	48.6	45	35.6	27	70.4	-34.8
	>20 weeks	65	37.0	47	34.0	18	44.4	-10
	All subjects	*190*	*45.2*	*35*	*38.6*	*18*	*58.7*	*-20.1*

Source: Statistical Reviewer

3.2.3 Late pre-term deliveries

In this section, I explore whether the deliveries that occurred between 35 and 37 weeks – "late pre-term deliveries" – can support the efficacy of 17P.

An examination of late pre-term deliveries is important for at least two reasons. First, prevention of late pre-term deliveries is deemed important and is part of the basis of the proposed approval under Subpart H. Second, because approval will be based on a single study, a determination of whether these results can be generalized is also important. For this review, I define late pre-term deliveries as those occurring between 35 weeks and 37 weeks gestations.

As noted earlier, the treatment effect using an endpoint for <32 weeks (treatment effect: -7.7%) was consistent with the treatment effect that used an endpoint of <35 weeks (treatment effect: -9.4%); see Table 5. The treatment effect almost doubles between 35 weeks and 37 weeks (treatment effect: -17.8%) and is statistically significant ($p<0.001$). This section of my review attempts to characterize the deliveries that occurred between 35 weeks and 37 weeks, and to explore reasons for why the treatment effect is statistically significant at 37 weeks but not persuasive at earlier times.

Eighty-six deliveries occurred between 35 weeks gestation and 37 weeks gestation (Table 8). Of these 86, 49 (57%) were from subjects randomized to 17P; 37 (43%) were from subjects randomized to Placebo. An examination of the distribution of the 86 deliveries among study centers did not reveal any unusual patterns. The University of Alabama accounted for 17.4% of the 86 deliveries followed by the University of Texas Southwestern (Center 18) which accounted for 12.8% of the deliveries.

Table 8. Percentage (number) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group.

Treatment	Total Randomized	% Delivered <35 weeks	% Delivered 35 to 36.9 weeks	% Delivered ≥37 weeks
17P	310	21.3 (66)	15.8 (49)	62.9 (195)
Placebo	153	30.7 (47)	24.2 (37)	45.1 (69)
All subjects	*463*	*24.4 (113)*	*18.6 (86)*	*57.0 (264)*

Source: Statistical Reviewer

Subjects who were black and who were non-black delivered in equal numbers between 35 and 37 weeks (Table 9). Among subjects who were black, approximately 16% of 17P-treated and Placebo-treated subjects delivered between 35 and 37 weeks (Table 9). However, among subjects who were non-black, the rate of deliveries among subjects randomized to Placebo (35%) was double the rate among subjects randomized to 17P.

This finding is consistent with the Kaplan-Meir graphs displayed in Figure 2. Through Week 35, these graphs suggest the lack of beneficial effect among subjects who were non-black cancels the beneficial effect observed among subjects who were black. Starting at Week 35, the emergence of a treatment effect favoring 17P among subjects who were non-black accounts for the statistically significant outcome for deliveries using a cutpoint of <37 weeks gestation.

Table 9. Percentage (# delivered) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group and race.

Race	Treatment	Total Randomized	% Delivered <35 weeks	% Delivered 35 to 36.9 weeks	% Delivered ≥37 weeks
Black	17P	183	20.8 (38)	15.3 (28)	63.9 (117)
	Placebo	90	35.6 (32)	16.7 (15)	47.8 (43)
	All subjects	*273*	*25.6 (70)*	*15.8 (43)*	*58.6 (160)*
Non-black	17P	127	22.0 (28)	16.5 (21)	61.4 (78)
	Placebo	63	23.8 (15)	34.9 (22)	41.3 (26)
	All subjects	*190*	*22.6 (43)*	*22.6 (43)*	*54.8 (104)*

Source: Statistical Reviewer

Among those randomized to 17P, the patterns in the rates of deliveries are consistent across the three categories of gestational age at randomization, as shown in Table 10. The rates of delivery <35 weeks are around 20%, decrease somewhat between 35 and 37 weeks and then increase to around 60% after 37 weeks. Among subjects randomized to Placebo, however, the rates of delivery appear to depend on the gestational age at randomization.

Table 10. Percentage (# delivered) of subjects who delivered <35 weeks, 35 weeks to 36.9 weeks, and ≥37 weeks, by treatment group and gestational age at randomization.

Gestational Age at Randomization	Treatment	Total Randomized	% Delivered <35 weeks	% Delivered 35 to 36.9 weeks	% Delivered ≥37 weeks
≤18 weeks	17P	108	22.0 (24)	14.8 (16)	63.0 (68)
	Placebo	56	42.9 (24)	17.9 (10)	39.2 (22)
>18 and ≤ 20 weeks	17P	113	22.1 (25)	19.5 (22)	58.4 (66)
	Placebo	57	29.8 (17)	33.3 (19)	36.8 (21)
>20 weeks	17P	89	19.1 (17)	12.4 (11)	68.5 (61)
	Placebo	40	15.0 (6)	20.0 (8)	65.0 (26)

Source: Statistical Reviewer

3.3 Evaluation of Safety

My review of 17P-CT002 discussed issues surrounding fetal deaths (i.e., miscarriages and stillbirths) and neonatal deaths. Briefly, subjects randomized to 17P experienced a higher rate of fetal deaths than did subjects randomized to Placebo.

Fetal and neonatal deaths appeared to depend on the gestational age at the time of randomization[5] (Table 11).

Table 11 Distribution of Miscarriages, Stillbirths and Neonatal Deaths, by Gestational Age at Randomization

	Gestational Age at Randomization		
	≤18 weeks	>18 and ≤ 20 weeks	>20 weeks
N	162	168	129
% deaths	10.5%	6.0%	2.3%

Note: This table excludes four subjects who were losses to follow-up.
Source: Statistical reviewer

Of the 30 fetal and neonatal losses, 10 occurred at the University of Alabama (Center 8).

[5] Note: The entries in Table 11 and Table 12 do not agree with my original review. I discovered a programming error and have updated these tables accordingly.

Table 12 Distribution of Fetal and Neonatal Deaths, by Center and Gestational Age at Randomization

	Gestational Age at Randomization		
	≤18 weeks	18.1 – 20 weeks	>20 weeks
Number of deaths:	17	10	3
Center #			
2	1 (5.9%)	-	-
4	3 (17.6%)	2 (20.0%)	-
8	7 (41.2%)	3 (30.0%)	-
9	-	-	1 (33.3%)
13	1 (5.9%)	1 (10.0%)	1 (33.3%)
14	-	1 (10.0%)	1 (33.3%)
15	3 (17.6%)	1 (10.0%)	-
17	-	1 (10.0%)	-
18	1 (5.9%)	-	-
21	1 (5.9%)	-	-
23	-	1 (10.0%)	-

Note: This table excludes four subjects who were losses to follow-up.
Source: Statistical reviewer

Overall, the crude rates of subjects who had a fetal or neonatal loss did not appear to vary by race (Table 13). Note, however, these are crude rates and do not account for duration of exposure to study treatment.

Table 13. Crude rates of fetal and neonatal losses, by race. Rates are not adjusted for duration of exposure to study treatment.

	17P		Placebo	
Race	Total number of subjects	% Deaths	Total number of subjects	% Deaths
Black	271	7.0 (19)	181	6.1 (11)	90	8.9 (8)
Non-black	188	5.9 (11)	125	6.4 (8)	63	4.8 (3)
All subjects	*459*	*6.5*	*306*	*6.2*	*153*	*7.2*

Note: This table excludes four subjects who were losses to follow-up.
Source: Statistical reviewer

When accounting for duration of exposure, however, the differences between 17P and Placebo in the rates of early deliveries appear related to the race of subjects (Figure 2 and Table 14). For both racial groups, the rate of early deliveries is increased among those randomized to 17P compared with those randomized to Placebo.

However, the patterns between subjects who are black and who are non-black are strikingly different. Among subjects who were black, by 24 weeks of gestation the rates of fetal and neonatal losses were similar for both treatment groups – approximately 6%[6]. A different picture emerges for subjects who were non-black. By 24 weeks of gestation, there were no fetal or neonatal losses among those randomized to Placebo, compared with an estimated rate of 9% among those randomized to 17P.

Table 14. Estimated Rates of Fetal Losses (miscarriages and stillbirths) and Neonatal Deaths, accounting for time on study drug, by race.

	Blacks				Non-blacks			
	17P		Placebo		17P		Placebo	
Week of Gestation	# at risk	% (n)	# at risk	% (n)	# at risk	% (n)	# at risk	% (n)
20	**140***	2.2% (2)	**76**	0.0% (0)	**61**	5.9% (3)	**52**	0.0% (0)
22	**174**	6.2% (8)	76	1.1% (1)	123	7.4% (5)	**63**	0.0% (0)
24	174	6.2% (8)	86	5.8% (5)	120	8.9% (7)	63	0.0% (0)

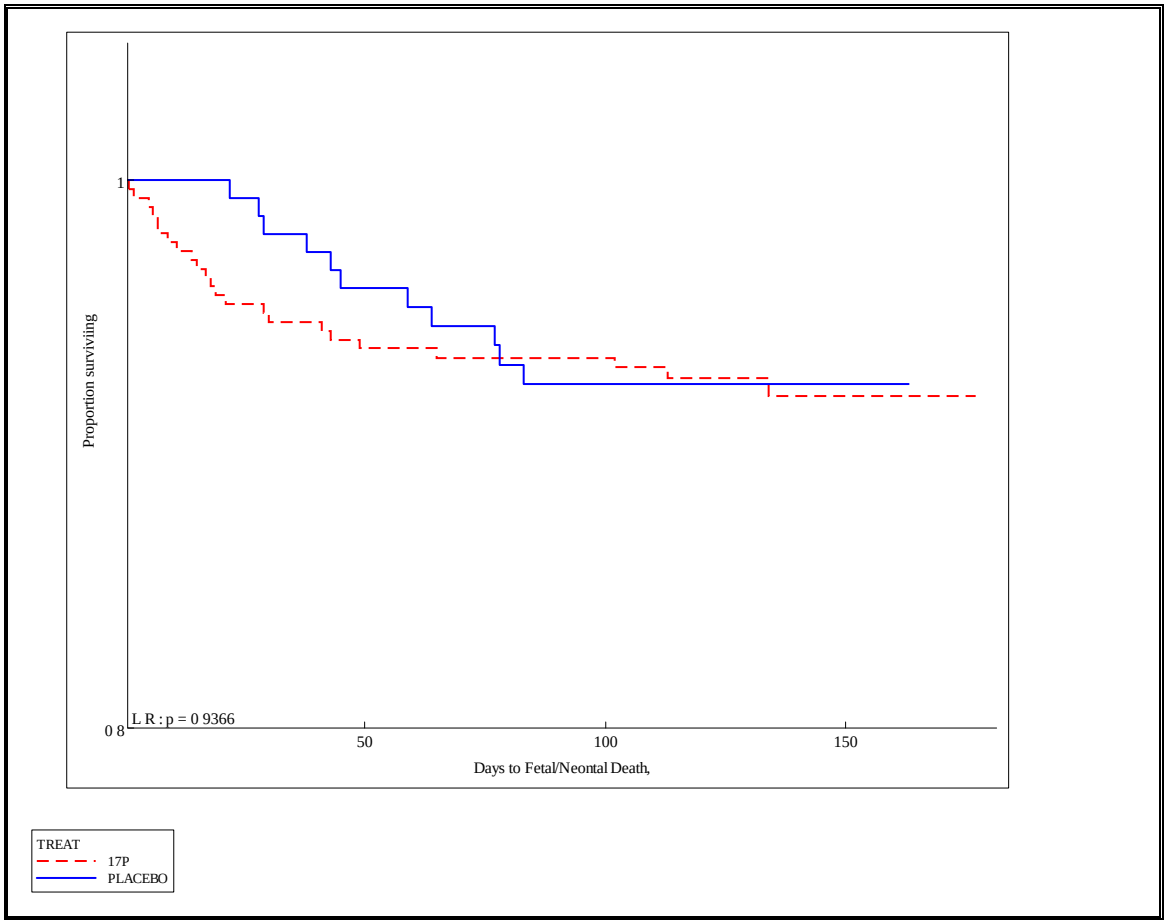
* Entries in **bold** are estimates of the # at risk.

Source: Kaplan-Meier estimates contained in Applicant's "Response to Information Request from 14 January 2011," submission dated 1/25/201.

The following graphs show time to fetal and neonatal deaths as a function of duration of exposure to study treatment by using date of randomization as the baseline (Figure 6 and Figure 7). These figures reinforce the potential signal of an increased rate of an increased rate of fetal and neonatal losses among subjects who were non-black and who were randomized to 17P. When interpreting the figures, note that the y-axis starts at 80%.

[6] The first delivery that was not a fetal loss or neonatal death occurred just after 24 weeks gestation.

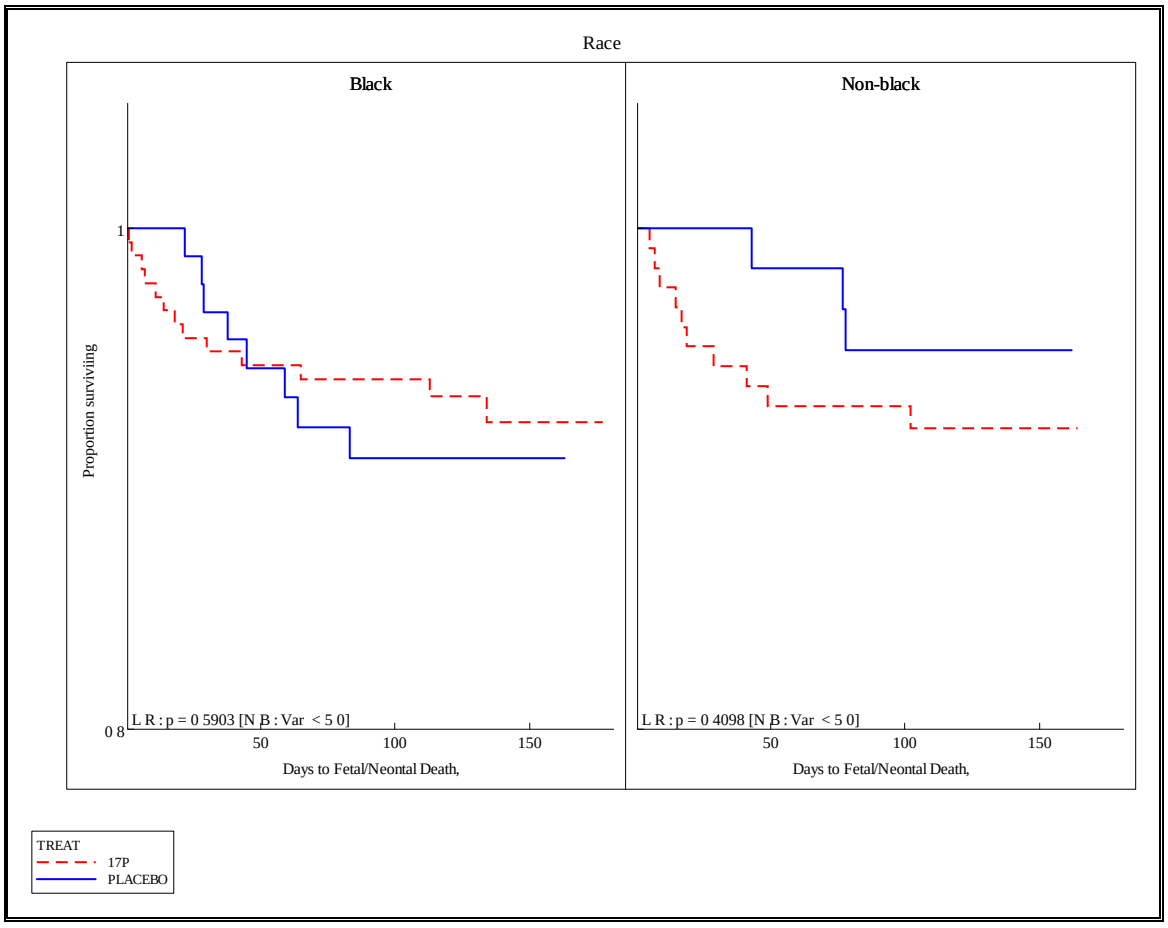
Figure 6. From time of randomization to fetal and neonatal deaths, by treatment group.



Note: y-axis starts at 80%

Source: Statistical reviewer

Figure 7. From time of randomization to fetal and neonatal deaths, by treatment group and race.

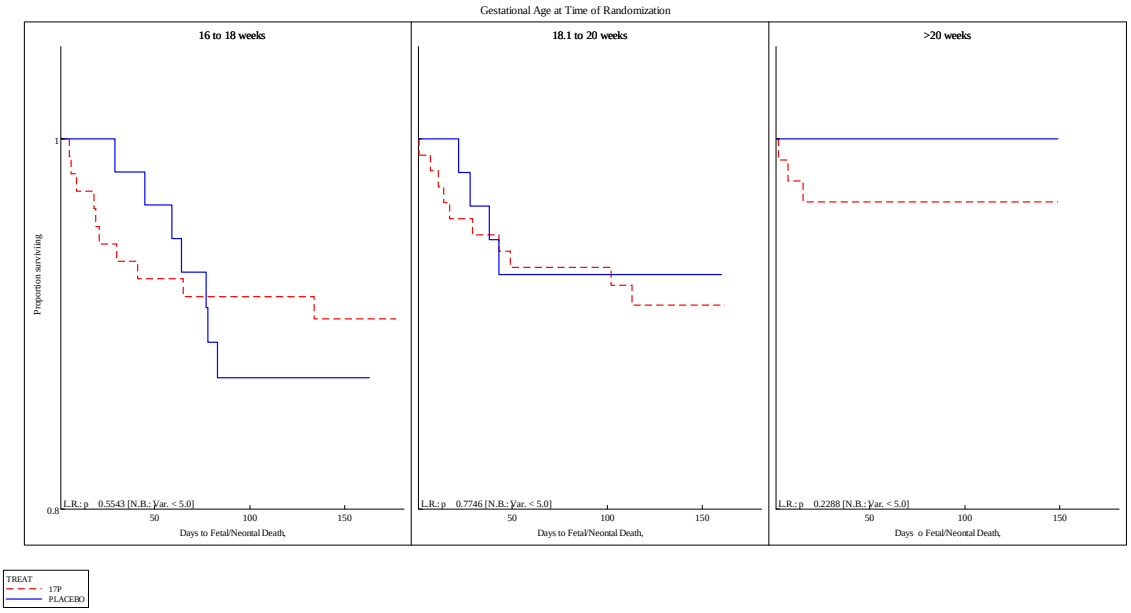


Note: y-axis starts at 80%

Source: Statistical reviewer

For completeness, I include the following figure that shows time to fetal and neonatal deaths by gestational age at randomization, although the sample sizes may be too small to make any meaningful conclusions.

Figure 8. From time of randomization to fetal and neonatal deaths, by treatment group and gestational age at the time of randomization.



Note: y-axis starts at 80%
Source: Statistical reviewer

3.4 Comments on labeling and promotional materials

I am concerned about the label's presentation of efficacy data for endpoints other than <37 weeks, the inclusion of information on the neonatal mortality/morbidity index and the lack of information on the findings for important subgroups.

3.4.1 Preterm deliveries

Because the approval of 17P will be based on deliveries <37 weeks gestation, I recommend that the label includes efficacy data for this endpoint only, and that the label excludes efficacy information pertaining to deliveries <35 weeks and to deliveries <32 weeks. As noted in my other reviews, the data for these two endpoints are insufficient to support the efficacy of 17P. Moreover, I recommend adding text indicating 17P has not been shown to be effective in reducing the risk of deliveries at earlier time points.

These recommendations are consistent with the guidance document, "Guidance for Industry: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content and Format"[7]. The guidance document states:

[7] Available at
www fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075059.pdf

- "The primary objective of the CLINICAL STUDIES section is to summarize (1) the evidence supporting effectiveness in the subjects who were studied, (2) the critical design aspects of the studies, including the populations studied and endpoints measured, and (3) the important limitations of the available evidence."

- "The CLINICAL STUDIES section should present those endpoints that establish the effectiveness of the drug or show the limitations of effectiveness."

These recommendation affect the table titled, "Proportion of Subjects Delivering at <37, <35 and <32 Weeks Gestational Age (ITT Population)", by removing the two rows corresponding to <35 weeks and <32 weeks.

Also affected is the text, "Compared to controls, treatment with Makena reduced the proportion of women who delivered preterm at < 37, (b) (4)", which I recommend deleting. Not only does this sentence duplicate information found elsewhere in the label, it also has a promotional quality. Information on deliveries at time points other than <37 weeks is contained in the Kaplan-Meier graph that is in the label; it is also in the table that I recommend modifying. If the medical division decides to include this text, it should be reworded to exclude "reduced", which has a promotional connotation. An alternative wording is: "Compared to women treated with placebo, the proportions of deliveries at < 37 (b) (4) weeks were lower among women treated with Makena."

If these changes are not adopted, I am concerned the labeling will overstate the effectiveness of 17P to prescribers reading the label, and that the effectiveness will be overstated in promotional materials. Promotional materials can include anything that appears in labeling. These concerns are evidenced by the draft press release and draft sales aid included in the DDMAC consult request dated 1/10/2011.

3.4.2 Neonatal mortality/morbidity index

The labeling information on the neonatal mortality/morbidity index is problematic, because it counts miscarriages and stillbirths as successes despite some of them occurring at the same time as a neonatal death, which is counted as a failure. For example, at 20.1 weeks a neonatal death occurred for Case 11 and a miscarriage/stillbirth for Case 12 (Table 15). Yet, Case 11 is counted as a failure while Case 12 is counted a success.

Table 15. Listing of Miscarriages, Stillbirths and Neonatal Deats, sorted by Treatment, Race and Gestational age at delivery

Treatment	Race	Case	Center ID	Patient ID	Gestational age at delivery (weeks)	Gestational age at Randomization (weeks.days)	Classification of death
17P	Nonblack	1	8	CT-008-110	18.3	17.5	Miscarriage or Stillbirth
		2	15	CT-015-014	18.9	16.2	Miscarriage or Stillbirth
		3	15	CT-015-023	19.1	18.0	Miscarriage or Stillbirth
		4	23	CT-023-007	21.0	18.5	Miscarriage or Stillbirth
		5	14	CT-014-012	21.1	20.2	Miscarriage or Stillbirth
		6	18	CT-018-024	22.1	16.3	Miscarriage or Stillbirth
		7	13	CT-013-014	22.6	20.4	Neonatal Death
		8	21	CT-021-033	24.7	17.6	Neonatal Death
	Black	9	4	CT-004-048	18.1	17.3	Miscarriage or Stillbirth
		10	8	CT-008-114	19.1	16.2	Miscarriage or Stillbirth
		11	4	CT-004-035	20.1	16.0	Neonatal Death
		12	17	CT-017-011	20.1	19.2	Miscarriage or Stillbirth
		13	8	CT-008-102	20.3	18.3	Miscarriage or Stillbirth
		14	9	CT-009-045	20.3	20.1	Neonatal Death
		15	15	CT-015-022	20.3	17.6	Miscarriage or Stillbirth
		16	4	CT-004-043	20.7	19.2	Neonatal Death
		17	8	CT-008-107	25.4	16.2	Neonatal Death
		18	14	CT-014-017	25.6	19.4	Neonatal Death
		19	8	CT-008-142	35.1	19.1	Neonatal Death
Placebo	Nonblack	20	13	CT-013-026	24.3	18.2	Neonatal Death
		21	2	CT-002-015	28.0	17.0	Neonatal Death
		22	13	CT-013-005	28.9	18.0	Miscarriage or Stillbirth
	Black	23	8	CT-008-171	20.4	16.3	Neonatal Death
		24	4	CT-004-054	22.9	19.6	Neonatal Death

	25	15	CT-015-032	23.4	19.4	Neonatal Death
	26	8	CT-008-075	23.6	17.2	Neonatal Death
	27	8	CT-008-060	23.9	18.4	Miscarriage or Stillbirth
	28	4	CT-004-023	25.0	16.5	Neonatal Death
	29	8	CT-008-087	25.1	16.1	Neonatal Death
	30	8	CT-008-091	28.1	16.3	Neonatal Death

I recommend replacing the neonatal index with an index that includes all cause mortality, not just deaths occurring among live births. By not including all deaths, the finding reported in the draft labeling overstates the efficacy of 17P. If the medical division decides to include the index, I recommend adding verbiage that indicates the index excludes fetal losses along with the number per treatment group that is excluded.

In addition, the inclusion of results in labeling with point estimates favoring a drug product even with the disclaimer "not statistically significant" is not advisable. Conceivably, the lower observed rate in 17P versus Placebo could be promoted as a benefit for 17P despite the phrase "not statistically significant".

Finally, including this information seems to contradict the statement contained in the indications and usage section: "There are no controlled trials demonstrating a direct clinical benefit, such as improvement in neonatal mortality and morbidity."

3.4.3 Racial subgroups

The guidance document also recommends the inclusion of summary statements about the results of required explorations. As such, I recommend including a statement that examination of racial subgroups suggests a larger treatment effect in African-American women, and a higher rate of early losses among women who are not African-American.

3.4.4 Gestational age at randomization

I also recommend including text indicating 17P may not be effective if treatment is started after 20 week of gestation. This finding in an important limitation of the study results.

3.5 Comments on ongoing confirmatory study, 17P-ES-003

In my review of the first Complete Response, I reviewed the protocol for the now ongoing confirmatory study. I made clear that a confirmatory study that fulfills the requirements for Subpart H approval needs to have a clinical endpoint as its primary endpoint – not a surrogate marker.

I reiterate the importance of having a true clinical endpoint as the primary endpoint for that study in order to meet the requirements of Subpart H.

As currently designed, the primary endpoint in the ongoing confirmatory study is a surrogate endpoint: deliveries <35 weeks of gestational age. The neonatal morbidity/mortality index is a secondary endpoint.

At the request of the medical division, the applicant proposes elevating the neonatal morbidity/mortality index to a co-primary endpoint[8]. The neonatal index includes neonatal deaths and neonatal morbidities among live births; miscarriages and still births are excluded from the analysis population. Delivery prior to 35 weeks of gestation is the other co-primary endpoint.

Because the results from the study will be used to confirm the clinical benefit of 17P in order to fulfill the Subpart H requirements, the primary endpoint needs to be a clinical endpoint. Deliveries <35 weeks is a surrogate endpoint, not a clinical endpoint, and isn't appropriate for confirming the clinical benefit of 17P.

Any clinical endpoint that will be used for the basis of approval needs to account for all subjects enrolled in the study. The proposed co-primary endpoint, the neonatal composite index, assesses live births only; miscarriages and stillbirths are excluded. An analysis set that excludes subjects based on post-randomization events violates the intention-to-treat principle. The primary analyses for the purpose of approval need to account for all subjects and their births.

An analysis limited to live births could be a secondary analysis.

3.6 Draft protocol for a follow-up study (Study 17P-FU-004) of children born to mothers who received 17P or placebo in the ongoing confirmatory study required for Subpart H approval

3.6.1 Summary of protocol

The division's approvable letter indicated, "additional developmental assessment is needed of children at ages 18-24 months whose mothers had been treated with HPC." To address this issue, the submission references the protocol for 17P-FU-004, which was submitted to IND 68,108 on 6/29/2009:

> "A prospective, noninterventional follow-up study of children aged 23 to 25 months, born to mothers who received hydroxyprogesterone, caproate injection, 250 mg/ml, or vehicle for the prevention of preterm birth."

The study objective of this ongoing study is to determine whether there is a difference in the achievement of developmental milestones between children whose mothers received

[8] Proposed Updates to Protocol 17P-ES-003, submitted on 1/7/11 by email from Robb Hesley, Hologic, Inc.

17P and those who received placebo in the ongoing confirmatory study required for Subpart H approval.

Informed consent will be obtained from the subject's mother/legal guardian between delivery and discharge from the delivery hospitalization. Mothers/legal guardians will be contacted periodically until the child nears the age of 18 months. If the mother/legal guardian is interested in continuing in the study, the Ages and Stages Questionnaire (ASQ) will be mailed in order to screen the child for a developmental delay.

If the questionnaire suggests a delay as measured by falling below a specified cutoff in at least 1 developmental area on the ASQ, the child will be referred for follow-up assessments. If more than one area is identified, secondary assessments will be done. Depending on the developmental area identified, the assessment may be the Bayley Scales of Infant Development, Modified Checklist for Autism in Toddlers, a neurological exam or the Gross Motor Function Classification System.

The primary outcome is the proportion of children who fall below the specified cutoff for at least one of the developmental areas assessed by the ASQ. Differences between treatment groups will be compared with a chi-square test. Secondary analyses will consider each of the five domains individually.

Subjects will be enrolled until 375 completed ASQs are obtained. An ASQ is considered complete if each of the 5 domains has no more than 2 unanswered questions. The protocol anticipates 450 to 500 children are expected to be enrolled to reach 375 completed ASQs.

To be enrolled in the study, a subject must be between 22 and 25 months of age adjusted for gestational age. Subjects born to women who are unblinded to study group assignment will be excluded from the study. Moreover, an investigator may withdraw a subject from the study if the subject's parent(s)/legal guardian is made aware of the mother's treatment assignment or if the subject's parent(s)/legal guardian fails to comply with the study protocol.

Assuming a completed ASQ is obtained for 375 children (250 17P and 125 placebo), the study will have 88% power to detect a difference of 15%, using an overall Type I error rate of 5% and assuming an outcome rate of 30% in the 17P group. The protocol also indicates the sample size will provide "sufficient power to detect a two-fold increase in the 17P group in the proportion of children with the primary outcome.

3.6.2 Statistical comments

The study hypotheses and sample size calculations are not appropriate for a trial designed to rule out differences in safety between treatment and placebo. Instead of a non-inferiority design, the study is designed as a superiority trial to show a difference of 15% between the treatment groups. With the proposed design, a finding of a non-significant difference cannot support a conclusion of "no difference" between treatment groups in long-term outcomes.

Although the protocol states the study will provide "sufficient power to detect a two-fold increase in the 17P group in the proportion of children with the primary outcome," in fact the study is designed as a superiority study and is powered to detect a difference of 15% between treatment groups.

Sample sizes need to be recalculated to rule out a clinically important increase in the risk of untoward outcomes among children exposed to 17P relative to those exposed to placebo.

The subjects will come from the ongoing confirmatory study, which plans to enroll 1700 women; 10% will be from the United States and Canada. The medical division may want to require follow-up for all children born to the 170 subjects expected to be enrolled in the United States and Canada.

Instead of excluding subjects born to women who are unblinded to treatment assignment, all subjects should be enrolled in the study. Similarly, subjects should remain in the study, even if their treatment assignment becomes unblinded during the course of the study.

In a response to a request for information to clarify these issues, the Applicant submitted an email message on 11/18/2010; see Memorandum to File signed by Ms. Williamson and dated 12/22/2010. I discuss the Applicant's responses in the following paragraphs.

- The Applicant's response indicates the study is large enough to rule out a doubling in the proportion of children with the primary outcome:

  "Based on your request, we have confirmed that the current study sample size is sufficiently powered to rule out a doubling in the 17P treatment group, relative to the vehicle group, in the proportion of children with the primary outcome. Based on data from the NICHD Follow-up study, 28% of children in both the 17P and vehicle groups fell below the specified cut-off for at least 1 developmental area on the ASQ. Thus, a completed ASQ obtained for at least 250 17P and at least 125 vehicle subjects will allow for 95% power to rule out a doubling in the proportion of children with the primary outcome, given a rate of 28%. Further if the rate for the vehicle arm is as low as 18%, there would be an 80% power to exclude a doubling in risk of adverse outcomes."

Although I did not confirm the power calculations, the Applicant's response appears appropriate.

- The Applicant indicated their intent to include as many children born to women enrolled in the US and Canada as possible in the follow-up study:

  "We are committed to ensuring that as many of the 375 subjects for study 17P-FU-004 as possible are entered in the US and Canada while meeting our post approval commitment date. To date, 65 subjects (all from the United States) have consented to be recontacted for participation in the Infant Follow-up Study.

  At all participating US/Canadian sites, we are encouraging every eligible patient to consent for participation in the 17P-FU-004 study. We recognize the importance of including as many North American subjects as possible in study 17P-FU-004 and will continue to pursue every available subject."

  The Applicant's response did not explicitly state that all children born to women enrolled in the US/Canadian sites will be enrolled in the follow-up study. If the infant follow-up study fails to enroll a sufficient number of children from the North American study sites, the generalisability of the results to the United States could be difficult.

- The Applicant's response suggests the potential for unblinding among subjects enrolled in the confirmatory study will be minimal, although the response recognizes the potential for selection bias:

  "If there is concern that the study may experience selection bias due to the potential that patients with certain pregnancy complications may be unblinded at the request of the investigator, we can provide assurance that the sponsor and investigators are committed to maintaining the study blind. To date with 171 patients randomized in the 17P-ES-003 study, only one patient has required unblinding and this was due to an protocol deviation in which the study was administered a 5X overdose of the blinded study medication."

  When the follow-up study is submitted for review, I recommend descriptive analyses summarizing the number of subjects who were unblinded and the reason for unblinding. This would apply both to subjects who were unblinded prior to enrollment in the follow-up study and those who were unblinded during participation in the follow-up study.

- The Applicant's response does not appear to explain the issue I raised regarding the withdrawal of a subject for non-compliance:

"This statement does not apply to the Confirmatory Study. The requirement is in section 5.6.1 of the study under the heading subject withdrawal. Per section 5, a subject will be considered enrolled once the ASQ has been mailed to their parent(s)/legal guardian. Section 5.6.1 applies only to enrolled subjects on study 17P-FU-004 and compliance only refers to study 17P-FU-004."

Although the Applicant's response states the non-compliance issue applies to the infant follow-up study – not the ongoing confirmatory study, the response does not describe the circumstances under which a subject might be non-compliant and, therefore, withdrawn from the study.

4. Conclusions and Recommendations

From a statistical perspective, the information and data submitted by the Applicant do not provide convincing evidence regarding the effectiveness of 17 α-hydroxyprogesterone, caproate injection (17P) for the prevention of preterm deliveries among women with a history of at least one spontaneous preterm delivery.

In addition to the issues surrounding the level of evidence provided by a single study, discussed below, the use of Subpart H as a pathway for approval does not seem appropriate for 17P. Unlike studies of HIV and cancer where the difference in time between the outcome of a surrogate endpoint and a clinical endpoint can be years, in this situation, the time between the clinical outcome of interest (i.e., mortality and neonatal morbidity) and the surrogate outcome (<37 weeks) is literally weeks. The fact that a confirmatory study is currently ongoing does not translate into a lesser standard of evidence needed to conclude efficacy based on the evaluation of an endpoint from a single study. The data from the single study submitted for approval, for the reasons summarized below, are insufficient to support the efficacy of 17P.

The Applicant is seeking approval based on the results from only one adequate and well-controlled study, which has been submitted for review. The study, submitted with the original NDA, had several features that do not allow the study to stand on its own to establish the efficacy of 17P on the surrogate endpoint of preterm deliveries, as described in the guidance document, "Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products."[9]

In my previous review of the study (see Attachment 2), I focused on the endpoints of delivery <35 weeks, delivery <32 weeks and time-to-delivery. My reasons for concluding that a single study was not sufficient to support the effectiveness of 17P in preventing preterm deliveries were:

[9] Available at www fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf

- Optimal time to start study drug was not identified.
    - 17P appeared most effective when started at 18 weeks of gestation or earlier; did not appear effective when started at 20 weeks of gestation or later.
    - Rate of fetal and neonatal deaths is most pronounced among births to women who started 17P at 18 weeks gestation or earlier (10%).
- Apparent confounding of study site and gestational age at randomization.
    - One center accounted for 44% of subjects enrolled at 18 weeks of gestation or earlier.
    - Some centers had a deficit of subjects enrolled at 18 weeks of gestation or earlier.
- Fetal and neonatal deaths among women treated with 17P occur earlier than among women treated with placebo.
- One center accounted for a relatively large proportion of all subjects enrolled.

However, recognizing an important public health need for the commercialization of this drug product, the medical division is currently recommending approval under Subpart H, based on a statistically significant treatment effect for the surrogate endpoint of deliveries prior to 37 weeks gestational age. This endpoint is a departure from the earlier review cycles that focused on the surrogate endpoints of deliveries prior to 32 weeks and deliveries prior to 35 weeks. My previous reviews did not sufficiently address the results at 37 weeks at the depth required to establish the efficacy of 17P based on a single study. In addition, my reviews did not explore whether the results from these endpoints were consistent among racial subgroups.

In this review of the second Complete Response, I have done additional analyses to address whether the data are sufficient to support approval if the endpoint of deliveries <37 weeks gestation is used as the surrogate endpoint. I have also done additional analyses exploring the effect of race on the efficacy results. However, the results from these analyses do not support the efficacy of 17P based on a single study.

My conclusion that the results from these additional analyses do not support the efficacy of 17P based on a single study are:

- The treatment effect at 37 weeks does not appear to be consistent among groups defined by gestational age at randomization. This finding may be confounded with race and study center.
- Lack of consistency of efficacy results among subgroups defined by race.
    - For subjects who were black, the benefit of 17P compared with Placebo appears to emerge at around 24 weeks.
    - For subjects who were non-blacks, a treatment benefit does not emerge until 35 weeks gestation.

- Lack of consistency of safety results at Week 24 among subgroups defined by race.
    - Among subjects who were black, the estimated rate of fetal and neonatal losses was 6% for subjects, regardless of treatment assignment.
    - Among subjects who were non-black, subjects randomized to Placebo did not have any fetal or neonatal losses compared with an estimated rate of 9% among those randomized to 17P.
- The doubling of the treatment effect from <35 weeks to <37 weeks is likely due to the increased number of deliveries among non-black subjects randomized to Placebo.

These exploratory analyses were necessary because of the reliance on a single study to support the approval of 17P. In some cases, the observed treatment effects may have been based on small numbers of subjects. However, the overall objective was to look at consistency among various endpoints and across various subgroups to determine whether the results could be extrapolated to a larger population in the absence of a second study, and these are the only data we have.

I recommend that the final label (1) include only those data on which approval will be based and (2) describe the limitations of the results. Because the approval of 17P will be based on the surrogate endpoint of deliveries <37 weeks gestation, I recommend that the label include efficacy data for this endpoint only and exclude efficacy information pertaining to deliveries <35 weeks and to deliveries <32 weeks. As noted in my previous reviews, the data for these two endpoints coming from a single study are insufficient to support the efficacy of 17P. Moreover, I recommend adding text indicating 17P has not been shown to be effective in reducing the risk of deliveries at earlier time points. These recommendations are consistent with the guidance document, "Guidance for Industry: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content and Format".

The guidance document also recommends the inclusion of summary statements about the results of required explorations. As such, I recommend including a statement that examination of racial subgroups suggests a larger treatment effect in African-American women, a higher rate of early losses among women who are not African-American, and the apparent absence of an effect when treatment is started after 20 weeks gestation.

I recommend excluding from the label the results from the neonatal mortality/morbidity index. First, including the index seems to contradict the statement contained in the indications and usage section: "There are no controlled trials demonstrating a direct clinical benefit, such as improvement in neonatal mortality and morbidity."

In addition, the index is limited to live births only and, therefore, counts miscarriages and stillbirths as successes regardless of their timing. For example, a neonatal death at 20.1 weeks gestation was counted a failure while a stillbirth at 20.1 weeks gestation was counted a success. If an index is needed, I recommend replacing the neonatal index with an index that includes all cause mortality, not just deaths occurring among live births. By not including all deaths, the finding reported in the draft labeling overstates the efficacy of 17P.

Finally, including results in labeling with point estimates favoring a drug product even with the disclaimer "not statistically significant" is not advisable. Conceivably, the lower observed rate in 17P versus Placebo could be promoted as a benefit for 17P despite the phrase "not statistically significant".

The ongoing confirmatory study, 17P-ES-003, will be used to confirm the clinical benefit of 17P in order to fulfill the Subpart H requirements. As currently designed the primary endpoint is a surrogate endpoint, deliveries <35 weeks. The use of this surrogate endpoint, instead of clinical endpoint, will not be sufficient for confirming the clinical benefit of 17P.

Any clinical endpoint that will be used for the basis of approval needs to account for all subjects enrolled in the study. The proposed co-primary endpoint, the neonatal composite index, assesses live births only; miscarriages and stillbirths are excluded. An analysis set that excludes subjects based on post-randomization events violates the intention-to-treat principle. The primary analysis for the purpose of approval needs to account for all subjects and their births. An analysis limited to live births could be a secondary analysis.

Attachment 1.
Statistical Review of NDA 21-945, Complete Response, dated 1/23/2009

APPEARS THIS WAY ON ORIGINAL

Attachment 2.
Statistical Review of NDA 21-945, dated 10/19/2006

APPEARS THIS WAY ON ORIGINAL

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

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number:	21-945
Drug Name:	Gestiva (17 α-hydroxyprogesterone; caproate injection)
Indication(s):	Prevention of preterm birth in pregnant women with a history of at least one spontaneous preterm birth
Applicant:	Cytyc Corporatioin
Date(s):	Date stamp: 4/25/2008
	PDUFA date: 1/25/2009 (with 3-month extension)
Review Priority:	6-month – Complete Response to Approvable Letter
Biometrics Division:	DB 3
Statistical Reviewer:	Lisa A. Kammerman, Ph.D.
Concurring Reviewers:	Mahboob Sobhan, Ph.D.
Medical Division:	Reproductive and Urological Products
Clinical Team:	Barbara Wesley, MD; Lisa Soule, MD
Project Manager:	Charlene Williamson

Keywords: clinical studies, NDA review, surrogate outcomes, generalisability, standard of evidence, non-inferiority.

Table of Contents

1. EXECUTIVE SUMMARY **3**

2. INTRODUCTION **4**

2.1 OVERVIEW OF COMPLETE RESPONSE 4
2.2 DATA SOURCES 4

3. STATISTICAL EVALUATION **5**

3.1 APPROVAL UNDER SUBPART H, 21 CFR 314.510 5
3.1.1 Surrogate endpoint 5
3.1.2 Evidence for an effect on preterm births (i.e., surrogate endpoint) 6
3.1.3 Appropriateness of Subpart H, Section 314.510 for this drug product 7
3.2 DRAFT PROTOCOL OF A PHASE 4 STUDY 8
3.2.1 Summary of protocol 8
3.2.2 Statistical Comments 9
3.3 DRAFT PROTOCOL OF A FOLLOW-UP STUDY OF CHILDREN BORN TO MOTHERS WHO RECEIVED 17P OR PLACEBO IN THE PHASE 4 STUDY. 12
3.3.1 Summary of protocol 12
3.4 PROMOTIONAL MATERIALS 13

4. CONCLUSIONS AND RECOMMENDATIONS **13**

1. EXECUTIVE SUMMARY

In this Complete Response to the Approvable Letter for NDA 21-945, the Applicant is seeking approval of 17 α-hydroxyprogesterone, caproate injection through Subpart H, Section 510 of the CFR. Under Subpart H, a drug product may be approved if adequate and well-controlled clinical trials establish the drug product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit.

The original NDA, which was a single study submission, is being used to support the efficacy of 17 α-hydroxyprogesterone, caproate injection in reducing preterm deliveries. The Complete Response does not contain any additional efficacy data.

From a clinical perspective, preterm delivery is reasonably likely to predict fetal and neonatal losses and neonatal morbidity. Thus, the use of preterm births as a surrogate endpoint appears to meet one of the requirements of Subpart H.

Study 17P-CT002, which was the singly study included in the original NDA, showed statistically significant reductions in preterm deliveries at <35 weeks and at <32 weeks. The medical team concluded these results were sufficient to support the efficacy of 17 α-hydroxyprogesterone, caproate injection.

However, from a statistical perspective, the effect of 17 α-hydroxyprogesterone, caproate injection on preterm births has not been established by adequate and well-controlled clinical trials -- a requirement of Subpart H approval. Although Study 17P-CT002 demonstrated statistically significantly reductions in preterm deliveries, it is my position that the level of evidence from this single study is not sufficient to support the effectiveness of 17 α-hydroxyprogesterone, caproate injection and, therefore, does not support the requirements for Subpart H; see Statistical Review of NDA 21-945, dated 10/19/2006.

Assuming Subpart H approval, the applicant's Complete Response includes a draft protocol for a "Phase 4" study to demonstrate the effectiveness of 17 α-hydroxyprogesterone, caproate injection on preterm births (the surrogate used in Study 17P-CT002) and on neonatal outcomes (the clinical endpoint required for final Subpart H approval). The proposed study, with some modifications, could be employed to develop the evidence for an alternative path to Subpart H approval. After a sufficient number of subjects have delivered, the results of the effect of 17 α-hydroxyprogesterone, caproate injection on preterm births could be submitted to us for review. If the data were sufficient to establish efficacy on preterm births, the application could be given Subpart H approval at that time. The ongoing study would need to be completed in order for us to review the effects on fetal and neonatal losses and on neonatal morbidity – the clinical endpoints of interest. If efficacy for the clinical endpoints were established, the drug could be given standard approval.

The practical limitation to this approach is the amount of savings in time between the submission of a study report of the effect on the surrogate (i.e., preterm births) and the submission of the final study report of the effects on the clinical endpoints of interest may not be sufficient to justify an early submission.

If the medical division does go forward with Subpart H approval for this Complete Response, I question whether a placebo-controlled study can be conducted in the United States if the drug product is approved for the indication under study. Also, the study is likely underpowered to rule out a difference of 2.5% between drug and placebo in the rate of fetal losses. Moreover, the primary endpoint needs to be changed from a surrogate endpoint to a clinical endpoint.

2. INTRODUCTION

2.1 Overview of Complete Response

This submission is a Complete Response to the Approvable Letter for NDA 21-945 (dated 10/20/06); see Appendix 1. The medical division's Approvable Letter raised the possibility for approval under Subpart H, 21 CFR 314.510. As a result, the complete response includes introductory promotional materials as requested in the Approvable Letter. Further, the response includes two draft protocols for a Phase 4 confirmatory study and a follow-up study of children aged 18 to 24 months, whose mothers received 17P[1].

My review of the Applicant's Complete Response focuses on these areas:

- Approval under Subpart H 21 CFR 314.510
- Draft protocol for a Phase 4 study
- Draft protocol for a follow-up study of children

2.2 Data Sources

Approvable Letter for NDA 21-945, dated 10/20/2006

Statistical Review of NDA 21-945, dated 10/19/2006

21 CFR 314.510 and 21 CFR 314.500

Transcripts from Reproductive Health Drugs Advisory Committee Meeting held on 8/29/2006

[1] The planned marketed drug product is 17 α-Hydroxyprogesterone; Caproate Injection, 250 mg/mL. This drug product is abbreviated as 17P throughout the NDA, Complete Response and this review.

3. STATISTICAL EVALUATION

3.1 Approval under Subpart H, 21 CFR 314.510

From my perspective, the complete response does not fulfill the requirements for approval under Subpart H, 21 CFR 314.510. The study submitted with the original NDA had several flaws, which did not allow the study to establish the efficacy of 17P on the surrogate endpoint. The complete response does not contain any that demonstrate an effect of 17P on the surrogate endpoint of preterm births.

To facilitate my discussion, I have reproduced Subpart H, Sections 314.500 and 314.510:

> **Subpart H – Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses**
>
> **21 CFR 314.500 Scope.**
>
> This subpart applies to certain new drug products that have been studied for their safety and effectiveness in treating serious or life-threatening illnesses and that provide meaningful therapeutic benefit to patients over existing treatment (e.g., ability to treat patients unresponsive to, or intolerant of, available therapy, or improved patient response over available therapy).
>
> **21 CFR 314.510 Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity.**
>
> FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity. Approval under this section will be subject to the requirement that the applicant study the drug further, to verify and describe its clinical benefit, where there is uncertainty as to the relation of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome. Post-marketing studies would usually be studies already underway. When required to be conducted, such studies must also be adequate and well-controlled. The applicant shall carry out any such studies with due diligence.

3.1.1 Surrogate endpoint

The consequences of preterm birth include significant neonatal morbidities and mortality. Moreover, children who are born prematurely are at higher risk for developmental and other delays. Therefore, the prevention of neonatal morbidity, mortality and development delays associated with prematurity is an important public health issue.

For this drug product, preterm delivery is the surrogate endpoint of interest. From a clinical perspective, this surrogate is reasonably likely to predict fetal and neonatal losses, neonatal morbidity and subsequent developmental delays.

The definition of a preterm delivery, when used as a surrogate endpoint, is not as clear. For that reason, the approvable letter suggests using deliveries prior to 32 weeks gestation and deliveries prior to 35 weeks gestation.

Preterm deliveries defined as deliveries prior to 32 weeks gestation and prior to 35 weeks gestation appear to meet the Subpart H requirements of a "surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit."

3.1.2 Evidence for an effect on preterm births (i.e., surrogate endpoint)

Under Subpart H, a drug product may be approved if adequate and well-controlled clinical trials establish the drug product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit. The effect was not established by Study 17P-CT002, which was submitted to the original NDA. That study was insufficient to support approval of 17P for the prevention of preterm births – the surrogate endpoint of interest; see my statistical review dated 10/19/2006. Moreover, the complete response does not include any new data from clinical trials that investigated the effect of 17P on preterm births.

My reasons for concluding that the original submission does not support the effectiveness of 17P in preventing preterm deliveries are:

- Reliance on a single study (17P-CT002), which did not yield the level of evidence needed for approval based on a single study.
- Optimal time to start study drug was not identified.
    - 17P appeared most effective when started at 18 weeks of gestation or earlier; did not appear effective when started at 20 weeks of gestation or later.
    - Rate of fetal and neonatal deaths is most pronounced among births to women who started 17P at 18 weeks gestation or earlier (10%).
- Apparent confounding of study site and gestational age at randomization.
    - One center accounted for 44% of subjects enrolled at 18 weeks of gestation or earlier.
    - Some centers had a deficit of subjects enrolled at 18 weeks of gestation or earlier.
- Fetal and neonatal deaths among women treated with 17P occur earlier than among women treated with placebo.

At the Advisory Committee meeting on 8/29/2006, Dr. Meis, the principal investigator of Study 17P-CT002, discussed the rationale for when to start study drug. He indicated that some trials of progesterone that did not show efficacy started drug relatively late in gestation. So a decision was made to start treatment earlier in this trial. They waited until 16 weeks to reduce the possibility of teratogenic effect; study treatment was not started after 21 weeks because the investigators felt there would be no efficacy after 21 weeks. His comments were consistent with my findings from the post-hoc analysis of the relationship between time of gestation and study outcome.

The current submission does not contain any new information or data to obviate those concerns. From a statistical perspective, the study did not meet the level of evidence needed to support the efficacy of 17P in the prevention of preterm births.

3.1.3 Appropriateness of Subpart H, Section 314.510 for this drug product

3.1.3.1 Short history of Subpart H

The prototype for Subpart H was the conditional approval in 1991 of didanosine for the treatment of HIV-infected individuals. Prior to didanosine, AZT was the only approved product for the treatment of HIV. Because of the AIDS public health crisis, additional treatment options were desperately needed. FDA sought ways to make promising drugs available as soon as possible for patients infected with HIV.

A commonly accepted endpoint for HIV trials was time to a new AIDS-defining event or death. Often, many years were needed to establish efficacy for this clinical endpoint. Those involved with the design of clinical studies explored the use of surrogate endpoints in order to dramatically reduce the time needed to approve promising drug products for the treatment of HIV. The assumption was that a treatment-induced change in a surrogate would translate into a clinical benefit. This assumption, however, is not always simple to evaluate.

Nonetheless, researchers were willing to assume that an improvement in CD4 counts would translate into a clinical benefit. In studies of AZT, changes in an individual's CD4 counts could be seen after several weeks of treatment. Moreover, clinical studies had shown that AZT, compared with placebo, reduced mortality. The thinking was that if similar changes in CD4 counts were seen in subjects treated with didanosine, then the changes would lead to a clinical benefit.

ACTG 116, conducted by NIH, was a double-blind study that compared two doses of didanosine with AZT in HIV-infected subjects who were diagnosed with AIDS or who had CD4 counts less than 300 at the time of study entry. The primary endpoint was time to a new AIDS-defining event or death. The study enrolled approximately 1000 subjects.

For this study, FDA accepted CD4 counts within the first six months of randomization as a surrogate for clinical efficacy. A look at data from an ongoing clinical study was unprecedented. Analyses of CD4 counts in ACTG 116 clearly showed drug-induced changes in CD4 counts. Analyses of these changes supported the conditional approval of didanosine. The sponsor (Bristol-Myers-Squibb) together with NIH was required to complete the on-going study and to analyze the clinical endpoint of interest in order to gain full approval. The results, presented at an advisory committee meeting in 1992, established the efficacy of the low dose of didanosine but not the higher dose.

3.1.3.2 Subpart H and 17P

Study 17P-CT-002, the subject of the original NDA, was designed to detect differences between treatment and placebo in the incidence of preterm birth. The medical division believes this to be an acceptable surrogate endpoint, likely to predict fetal and neonatal losses and neonatal morbidity. Fetal and neonatal losses and morbidity are an important public health issue for which there are no approved products. Because the study was not powered to detect difference in these clinical endpoints, the division was willing to accept differences in the surrogate endpoint, preterm births, as the basis for approval.

In my statistical review of 17P-CT-002, I concluded the level of evidence from Study 17P-CT002 was not sufficient to support the effectiveness of 17P. I stated the need for a second study of the effect of 17P on preterm births. Although this complete response provides a draft protocol for a second study, the response does not provide any additional data that establishes the efficacy of 17P preterm births. Therefore, the complete response has not convinced me of the efficacy of 17P in preventing preterm births. The study of the surrogate endpoint will need to be completed and reviewed in order for me to address whether this 2nd study has demonstrated an effect on the endpoint.

In my mind the question then becomes, is there a situation where 17P might be appropriate for Subpart H approval? Unlike studies of HIV and cancer where the difference between subject-level evaluation of a surrogate and clinical endpoint can be years, the clinical endpoint (neonatal mortality and morbidity) is known within weeks at most. From that perspective, 17P is not an appropriate candidate for Subpart H.

I can conceive of one approach to gaining Subpart H approval. Because the incidence of neonatal mortality and morbidity is much less than that of preterm births, one approach would be to power a study for neonatal outcomes. Using that sample size, one can estimate the power needed to detect a difference in preterm births. This is what the draft protocol for a Phase 4 study is proposing. Possibly an analysis of preterm births could be done when enough subjects have accrued to detect with 80% power a difference between treatment arms in the incidence of preterm births. If the study shows a difference in preterm births, approval could be granted under Subpart H. The study would need to continue to completion and the neonatal outcomes analyzed.

Although this approach to Subpart H approval is theoretically possible, it may not be practical. By the time the database is locked, data are analyzed, study reports are written and submitted to FDA for review, it is possible the study would be complete. In such a situation, the savings in time for a full approval may not be significant.

3.2 Draft protocol of a Phase 4 study

3.2.1 Summary of protocol

The submission includes a draft study protocol entitled, "A Phase 4, multi-center, randomized, double-blind study of 17 α-hydroxyprogesterone caproate (17P) versus placebo for the prevention of preterm birth in women with a previous singleton

spontaneous preterm delivery." This study represents a confirmatory study of the findings from Study 17P-CT-002.

The study objective is to determine if treatment with 17P reduces the rate of preterm birth $<35^0$ weeks of gestation in women with a previous singleton spontaneous preterm delivery. The study will enroll a total of 1230 women (820 17P and 410 placebo) with a singleton pregnancy. Subjects will receive weekly injections of study drug from randomization (16^0 through 20^6 weeks of gestation) until 36^6 weeks of gestation or delivery, whichever occurs first. Subjects will be followed up to around 30 days after the last dose of study drug of discharge from the delivery hospitalization, whichever occurs later. Neonates will be followed until discharge from the birth hospitalization or 120 days after birth, whichever occurs first.

Secondary endpoints are preterm birth prior to 32^0 weeks of gestation, earl fetal loss, and a composite neonatal morbidity and mortality index. Although the protocol indicates the index includes neonatal death, IVH, RDS, BPH, NEC and proven sepsis, the index is not defined.

Subjects will randomized in a 2:1 ratio to 17P or placebo using a blocked randomization stratified by study site. A sample size of 1230 subjects (820 17P and 410 placebo) yields 90% power at a Type I error rate (two-sided) of 5% to detect a reduction in the rate of preterm births (delivery $<35^0$ weeks of gestation) from 30% to 21.4%, and 82% power to detect a reduction in the rate of preterm births (delivery $<32^0$ weeks of gestation) from 20% to 14%. The protocol also indicates this sample size has 80% power to detect a reduction in the Neonatal Composite Index from 17% to 11%.

The study is designed also to show non-inferiority for early fetal losses. According to the protocol, assuming a 1.5% early fetal loss rate in both treatment groups with a one-sided alpha of 2.5%, a sample size of 1230 subjects provides 92% power to show non-inferiority of the fetal loss rate in the placebo and 17P groups with a margin of 2.5%. In Study 17P-CT-002, the 17P group had a fetal loss rate of 3.6%; the placebo group had a fetal loss rate of 1.3%.

3.2.2 Statistical Comments

3.2.2.1 Primary endpoint

I recommend changing the primary endpoint from a surrogate endpoint (preterm birth) to a clinical endpoint of interest, which appears to be the composite index. If this study is to be used in the future as the basis for Subpart H approval, the currently specified primary endpoint (preterm birth) could be evaluated at an interim analysis. If those results show efficacy, the study could be continued to its conclusion at which time the clinical endpoint of interest could be evaluated to determine the efficacy of 17P in preventing fetal and neonatal losses and neonatal morbidity.

The protocol does not describe the composite neonatal morbidity and mortality index, other than to say that it includes neonatal death, Grade 3 or 4 IVH, RDS, BPH, NEC and proven sepsis. Based on Study 17P-CT-002, the index yields a binary outcome; presumably, a "yes" if any of the components is present. The protocol needs to describe the index in detail. This information is especially important if the applicant decides to use the index as the primary endpoint in the study.

3.2.2.2 Analyses

The analysis of preterm births needs to account for time on study drug, since women enter and start study treatment at staggered times relative to gestation. One way to accomplish this analysis is to construct Kaplan-Meier estimates of the rates of preterm births that are adjusted for left-censoring. These estimates and their standard errors can be compared between treatment groups.

Analyses of fetal losses and neonatal deaths and other safety outcomes need to be adjusted for time on study drug. A Kaplan-Meier approach can be used here as well.

3.2.2.3 Sample size calculations

The assumptions for the sample size calculation for non-inferiority are incorrect.
The protocol gives the following rationale for the selection of 1.5% as the common rate for losses prior to 20 weeks gestation for both 17P and placebo:

"An early fetal loss of 1.5% and a non-inferiority margin of 2.5% were chosen based on the results of Study 17P-CT-002 (the NICHD 17P trial). In that study the 17P group had a higher, but not statistically significant, rate of fetal loss (17P 3.6% vs placebo 1.3%; $p>0.05$). …"

These are inaccurate estimates of fetal losses. They are crude estimates and do not account for staggered entry into the study. For example, women who entered during week 20 of gestation would not have been eligible for fetal loss at earlier gestation times, resulting in a denominator that is too large for the calculation of the crude rate.. My review of Study 17P-CT-002 shows the following (this information is also in the Advisory Committee transcripts):

Estimated Rates of Fetal and Neonatal Deaths, accounting for time on study drug.

Week of Gestation	17P %	Placebo %
16	0.0%	0.0%
17	0.0%	0.0%
18	0.0%	0.0%
19	2.3%	0.0%
20	3.5%	0.0%
21	6.3%	0.8%
22	6.6%	0.8%
23	7.2%	1.4%
24	7.2%	3.3%

Source: Statistical Review of NDA 21-94, Table 3.2, Estimated rates of fetal and neonatal Deaths, accounting for time on study drug.

For example, at 20 weeks gestation, the estimated rates of fetal loss are 3.5% for 17P and 0% for placebo; at 24 weeks gestation the estimated rates are 7.2% for 17P and 3.3% for placebo.

Even if the sample size calculations assume a common fetal loss rate of 1.5% at 20 weeks gestation, the study may be underpowered to show non-inferiority if the rate of losses for 17P is greater than the rate of losses for placebo, as suggested by Study 17P-CT-002. The applicant should reconsider power and sample sizes for scenarios where the true rate of losses for women receiving 17P is greater than the rate of losses for women receiving placebo. For example, if the true rate for fetal losses is 1.75% among women receiving 17P, the proposed sample size will have about 87% power to rule out a difference of more than 2.5%; if the true rate is 2.0%, the power decreases to 76%.

The medical reviewer indicates Week 24 is a more appropriate cutoff for defining early losses. At Week 24, the estimated rates of losses are 7.2% for the 17P treatment group and 3.3% for placebo. The best case scenario assumes a common rate of losses of 3% for each group. With the planned sample size, the study will have 72% power to rule out a

difference of 2.5%. If the true rate for 17P is greater than that for the placebo, the power will be even less than 72%.

3.2.2.4 Withdrawal from study

The protocol needs to make a distinction between a subject withdrawn from the study and a subject withdrawn from treatment. Subjects should be withdrawn from treatment for reasons of withdrawal from consent or for safety only. Non-compliance is not a sufficient reason. All subjects, whether they are receiving treatment or not, should remain in the study and receive all study visits and evaluations as specified in the protocol.

3.2.2.5 Other comments

The data monitoring committee charter should be submitted for review.

The protocol needs to include a copy of the informed consent that will be given to patients.

If 17P is approved under Subpart H with the commitment that this Phase 4 study will be conducted, study enrollment may be difficult. With the drug product approved, women may be reluctant to enroll in a placebo-controlled study.

3.3 Draft protocol of a follow-up study of children born to mothers who received 17P or placebo in the Phase 4 study.

3.3.1 Summary of protocol

The submission includes a draft study protocol entitled, "A prospective, noninterventional follow-up study of children aged 18 to 24 months born to mothers who received 17 α-hydroxyprogesterone caproate (17P) or placebo in the Phase 4 17P efficacy trial." The study objective is to determine whether there is a difference in the achievement of developmental milestones between children whose mothers received 17P and those who received placebo in the Phase 4 study discussed above.

Informed consent will be obtained from the subject's mother/legal guardian between delivery and discharge from the delivery hospitalization. Mothers/legal guardians will be contacted periodically until the child nears the age of 18 months. If the mother/legal guardian is interested in continuing in the study, the Ages and Stages Questionnaire (ASQ) will be mailed in order to screen the child for a developmental delay.

If the questionnaire suggests a delay as measured by falling below a specified cutoff in at least 1 developmental area on the ASQ, the child will be referred for follow-up assessments. If more than one area is identified, secondary assessments will be done. Depending on the developmental area idenitified, the assessment may be the Bayley

Scales of Infant Development, Modified Checklist for Autism in Toddlers, a neurological exam or the Gross Motor Function Classification System.

The primary outcome is the proportion of children who fall below the specified cutoff for at least one of the developmental areas assessed by the ASQ. Differences between treatment groups will be compared with a chi-square test. Secondary analyses will consider each of the five domains individually.

Assuming a completed ASQ is obtained for 375 children (250 17P and 125 placebo), the study will have 80% power to detect a difference of 15%, using an overall Type I error rate of 5% and assuming an outcome rate of 30% in the 17P group.

3.4 Promotional materials

The promotional materials fail to highlight the higher rate of fetal losses seen in the women who received 17P as compared with women who received placebo.

4. Conclusions and Recommendations

In this Complete Response to the Approvable Letter for NDA 21-945, the Applicant is seeking approval of 17 α-hydroxyprogesterone, caproate injection through Subpart H, Section 510 of the CFR. Under Subpart H, a drug product may be approved if adequate and well-controlled clinical trials establish the drug product has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit.

The original NDA, which was a single study submission, is being used to support the efficacy of 17 α-hydroxyprogesterone, caproate injection in reducing preterm deliveries. The Complete Response does not contain any additional efficacy data.

From a clinical perspective, preterm delivery is reasonably likely to predict fetal and neonatal losses and neonatal morbidity. Thus, the use of preterm births as a surrogate endpoint appears to meet one of the requirements of Subpart H.

Study 17P-CT002, which was the singly study included in the original NDA, showed statistically significant reductions in preterm deliveries at <35 weeks and at <32 weeks. The medical team concluded these results were sufficient to support the efficacy of 17 α-hydroxyprogesterone, caproate injection.

However, from a statistical perspective, the effect of 17 α-hydroxyprogesterone, caproate injection on preterm births has not been established by adequate and well-controlled clinical trials -- a requirement of Subpart H approval. Although Study 17P-CT002 demonstrated statistically significantly reductions in preterm deliveries, it is my position that the level of evidence from this single study is not sufficient to support the effectiveness of 17 α-hydroxyprogesterone, caproate injection and, therefore, does not support the requirements for Subpart H; see Statistical Review of NDA 21-945, dated 10/19/2006.

Assuming Subpart H approval, the applicant's Complete Response includes a draft protocol for a "Phase 4" study to demonstrate the effectiveness of 17 α-hydroxyprogesterone, caproate injection on preterm births (the surrogate used in Study 17P-CT002) and on neonatal outcomes (the clinical endpoint required for final Subpart H approval). The proposed study, with some modifications, could be employed to develop the evidence for an alternative path to Subpart H approval. After a sufficient number of subjects have delivered, the results of the effect of 17 α-hydroxyprogesterone, caproate injection on preterm births could be submitted to us for review. If the data were sufficient to establish efficacy on preterm births, the application could be given Subpart H approval at that time. The ongoing study would need to be completed in order for us to review the effects on fetal and neonatal losses and on neonatal morbidity – the clinical endpoints of interest. If efficacy for the clinical endpoints were established, the drug could be given standard approval.

The practical limitation is the amount of savings in time between the submission of a clinical study report of the effect on preterm births and the submission of the final clinical study report of the effects on the clinical endpoints of interest may not be sufficient to justify an early submission.

If the medical division does go forward with Subpart H approval for this Complete Response, I question whether a placebo-controlled study can be conducted in the United States if the drug product is approved for the indication under study.

For the proposed Phase 4 study, I recommend changing the primary endpoint from a surrogate endpoint (preterm birth) to a clinical endpoint of interest, which appears to be the composite index. If this study is to be used in the future as the basis for Subpart H approval, the currently specified primary endpoint (preterm birth) could be evaluated at an interim analysis. If those results show efficacy, the study could be continued to its conclusion at which time the clinical endpoint of interest could be evaluated to determine the efficacy of 17P in preventing fetal and neonatal losses and neonatal morbidity.

7 Pages of Appendix 1 that contains the 10/20/06 Approvable Letter has been removed as a duplicate copy of the original Approvable Letter that is located in the "Other Action Letters" section of the redacted Approval Package.

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

/s/

Lisa A. Kammerman
1/23/2009 01:58:10 PM
BIOMETRICS

Mahboob Sobhan
1/23/2009 02:27:16 PM
BIOMETRICS

Steve Wilson
1/23/2009 02:48:41 PM
BIOMETRICS