

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*  
**21945Orig1s000**

**STATISTICAL REVIEW(S)**



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

## Statistical Review and Evaluation

### CLINICAL STUDIES

NDA/Serial Number: 21-945

Drug Name: Makena (formerly Gestiva; 17  $\alpha$ -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

<b>1. EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>2. INTRODUCTION .....</b>	<b>8</b>
2.1 Overview of Complete Response.....	8
2.2 Data Sources .....	8
<b>3. STATISTICAL EVALUATION .....</b>	<b>9</b>
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
<b>4. CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>37</b>

## 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 <sup>0</sup> Weeks Gestation.....	19
Table 5. Delivery <37 <sup>0</sup> Weeks, <35 <sup>0</sup> Weeks, <32 <sup>0</sup> Weeks, <28 <sup>0</sup> 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 <sup>0</sup> Weeks, by gestational age at randomization .....	21
Table 7. Proportion of subjects who delivered <37 <sup>0</sup> 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 Deaths, 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  $\alpha$ -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.”<sup>1</sup>

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.

---

<sup>1</sup> Available at [www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf](http://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 (b) (4)

[REDACTED]

[REDACTED] (b) (4)

[REDACTED] (b) (4)

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<sup>2</sup> 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

---

<sup>2</sup> The planned marketed drug product is 17  $\alpha$ -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."<sup>3</sup>

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

---

<sup>3</sup> Available at [www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf](http://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.

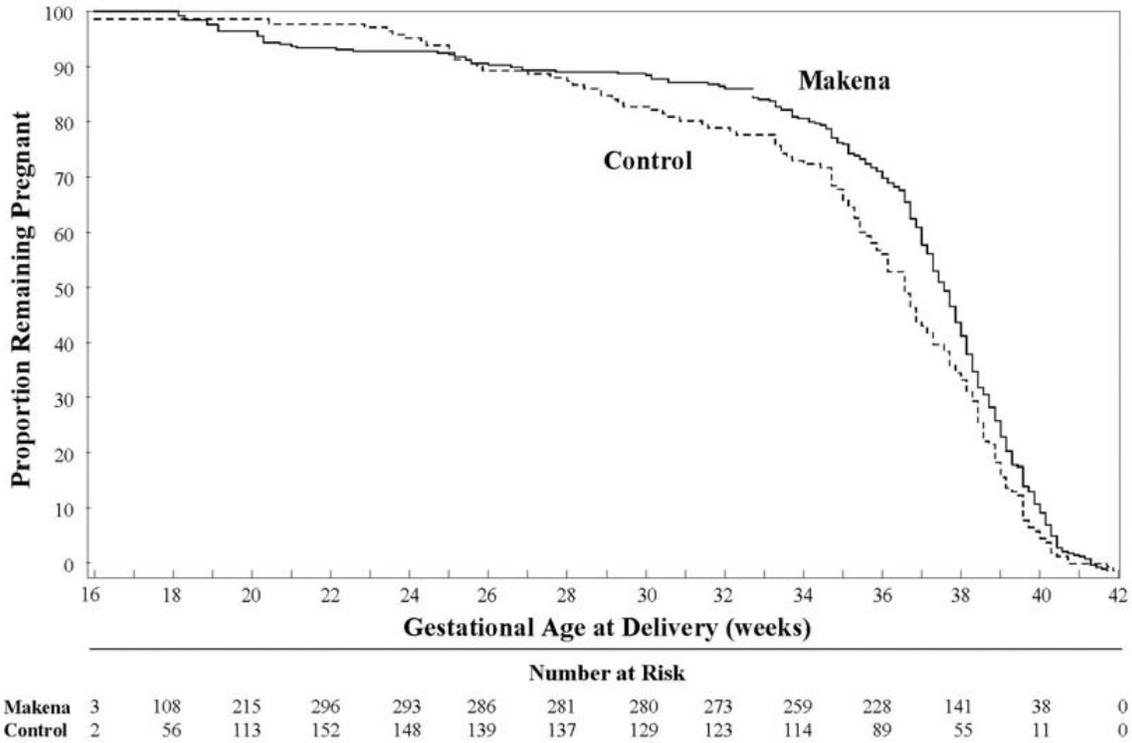
## **3.2 Evaluation of efficacy**

### **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**



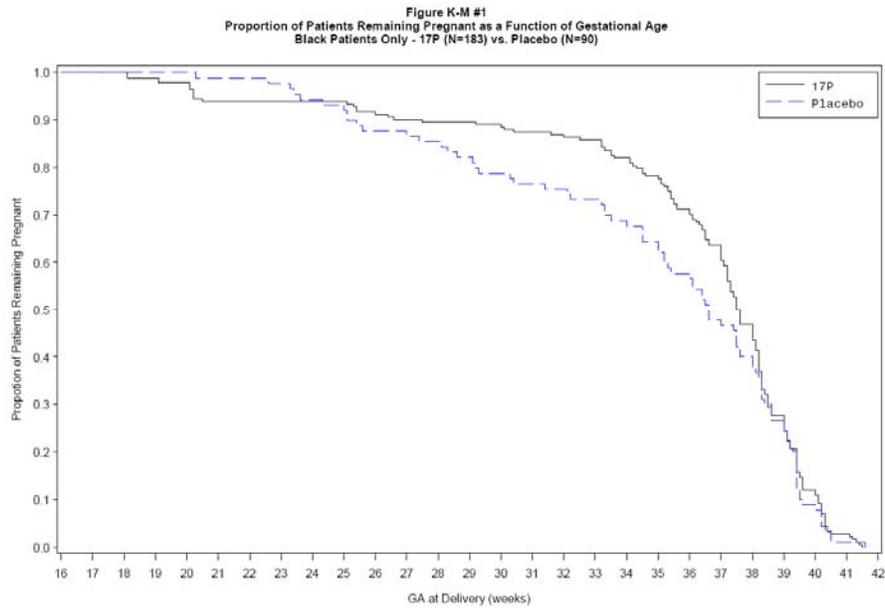
Source: (b) (4)

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).

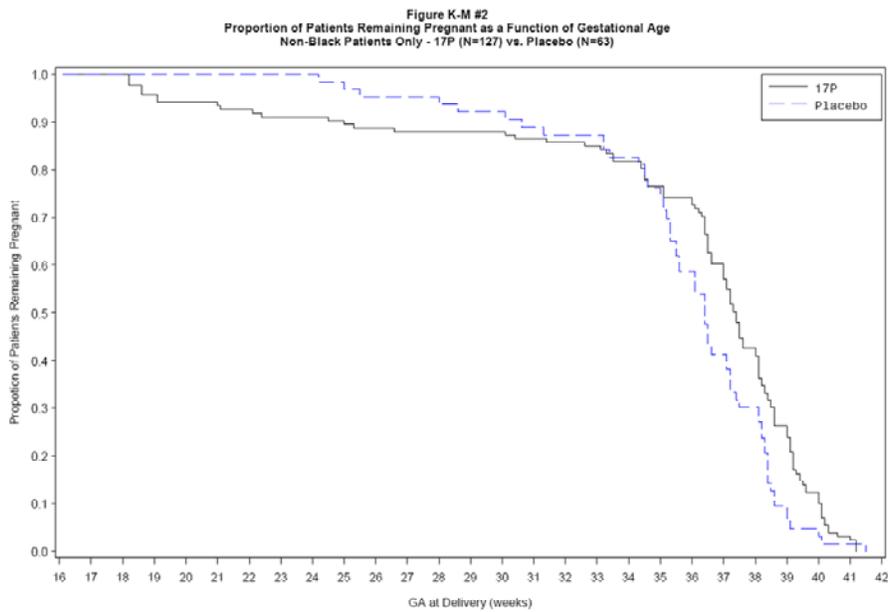
Although both racial subgroups mimic the pattern seen in the overall population, the differences between subjects who were black and non-black is notable. Overall, subjects who were black appeared to receive a greater benefit from 17P than did subjects who were non-black. Initially, in both subgroups, 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**

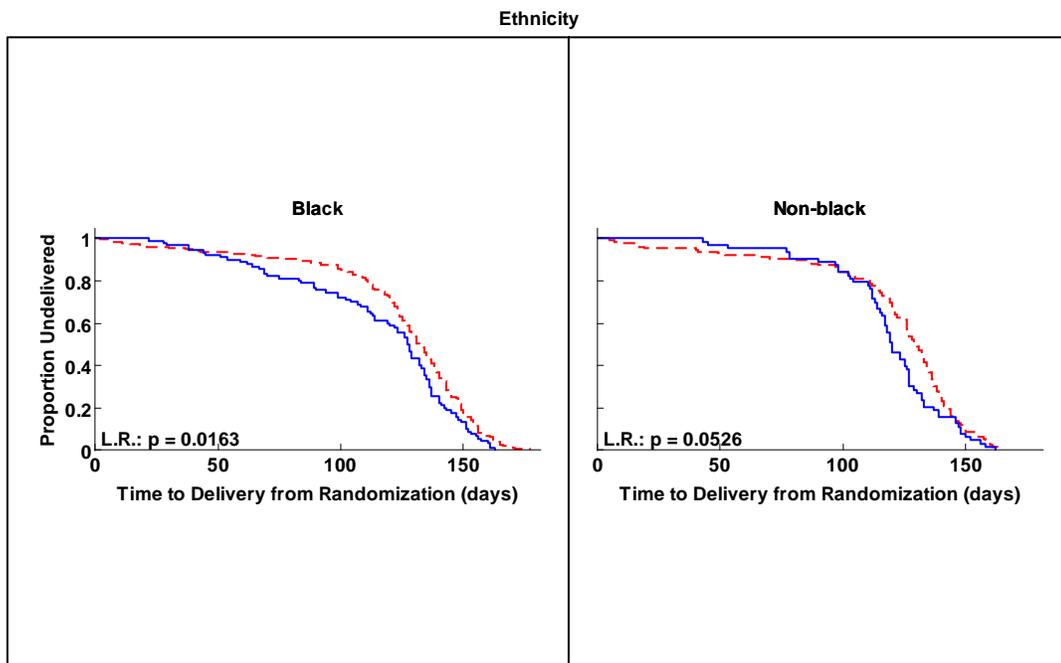


**Non-black**



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



TREAT:  
- - - 17P  
— PLACEBO

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**

<u>Center #</u>	<u>Number_of subjects enrolled at center</u>	<u>% of all subjects enrolled in study (N=463)</u>	<u>% of all subjects who were black (N=273)</u>	<u>% of all subjects who were non-black (N=190)</u>
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
<i>All subjects</i>	<i>463</i>	<i>100%</i>	<i>59%</i>	<i>41%</i>

*Source: Statistical Reviewer*

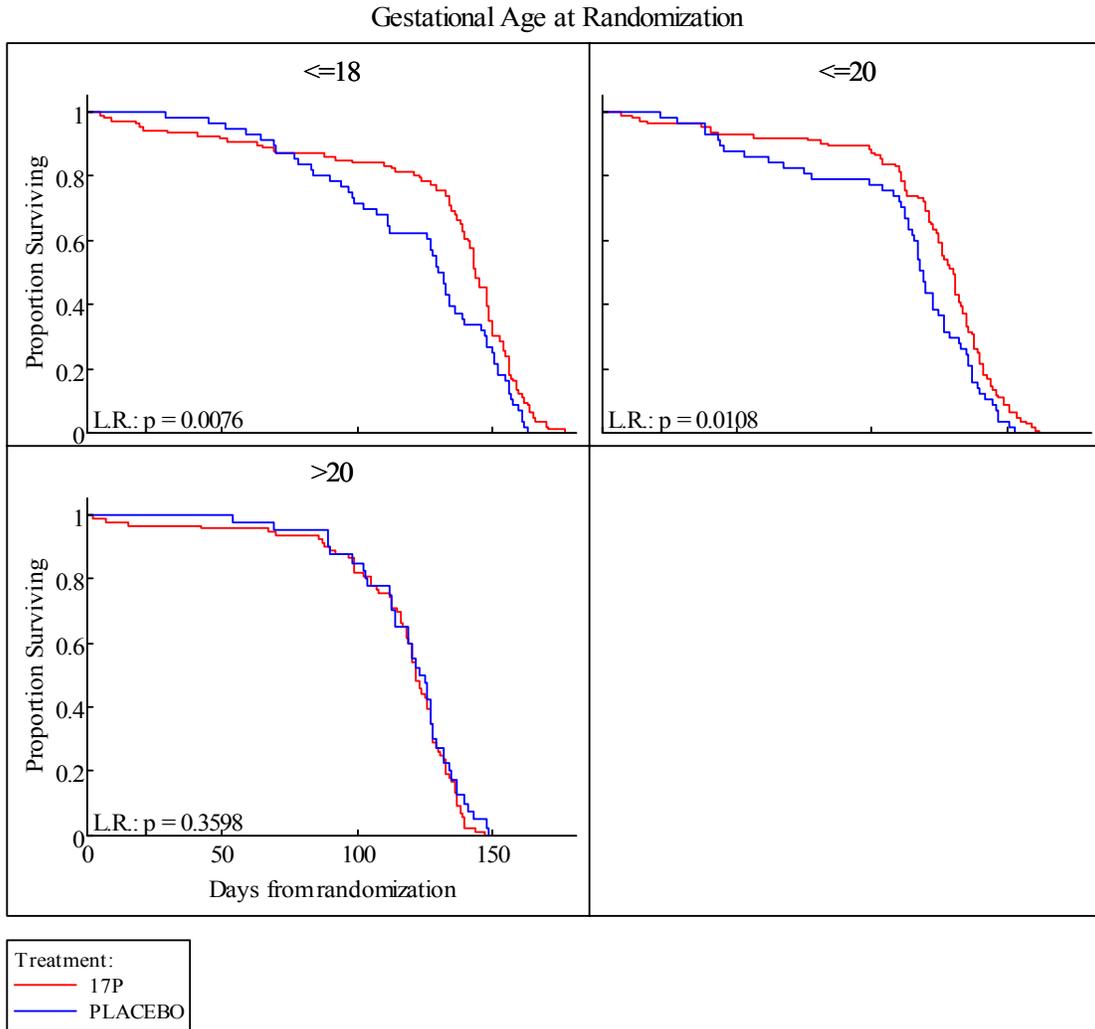
**Table 2. Distribution of blacks, by center. Entries are percentages of each center's enrollment.**

Center #	Number_of subjects enrolled at <u>center</u>	% of subjects at center who were <u>black</u>
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
<i>All subjects</i>	<i>463</i>	<i>59%</i>

*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 do 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**

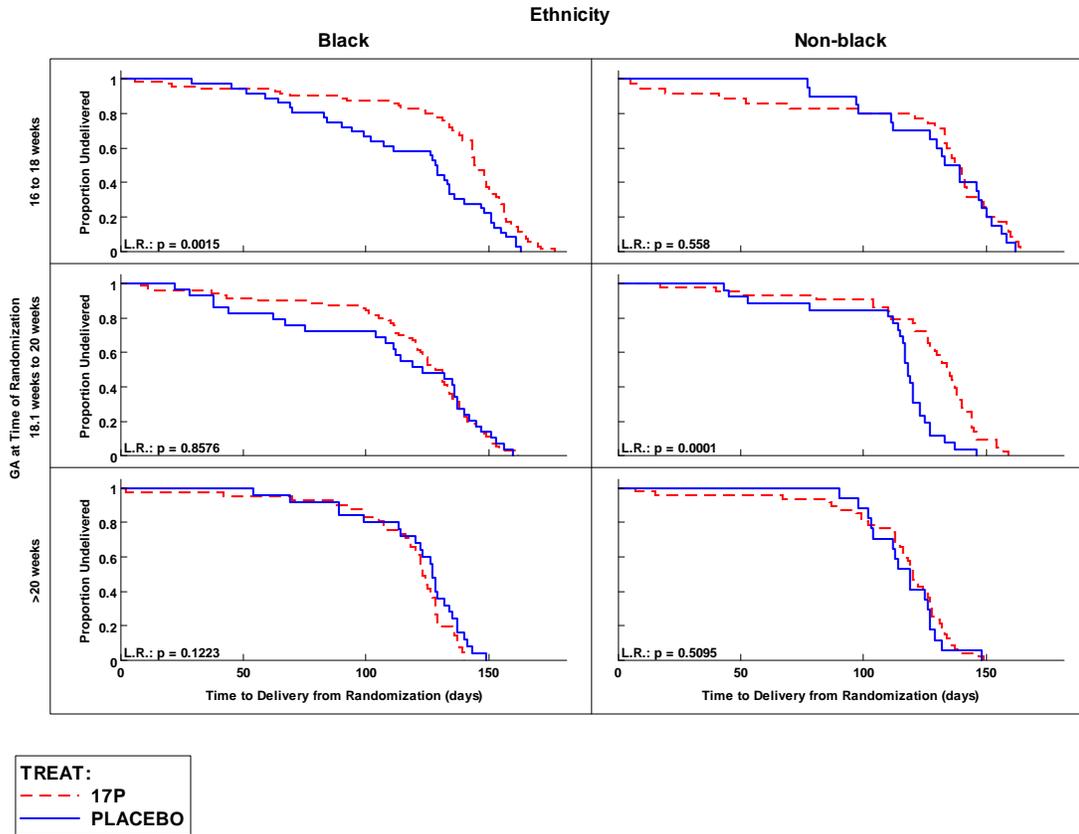
<u>Gestational Age at Randomization</u>	Total number of subjects	% Black	% Non-black
≤18 weeks	164	68%	32%
>18 and ≤ 20 weeks	170	58%	42%
>20 weeks	129	50%	50%
<i>All subjects</i>	<i>463</i>	<i>59%</i>	<i>41%</i>

*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<sup>0</sup> 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<sup>0</sup> 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<sup>0</sup> Weeks Gestation**

Data Source	17P		Placebo		Nominal P-value <sup>a</sup>	Treatment difference and its 96.6% Confidence Interval <sup>b</sup>
	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%]

<sup>a</sup> Chi-square test. To account for the interim analyses, the nominal p-values need to be compared to 0.0345.

<sup>b</sup> 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 be 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 why 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<sup>0</sup> Weeks, <35<sup>0</sup> Weeks, <32<sup>0</sup> Weeks, <28<sup>0</sup> Weeks Gestation, ITT population. The estimates of rates of delivery do not account for duration of drug exposure.**

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

<sup>a</sup> 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).

<sup>b</sup> 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. <sup>4</sup>

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

<sup>4</sup> 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<sup>0</sup> Weeks, by gestational age at randomization**

Gestational Age at Randomization	All Subjects		17P		Placebo		Treatment Difference
	N	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	
≤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<sup>0</sup> Weeks, by gestational age at randomization and race**

Race	Gestational Age at Randomization	All Subjects		17P		Placebo		Treatment Difference
		N	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	Number Randomized	% Delivered <37 weeks	
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-Meier 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.**

<u>Gestational Age at Randomization</u>	<u>Treatment</u>	<u>Total Randomized</u>	<u>% Delivered &lt;35 weeks</u>	<u>% Delivered 35 to 36.9 weeks</u>	<u>% Delivered ≥37 weeks</u>
≤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<sup>5</sup> (Table 11).

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

	<u>Gestational Age at Randomization</u>		
	<u>≤18 weeks</u>	<u>&gt;18 and ≤ 20 weeks</u>	<u>&gt;20 weeks</u>
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).

<sup>5</sup> 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
<i>Number of deaths:</i>	17	10	3
<b>Center #</b>			
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.**

<u>Race</u>	<u>Total number of subjects</u>	<u>% Deaths</u>	<u>17P</u>		<u>Placebo</u>	
			<u>Total number of subjects</u>	<u>% Deaths</u>	<u>Total number of subjects</u>	<u>% Deaths</u>
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)
<i>All subjects</i>	<i>459</i>	<i>6.5</i>	<i>306</i>	<i>6.2</i>	<i>153</i>	<i>7.2</i>

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%<sup>6</sup>. 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.**

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

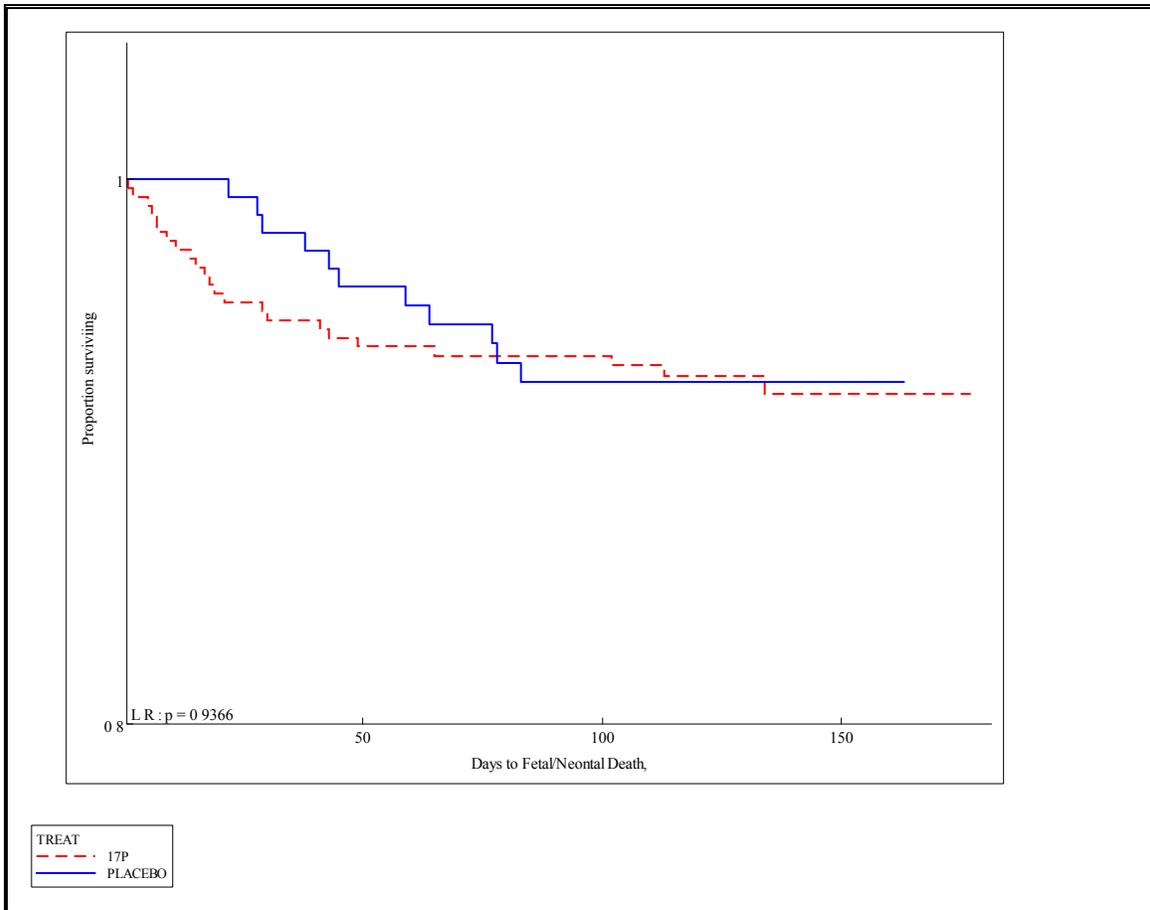
\* Entries in **bold** are estimates of the # at risk.

*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%.

<sup>6</sup> The first delivery that was not a fetal loss or neonatal death occurred just after 24 weeks gestation.

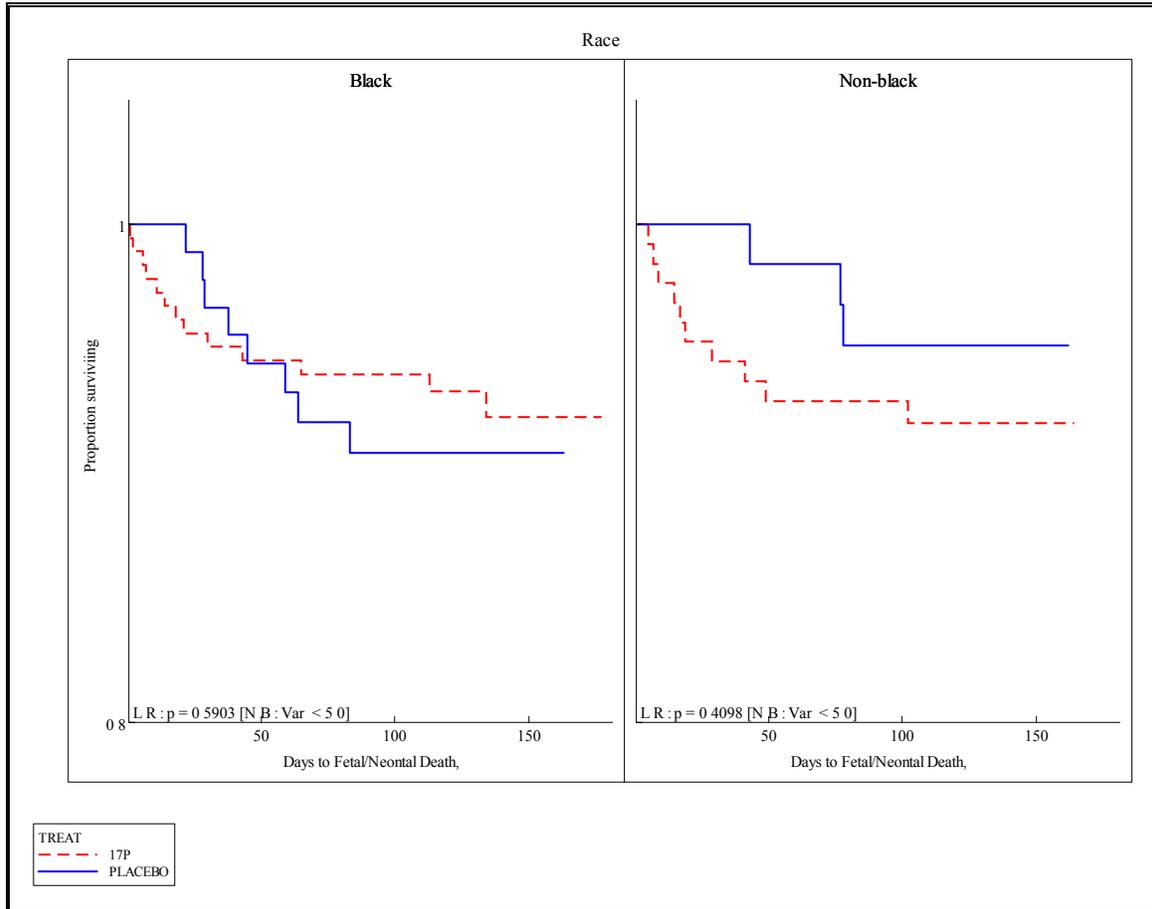
**Figure 6. From time of randomization to fetal and neonatal deaths, by treatment group.**



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.**

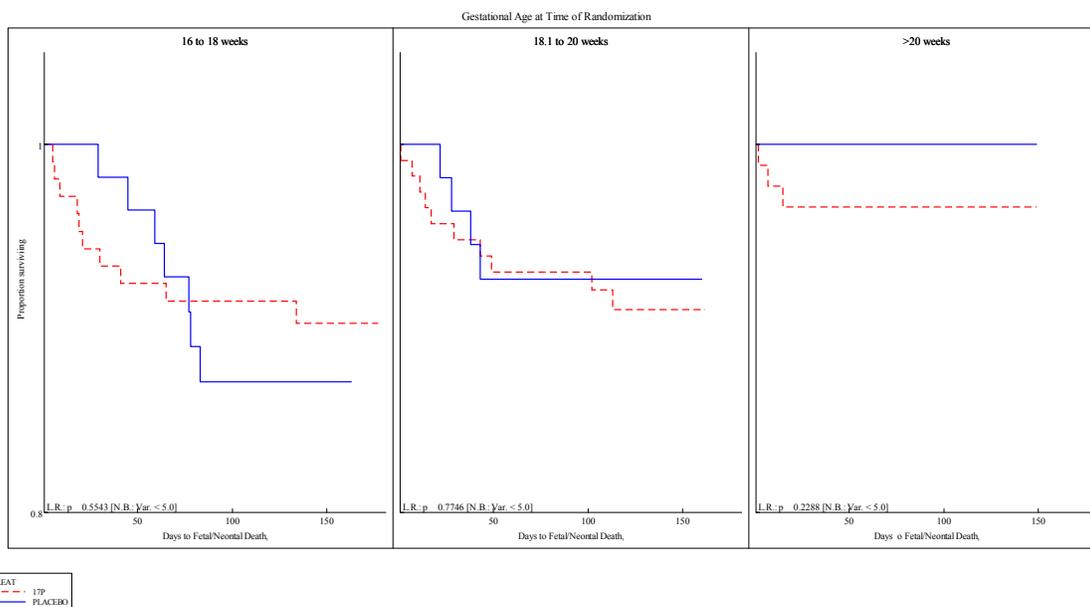


Note: y-axis starts at 80%

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”<sup>7</sup>. The guidance document states:

<sup>7</sup> Available at [www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075059.pdf](http://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)”, (b) (4)

Also affected is the text, (b) (4)

(b) (4)

If these changes are not adopted, I am concerned the labeling will (b) (4)

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 Deaths, sorted by Treatment, Race and Gestational age at delivery**

<u>Treatment</u>	<u>Race</u>	<u>Case</u>	<u>Center ID</u>	<u>Patient ID</u>	<u>Gestational age at delivery (weeks)</u>	<u>Gestational age at Randomization (weeks.days)</u>	<u>Classification of death</u>
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 is 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<sup>8</sup>. 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

---

<sup>8</sup> 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  $\alpha$ -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.”<sup>9</sup>

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:

---

<sup>9</sup> Available at [www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf](http://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 (b) (4)

[REDACTED]

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**

22 Pages Have Been Withheld As A Duplicate Copy Of The Statistical Review Dated 1/23/09  
Which Is Already Located In This Section

Appears This Way On Original

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

22 Pages Have Been Withheld In Full As A Duplicate Copy Of The Statistical Review  
Dated 10/19/06 Which Is Already Located In This Section

Appears This Way On Original

-----  
**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  
02/03/2011

STEPHEN E WILSON  
02/03/2011



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  $\alpha$ -hydroxyprogesterone; caproate injection)  
Indication(s): Prevention of preterm birth in pregnant women with a history of at least one spontaneous preterm birth  
Applicant: Cytoc Corporation  
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

<b>1. EXECUTIVE SUMMARY</b>	<b>3</b>
<b>2. INTRODUCTION</b>	<b>4</b>
2.1 OVERVIEW OF COMPLETE RESPONSE	4
2.2 DATA SOURCES	4
<b>3. STATISTICAL EVALUATION</b>	<b>5</b>
3.1 APPROVAL UNDER SUBPART H, 21 CFR 314.510	5
3.1.1 <i>Surrogate endpoint</i>	5
3.1.2 <i>Evidence for an effect on preterm births (i.e., surrogate endpoint)</i>	6
3.1.3 <i>Appropriateness of Subpart H, Section 314.510 for this drug product</i>	7
3.2 DRAFT PROTOCOL OF A PHASE 4 STUDY	8
3.2.1 <i>Summary of protocol</i>	8
3.2.2 <i>Statistical Comments</i>	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 <i>Summary of protocol</i>	12
3.4 PROMOTIONAL MATERIALS	13
<b>4. CONCLUSIONS AND RECOMMENDATIONS</b>	<b>13</b>

## 1. EXECUTIVE SUMMARY

In this Complete Response to the Approvable Letter for NDA 21-945, the Applicant is seeking approval of 17  $\alpha$ -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  $\alpha$ -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  $\alpha$ -hydroxyprogesterone, caproate injection.

However, from a statistical perspective, the effect of 17  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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<sup>1</sup>.

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](#)

---

<sup>1</sup> The planned marketed drug product is 17  $\alpha$ -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 2<sup>nd</sup> 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  $\alpha$ -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<sup>0</sup> 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<sup>0</sup> through 20<sup>6</sup> weeks of gestation) until 36<sup>6</sup> weeks of gestation or delivery, whichever occurs first. Subjects will be followed up to around 30 days after the last dose of study drug or 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<sup>0</sup> weeks of gestation, early 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 be 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<sup>0</sup> weeks of gestation) from 30% to 21.4%, and 82% power to detect a reduction in the rate of preterm births (delivery <32<sup>0</sup> 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  $\alpha$ -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 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.

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  $\alpha$ -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  $\alpha$ -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  $\alpha$ -hydroxyprogesterone, caproate injection.

However, from a statistical perspective, the effect of 17  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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.

Appendix 1: Approvable Letter for NDA 21-945, dated 10/20/2006

(b) (4)



6 Pages Have Been Withheld As A Duplicate Copy Of The Approvable Letter Dated 10/20/06 Which Is Located in the "Other Action Letters" Section of this NDA

-----  
**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



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  $\alpha$ -hydroxyprogesterone; caproate injection)  
Indication(s): Prevention of preterm birth in pregnant women with a history of at least one spontaneous preterm birth  
Applicant: Cytoc Corporation  
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

**ADDENDUM TO STATISTICAL REVIEW OF COMPLETE  
RESPONSE TO THE APPROVAL LETTER FOR NDA 21-945**

Since completing my statistical review of the applicant's complete response to the approval letter for NDA 21-945, the medical division and I have had numerous discussions with the applicant regarding their draft study protocol:

“A Phase 4, multi-center, randomized, double-blind study of 17  $\alpha$ -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.

As a result of these discussions, the applicant submitted a revised protocol on 12/12/2008. I agree with changes made to the protocol and do not have any additional statistical comments.

-----  
**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  
1/14/2009 03:04:54 PM  
BIOMETRICS

Mahboob Sobhan  
1/14/2009 03:17:48 PM  
BIOMETRICS

Steve Wilson  
1/21/2009 12:48:25 PM  
BIOMETRICS



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

## Statistical Review and Evaluation

### CLINICAL STUDIES

NDA/Serial Number: 21-945  
Drug Name: Gestiva (17  $\alpha$ -hydroxyprogesterone; caproate injection)  
Indication(s): Prevention of preterm birth in pregnant women with a history of at least one spontaneous preterm birth  
Applicant: Adeza Biomedical  
Date(s): Date stamp: 4/20/06  
PDUFA date: 10/20/06  
Review Priority: Priority  
Biometrics Division: DB 3  
Statistical Reviewer: Lisa A. Kammerman, Ph.D.  
Concurring Reviewers: Not needed  
Medical Division: Reproductive and Urological Products  
Clinical Team: Barbara Wesley, MD  
Project Manager: Freshnie DeGuia

**Keywords:** clinical studies, NDA review, logrank test, censored observations, one study application, multi-center, post-hoc/prospective analyses, generalisability, treat.-by-center interaction, treat.-by-baseline interaction, survival analysis, interim analysis, Kaplan-Meier product limit, surrogate outcomes, standard of evidence.

## Table of Contents

<b>1. EXECUTIVE SUMMARY</b>	<b>4</b>
1.1 CONCLUSIONS AND RECOMMENDATIONS	4
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	6
<b>2. INTRODUCTION</b>	<b>6</b>
2.1 OVERVIEW OF STUDY 17P-CT002	6
2.1.1 <i>Study Design</i>	6
2.1.2 <i>Subject Disposition</i>	7
2.1.3 <i>Data Monitoring and Safety Committee Meetings</i>	10
2.1.4 <i>Results for the primary efficacy variable: Delivery &lt;37<sup>0</sup> weeks gestation</i>	11
2.2 DATA SOURCES	12
<b>3. STATISTICAL EVALUATION</b>	<b>13</b>
3.1 EVALUATION OF EFFICACY	13
3.1.1 <i>Prevention of Delivery &lt;28<sup>0</sup>, &lt;32<sup>0</sup>, &lt;35<sup>0</sup>, &lt;37<sup>0</sup> Weeks Gestation</i>	13
3.1.2 <i>Study Not Designed for Drug Approval</i>	14
3.1.3 <i>Interim Analyses and the Secondary Endpoints</i>	15
3.1.4 <i>Placebo Response</i>	16
3.1.5 <i>An Alternative View of the Data – Prolongation of Pregnancy (Delaying Time-to-delivery)</i>	16
3.1.6 <i>Fetal Loss</i>	20
3.1.7 <i>Summary Statistics: Mean and Median</i>	21
3.1.8 <i>A Single Study Submission</i>	23
3.2 SURROGATE OUTCOME	37
3.3 EVALUATION OF SAFETY	38
<b>4. SUMMARY AND CONCLUSIONS</b>	<b>38</b>
4.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	38
4.2 CONCLUSIONS AND RECOMMENDATIONS	38
<b>5. APPENDICES</b>	<b>38</b>
5.1 LISTINGS OF KAPLAN-MEIER ESTIMATES OF TIME TO DELIVERY, INCORPORATING GESTATIONAL AGE AT RANDOMIZATION AS A LEFT-CENSORED VARIABLE	41
5.2 PREVENTION OF DELIVERY <37 <sup>0</sup> , <35 <sup>0</sup> , <32 <sup>0</sup> WEEKS, BY CENTER	46
5.3 TIME-TO-DELIVERY, INCORPORATING GESTATIONAL AGE AT RANDOMIZATION AS A LEFT-CENSORED VARIABLE, BY CENTER	51
5.4 TIME-TO-DELIVERY FROM RANDOMIZATION, BY CENTER AND BY GESTATIONAL AGE AT RANDOMIZATION	61

## LIST OF TABLES

Table 2.1 Distribution of Subjects by Study Center, Sorted by Size of Center .....	8
Table 2.2 The four subjects who were lost to follow-up .....	10
Table 2.3 Distribution of gestational age (weeks) at the time of randomization .....	10
Table 2.4 Applicant's Analysis: Delivery <37 <sup>0</sup> Weeks Gestation .....	12
Table 3.1 Delivery <37 <sup>0</sup> Weeks, <35 <sup>0</sup> Weeks, <32 <sup>0</sup> Weeks, <28 <sup>0</sup> Weeks Gestation, ITT population. The estimates of rates of delivery do not account for duration of drug exposure. ....	14
Table 3.2 Estimated Rates of Fetal and Neonatal Deaths, accounting for time on study drug. ....	20
Table 3.3 Mean Gestational Age at Time of Delivery, by Treatment Group. Excludes four 17P-treated subjects who were lost to follow-up. ....	22
Table 3.4 Mean Time to Delivery, by Treatment Group. Based on Kaplan-Meier analysis. ....	22
Table 3.5 Delivery <37 weeks, <35 weeks, <32 weeks: University of Alabama versus All Other Centers .....	26
Table 3.6 Time to Delivery: Summary statistics for University of Alabama and All Other Centers Combined. [95% Confidence Intervals Adjusted for Interim Analyses]... ..	28
Table 3.7 Distribution of Fetal and Neonatal Deaths, by Gestational Age at Randomization .....	35
Table 3.8 Distribution of Fetal and Neonatal Deaths, by Center and Gestational Age at Randomization .....	35
Table 3.9 Distribution of Gestational Age at Randomization: University of Alabama versus All Other Centers Combined. ....	36
Table 3.10 Delivery <37 <sup>0</sup> Weeks, <35 <sup>0</sup> Weeks, <32 <sup>0</sup> Weeks Gestation, by Gestational Age at Randomization: University of Alabama versus All Other Centers .....	36
Table 3.11 Distribution of Gestational Age at Randomization, by Center .....	37

## LIST OF FIGURES

Figure 2-1 Subject Disposition .....	9
Figure 3-1 Days from randomization to delivery .....	18
Figure 3-2. Time to delivery as a function of gestational age, using staggered entry based on the gestational age at randomization. ....	19
Figure 3-3 Time to fetal or neonatal deaths, by treatment group.....	21
Figure 3-4. Time to Delivery: University of Alabama (n=126) vs. All Other Centers Combined (n=337).....	27
Figure 3-5 Time to Delivery: University of Alabama vs. All Other Centers Combined, by Treatment Group.....	29
Figure 3-6 Time-to-delivery, by Center.....	30
Figure 3-7 Cumulative distribution plot of gestational age at randomization. The horizontal lines denote the 33 <sup>rd</sup> and 67 <sup>th</sup> percentiles. ....	32
Figure 3-8 Time to delivery, by gestational age at randomization .....	33
Figure 3-9. Time to Delivery, by Gestational Age at Randomization: University of Alabama versus All Other Centers Combined.....	34

# 1. EXECUTIVE SUMMARY

## 1.1 Conclusions and Recommendations

From a statistical perspective, the level of evidence from Study 17P-CT002 is not sufficient to support the effectiveness of 17P. The primary reason is the absence of a second, confirmatory study. Without a second study, the generalizability of the study results to a larger population cannot be assessed.

This submission contains a single study to support the claim of effectiveness of 17P. Prior to Study 17P-CT002 another study was initiated but was halted due to drug product manufacturing issues. Because of its small size and issues regarding drug potency, I did not review that study.

Study 17P-CT002 was stopped after the second interim analysis, which showed that Delivery <37 weeks gestation had met the stopping rules in favor of 17P. Subsequently, analyses showed that Delivery < 35 weeks gestation and Delivery <32 weeks gestation were statistically significant when accounting for the interim analyses.

Study 17P-CT002 was not designed for drug approval. FDA and the applicant did not have the usual meetings and discussions regarding the choice of endpoint needed to establish efficacy in a regulatory environment. As a result, the primary endpoint for the study – Delivery <37 weeks gestation – is not what the FDA would have advised.

After the results of the study were published, the FDA and the applicant discussed the analyses that would be submitted as part of an NDA. The FDA requested analyses of Delivery <35 weeks gestation as the primary basis of approval. The Advisory Committee reiterated the clinical importance of this endpoint as a preferred surrogate for neonatal morbidity and mortality.

Although the results are statistically significant for Delivery < 35 weeks gestation and Delivery <32 weeks gestation when accounting for interim analyses, the confidence intervals for the treatment effects are not convincing when considering that only one study was submitted to support the claim of effectiveness for 17P.

“Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products” sets forth guidance needed for the FDA to accept results from a single, clinical study. Using the guidance document, I focused my review on whether the results could be generalized to a larger population, or not.

The guidance on clinical evidence stresses the importance of a large multi-center study to help establish the credibility of a single study submission and that the credibility is enhanced if no single center accounts for an unusually large proportion of the subjects.

When compared with all other centers, one center, the University of Alabama, is disproportionately represented in the study. The University of Alabama accounts for about 25% of all subjects enrolled (126/463) and is about three times the size of the next largest center, the University of Tennessee ( $45/463 = 9.7\%$ ).

The effect of 17P is most pronounced when started at 18 weeks gestation or earlier and does not appear effective when started at 20 weeks of gestation or later. The rate of fetal and neonatal deaths is also most pronounced among women who started study drug at 18 weeks gestation or earlier (10%). The rate decreases to 2% when study drug is started at 20 weeks of gestation or later.

These results need to be interpreted in the larger context of confounding with study center. The results of my analyses suggest the presence of confounding between center and gestational age at randomization. For example, the University of Alabama accounts for 44% of subjects enrolled at 18 weeks gestation or earlier and had relatively few patients at later ages. At other centers, the gestational age at randomization is skewed towards later gestational ages at the time of randomization.

Moreover, the University of Alabama accounts for about 50% of the fetal and neonatal deaths that occurred among women who started study drug at 18 weeks of gestation or earlier.

Thus, the apparent age trends in treatment effect, and fetal and neonatal deaths simply could be unique to the patient population enrolled at the University of Alabama.

When two studies are submitted, the chance of both studies yielding a false positive result is 1/1600. In the case of a single study, the results must be less than a nominal p-value of 0.00125 to ensure the same false positive rate. In Study 17P-CT002, the only endpoint that meets this criterion is Delivery <37 weeks gestation. Deliveries at times earlier than 37 weeks gestation were not statistically significant at 0.001. The results of the analyses of the 32 and 35 week endpoints suggest their false positive rates could be as great as 1/40.

Because of the public health need for a drug product to prevent preterm deliveries, we might be willing to accept a false positive rate that is somewhat greater than 1/1600 if the results appear to be generalizable. However, because of the issues introduced by the size of the University of Alabama and its findings, together with the 1/40 false positive rates for the 32 and 35 week endpoints, I do not believe the study results can be generalized to a larger population.

Therefore, from a statistical perspective, I do not believe this study meets the level of evidence needed to support the efficacy of 17P.

## 1.2 Brief Overview of Clinical Studies

A single Phase 3 study, submitted by the applicant to demonstrate the efficacy and safety of 17  $\alpha$ -Hydroxyprogesterone; Caproate Injection, 250 mg/mL for the prevention of recurrent preterm birth, is the focus of this review. This planned marketed drug product is abbreviated as 17P throughout the NDA and this review. The applicant states 17P is the identical formulation as both the 17-HPC used in the NICHD clinical trial and the identical formulation as the previously marketed product, Delalutin 250 mg/mL.

17-HPC is the abbreviation for the drug substance, 17  $\alpha$ -hydroxyprogesterone caproate, and identifies the product administered in previously conducted clinical trials and animal studies.

## 2. INTRODUCTION

### 2.1 Overview of Study 17P-CT002

#### 2.1.1 Study Design

The National Institute of Child Health and Human Development (NICHD) conducted Study 17P-CT002 through its Maternal Fetal Medicine Units (MFMU) Network. Drug approval was not part of the study objectives. This aspect has implications for the data analyses and conclusions; see 3.1.2.

Study 17P-CT002 is a multicenter, randomized, placebo-controlled, double-blinded clinical study of women who had at least one documented prior spontaneous preterm birth of a singleton, nonanomalous fetus. The definition of spontaneous preterm delivery (SPTD) is delivery from 20<sup>0</sup> weeks gestation (20 weeks, 0 days) to 36<sup>6</sup> weeks gestation (36 weeks, 6 days) following spontaneous preterm labor or premature rupture of membranes<sup>1</sup>.

All patients who presented for prenatal care before 20<sup>3</sup> weeks gestational age were eligible for screening. After signing the informed consent form, the subject was to receive an injection of the placebo to assess compliance and for any unusual reactions to the injection.

Randomization was planned to occur from 16<sup>0</sup> to 20<sup>6</sup> weeks gestational age. Subjects were randomized, using a simple urn method, in a 2 to 1 ratio to receive intramuscular injections (1 mL) of either 17  $\alpha$ -hydroxyprogesterone caproate injection, 250 mg/mL (17P) or Placebo. Randomization was stratified by center to ensure balance between the

---

<sup>1</sup> Throughout this review, the number of weeks and number of days for gestational age is expressed as weeks<sup>days</sup>. Thus, “20<sup>0</sup> weeks” means 20 weeks, 0 days and 36<sup>6</sup> weeks means 36 weeks and 6 days.

two treatment groups with respect to anticipated differences in the clinic population and possible difference in patient management.

The study protocol estimated 1500 women would need to be screened to achieve the desired sample size of 500 patients (334 to 17P; 166 to Placebo). Five hundred subjects were needed to detect a reduction of 33% in the rate of preterm birth (from 37% to 25%), assuming a Type I error (2-sided) of 5% and a power of at least 80%.

Study personnel administered injections of 17P or Placebo weekly through 36<sup>6</sup> weeks gestation or delivery, whichever occurred first. The first injection was given on the day of randomization.

The primary efficacy outcome was delivery <37<sup>0</sup> weeks. The primary outcome counted all deliveries occurring from randomization through 36<sup>6</sup> weeks gestation, including miscarriages, stillbirths and elective abortions. The study also measured neonatal outcomes up until the time that the mothers and infants were discharged from the hospital.

The analysis plan in the study protocol indicated an external Data Monitoring and Safety Committee would meet “periodically to review trial results.” Although the number of meetings and analyses were not stated, the protocol specified the use of the Lan and DeMets implementation of the O’Brien-Fleming boundaries.

### **2.1.2 Subject Disposition**

At 19 centers, a total of 463 subjects were randomized to treatment (310 to 17P and 153 to Placebo). One center (Center 8, University of Alabama) enrolled 126 subjects, constituting about ¼ of the subjects in the study; see Table 2.1. Four centers enrolled 36 to 45 subjects, or about half of the subjects. The remaining centers enrolled anywhere from 2 to 28 subjects each.

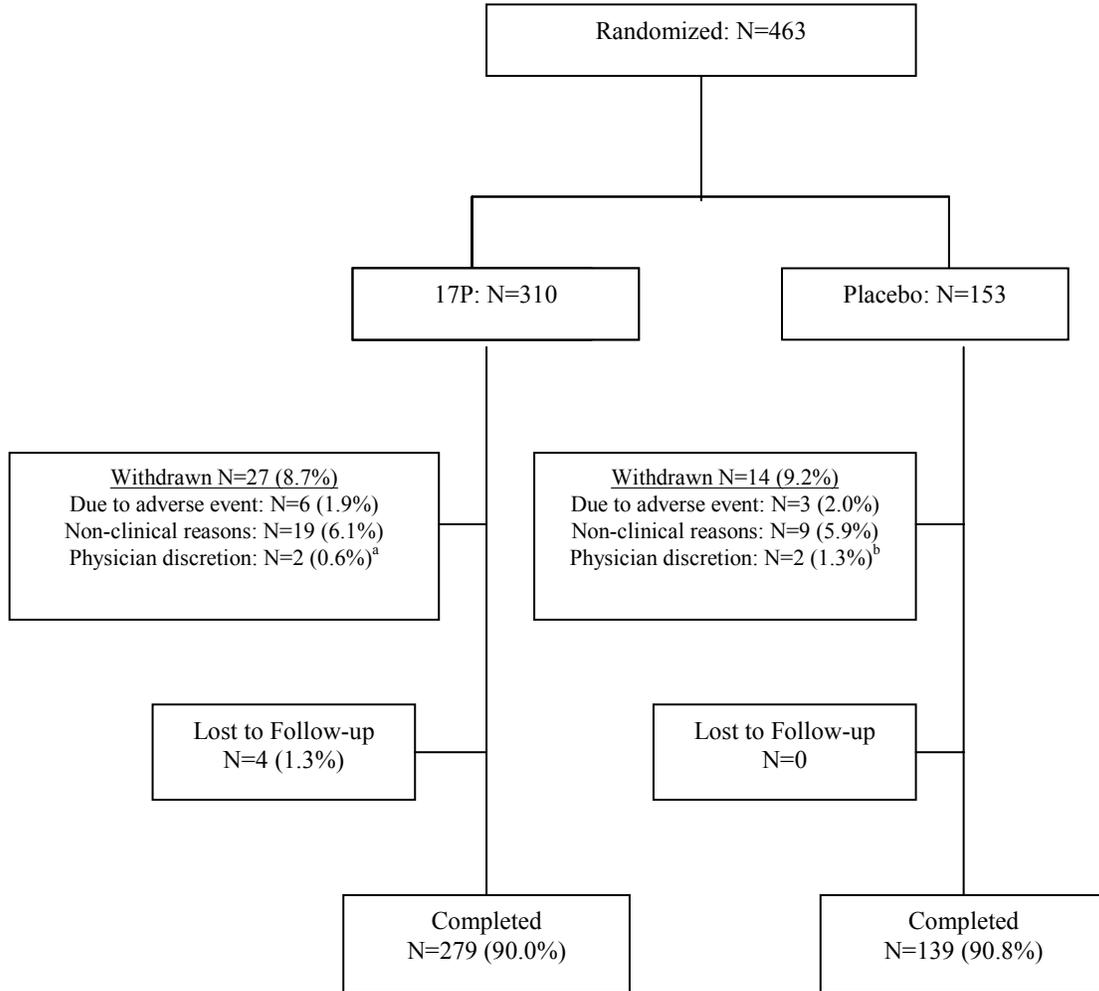
**Table 2.1 Distribution of Subjects by Study Center, Sorted by Size of Center**

Center #	Name	# enrolled
8	University of Alabama	126
4	University of Tennessee	45
20	University of Utah	43
18	University of Texas Southwestern	39
2	University of Pittsburgh	36
15	Ohio State University	28
9	Wayne State University	24
21	Thomas Jefferson University	24
13	Wake Forest University	22
11	University of Cincinnati	13
19	University of Texas San Antonio	13
17	University of Miami	11
23	Columbia University	11
14	University of Chicago	7
25	Case Western University	6
22	Brown University	5
26	University of Texas Houston	4
27	University of North Carolina, Chapel Hill	4
28	Northwestern University	2

Only four subjects were lost to follow-up, meaning their delivery data could not be obtained (Figure 2-1 and Table 2.2). The four came from the same center, Center 018, and were randomized to 17P. They are considered failures in the intent-to-treat analyses of the primary efficacy endpoint, and failures in the analyses of the secondary endpoints on the basis of the gestational age at the time of last contact. Time to delivery analyses count these subjects as censored at last known status.

Forty-one subjects were withdrawn from treatment (either 17P or Placebo), but were not withdrawn from the study. These subjects have a complete set of follow-up information.

**Figure 2-1 Subject Disposition**



Reference: Section 14.1, Post-Text Table 1 and Post-Text Table 2

Note: “Withdrawn from the study” was defined as the patient no longer received study drug. “Lost to follow-up” was defined as the patient’s delivery data could not be obtained. “Completed the study” was defined as the patient did not withdraw from the study and was not lost to follow-up.

<sup>a</sup> In the 17P group, Investigators stopped the participation of one patient due to injection site reactions and another patient due to pPROM, which was not considered an AE. Therefore, 7 (2.2%) patients in the 17P group discontinued due to AEs.

<sup>b</sup> In the Placebo group, Investigators stopped the participation of one patient due to a potential allergic reaction and another patient due to pPROM, which was not considered an AE. Therefore, 4 (2.6%) patients in the Placebo group discontinued due to AEs.

**Table 2.2 The four subjects who were lost to follow-up**

Subject ID	GA at Randomization (weeks)	Time on Study (days)	GA at Time of Last Contact
CT-018-022	20.0	102	34.6
CT-018-033	18.0	29	22.1
CT-018-035	17.6	134	36.7
CT-018-038	18.6	1	18.7

Approximately half of the subjects enrolled between 18 and 20 weeks (Table 2.3). The distributions for each treatment group were essentially identical.

**Table 2.3 Distribution of gestational age (weeks) at the time of randomization**

Percentiles	Weeks of Gestation
100.0% maximum	21.0
99.5%	20.9
97.5%	20.9
90.0%	20.7
75.0% quartile	20.3
50.0% median	19.0
25.0% quartile	17.6
10.0%	17.0
2.5%	16.3
0.5%	16.0
0.0% minimum	16.0

### **2.1.3 Data Monitoring and Safety Committee Meetings**

**Delivery <37<sup>0</sup> Weeks Gestation (yes/no)** is the primary endpoint specified in the study protocol. The protocol indicated an external Data Monitoring and Safety Committee (DMSC) would meet periodically to review trial results. The timing of the interim analyses would be at their discretion.

The analyses were conducted when 15.2% (176 patients) and 70.2% (351 patients) were randomized and had outcome data<sup>2</sup>.

A Lan-DeMets implementation of the O'Brien-Fleming stopping boundaries was used. A nominal p-value <0.0001 was required to show statistical significance at the first look and a nominal p-value <0.015 was required at the second analysis.

The DSMC met twice to discuss the results of the interim analyses<sup>3</sup>.

At the first meeting, held on 10/3/2000, the committee reviewed an interim report based on 176 patients randomized before 9/1/2000 and recommended continuation of the study.

At the second meeting, held on 2/21/2002, the DSMC reviewed an interim report of 446 subjects randomized before 1/16/2001. The boundary (p=0.015) was crossed at this second interim analysis. The committee recommended:

- Discontinuation of recruitment because 17P had demonstrated benefit for the primary outcome.
- Subjects who were in the process of being screened, including those who had received the placebo injection, should not be enrolled.
- For subjects who were currently on study, consent should be requested to continue on blinded, study medications in order to gather data that could address "important secondary questions".

At the time the study enrollment was stopped, 463 of the planned 500 women had been randomized.

#### **2.1.4 Results for the primary efficacy variable: Delivery <37<sup>0</sup> weeks gestation**

The study protocol specifies the analyses of the primary endpoint, *Delivery <37<sup>0</sup> Weeks Gestation (yes/no)*, would be based upon the total cohort of patients randomized, regardless of whether subjects took any study medication or not. The statistical analysis plan (SAP) further specifies that missing outcomes would be classified as a treatment failure (i.e., delivery < 37<sup>0</sup> weeks gestation). This affects four subjects, each of whom came from Center 18 and was randomized to 17P; see Table 2.2 above.

Moreover, because of the interim analyses and the decision to stop the study early, the final analyses uses 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.)

---

<sup>2</sup> See Volume 5.17, page 85 of 362.

<sup>3</sup> See Volume 5.17, page 188 of 362.

The applicant’s analyses show the incidence of deliveries prior to 37<sup>0</sup> weeks gestation was significantly lower in the 17P group than in the Placebo group (37.1% vs 54.9%; nominal p=0.0003); see Table 2.4. The results of the applicant’s logistic regression analyses, which adjusted for an imbalance in the number of previous preterm deliveries, are consistent with these findings.

The ITT population is all randomized patients. Patients with missing outcome data are classified as having a preterm delivery <37<sup>0</sup> weeks (treatment failure). Per-protocol are defined as the patient was eligible for the trial, was at least 90% compliant, and outcome data were available.

**Table 2.4 Applicant’s Analysis: Delivery <37<sup>0</sup> Weeks Gestation**

Data Source	17P		Placebo		Nominal P-value <sup>a</sup>	Treatment difference and its 95% Confidence Interval, adjusted for interim analyses <sup>b</sup>
	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%]

<sup>a</sup> Chi-square test. To account for the interim analyses, the nominal p-values need to be compared to 0.0345.

<sup>b</sup> 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 adjusted confidence intervals (equivalent to a 96.6% confidence interval) use the final p-value boundary of 0.0345.

The applicant’s analyses did not identify “qualitative” treatment-by-center interactions.

## 2.2 Data Sources

Volumes 5.16 and 5.17.

Datasets submitted to the EDR.

Amendment 16-1, dated 7/30/06: “NDA 21-945 Staggered KM and 3 new analyses”

Amendment 17-1, dated 9/25/06: “Conference Call Document”

Amendment 20-1, dated 8/22/06: “RE: NDA 21-945 Staggered KM and 3 new analyses”

Background Package submitted by Adeza to the Advisory Committee

Background Package submitted by FDA to the Advisory Committee

### **3. STATISTICAL EVALUATION**

#### **3.1 Evaluation of Efficacy**

##### **3.1.1 Prevention of Delivery <28<sup>0</sup>, <32<sup>0</sup>, <35<sup>0</sup>, <37<sup>0</sup> Weeks Gestation**

According to the FDA Background Document for the DRUP Advisory Committee (dated 8/2/06):

“Although preterm birth is defined as a birth prior to 37 weeks gestation, the clinical significance of preterm birth is more pronounced prior to 32 weeks gestation. In the U.S., infants born after 32 weeks have very low mortality rates, and relatively low long-term morbidity.”

Based on communications with FDA, the applicant added analyses of the following endpoints:

- Prevention of Delivery <35<sup>0</sup> weeks gestation
- Prevention of Delivery <32<sup>0</sup> weeks gestation

The analyses of these endpoints, which were specified after the study had ended and the results published, are considered *post hoc* analyses.

Some of the statistical review issues surrounding the Prevention of Delivery endpoints are:

- The study was stopped after an interim analysis of the primary endpoint, delivery <37<sup>0</sup> weeks gestation.
- The study was powered for the primary endpoint, delivery <37<sup>0</sup> weeks gestation, and not for delivery at earlier time points.
- Fetal and neonatal deaths are counted as preterm deliveries.
- The analyses do not account for time on study drug.

The proportions of deliveries prior to various gestation ages, along with confidence intervals for the treatment differences, are:

**Table 3.1 Delivery <37<sup>0</sup> Weeks, <35<sup>0</sup> Weeks, <32<sup>0</sup> Weeks, <28<sup>0</sup> Weeks Gestation, ITT population. The estimates of rates of delivery do not account for duration of drug exposure.**

Data Source	17P <sup>a</sup> (N=310) %	Placebo (N=153) %	Treatment difference and its 95% Confidence Interval, adjusted for interim analyses <sup>b</sup>
<37 <sup>0</sup> weeks	37.1	54.9	-17.8% [-28%, -7%]
<35 <sup>0</sup> weeks	21.3	30.7	-9.4% [-18.7%, -0.2%]
<32 <sup>0</sup> weeks	11.9	19.6	-7.7% [-16.1%, -0.3%]
<28 <sup>0</sup> weeks	9.4	10.5	-1.1% [-7.4%, 5.2%]

<sup>a</sup> 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).

<sup>b</sup> 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. <sup>4</sup>

The analyses in this table address the question, “Among women who started treatment (17P or Placebo) between 16 weeks and 21 weeks gestation, what is the benefit of 17P in reducing preterm deliveries by 28 weeks, 32 weeks, 35 weeks or 37 weeks gestation?” Based on the treatment effects observed for each endpoint, the results suggest the treatment difference attenuates between 35 and 37 weeks gestation.

When answering this question, two features of the analyses summarized in Table 3.1 require consideration.

- The rates of delivery include live births, neonatal deaths and fetal loss. The analyses count them equally.
- The analyses do not address the duration of drug exposure and its relationship to preterm deliveries. The characterization of the relationship between duration of treatment and time to delivery could be of interest, especially for preterm deliveries at early weeks. For instance, consider 28 weeks of gestation. At 28 weeks, the time on study drug ranges from 7 to 12 weeks, potentially affecting the outcomes at 28 weeks.

### 3.1.2 Study Not Designed for Drug Approval

From a regulatory perspective, a difficulty in reviewing this study is that it was designed for objectives other than drug approval. A different primary endpoint would have been

<sup>4</sup> 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%].

used. It is for this reason that the medical division requested analyses of deliveries <35<sup>0</sup> weeks gestation.

The study was not powered for endpoints earlier than 37 weeks. This lack of power may explain the weaker results shown for the secondary endpoints.

### **3.1.3 Interim Analyses and the Secondary Endpoints**

The Applicant and I hold different views on whether to report the results of the adjusted analyses of the secondary endpoints or the results of the unadjusted analyses endpoints.

The Applicant maintains<sup>5</sup> the secondary analyses do not need adjustments because the secondary endpoints were not assessed by the Data Monitory Committee at the interim looks of the data, whereas I believe the analyses of the secondary endpoints do need to be adjusted.

Here is my reasoning for reporting results that are adjusted for interim analyses.

#### **3.1.3.1 Correlation with Primary Endpoint**

All of the secondary endpoints are correlated with the primary endpoint, <37 weeks, which was the basis for stopping the study. The primary and secondary endpoints are not independent.

Each endpoint represents preterm deliveries cumulated from 16 weeks gestation. Thus, a delivery at <28 weeks counts as a delivery at <37 weeks; a delivery at <32 weeks counts as a delivery at <37 weeks; and a delivery at <35 weeks counts as a delivery at <37 weeks.

In the extreme, pre-specified secondary endpoints could have included <28 weeks, <29 weeks, <30 weeks, ..., <34 weeks, <35 weeks, <36 weeks. This perspective illustrates the presence of correlations among the secondary endpoints in addition to correlations with the primary endpoint.

#### **3.1.3.2 Biased Estimates**

Estimates of treatment effects, which are the basis for early termination of a study, are biased. The observed treatment effects overestimate the “true” effect. Because the secondary endpoints in this study are correlated with the primary endpoint, they too are overestimates.

#### **3.1.3.3 Single Study**

Our differences in views also need to put into a larger context. The demonstration of efficacy of 17P rests on a single study.

---

<sup>5</sup> Amendment (17-1) dated 9/25/06

The results from a single study submission need to be robust in establishing efficacy of a drug product. If the results of the secondary endpoints from this study were overwhelming, they would remain statistically significant regardless of whether a level of 0.0345 or 0.05 is used to declare significance. Further, an assessment of efficacy does not rest on the statistical significance of a single endpoint. I discuss the level of evidence from this single study submission in Section 3.1.8, A Single Study Submission.

### **3.1.4 Placebo Response**

The Medical Review Division notes the rate of pre-term deliveries <37 weeks gestation among Placebo-treated women (55%) is higher than those observed in other studies conducted by NIH, and is higher than what was used to power the study (37%). Potentially, they believe this higher than anticipated response rate could contribute to the observed benefit of 17P in preventing pre-term deliveries prior to 37 weeks gestation.

The Placebo response rate, however, does not appear to be responsible for the statistically significant treatment effect at 37 weeks gestation. Not only is the response rate higher than expected for the Placebo-treated women (55% vs 37%), it is also higher than expected for the 17P-treated women (37% vs 25%)<sup>6</sup>.

The decrease in pre-term deliveries among 17P-treated women, relative to the Placebo-treated women is 33%. This relative decrease is consistent with what the protocol states is an important reduction in risk.

The reasons for the unanticipated rates of preterm deliveries are not clear. Possibly, the women in the study are at higher risk than anticipated.

### **3.1.5 An Alternative View of the Data – Prolongation of Pregnancy (Delaying Time-to-delivery)**

The use of categories to define preterm deliveries provides looks of slices in time. The results suggest an important treatment difference emerges between 35 and 37 weeks. The use of categories, however, do not tell a complete story. In fact, Kaplan-Meier graphs of the entire time course from the start of study treatment through 40 weeks gestation age give a different impression.

The evolution of the protocol, the secondary analyses requested by FDA prior to the submission of the NDA, the analyses included in the Advisory Committee background packages and the analyses presented to the Advisory Committee by the applicant and by FDA provide evidence for interest in fine gradations of the preterm delivery categories.

---

<sup>6</sup> The expected rates of 37% for the Placebo treatment group and 25% for the 17P treatment group were used to power the study.

Originally the study protocol specified preterm delivery <37<sup>0</sup> weeks as the primary endpoint and delivery <35<sup>0</sup> weeks as a secondary endpoint. During the review process, including the Advisory Committee background documents and presentations to the Advisory Committee, deliveries at <37<sup>0</sup>, <35<sup>0</sup>, <32<sup>0</sup>, <30<sup>0</sup>, <28<sup>0</sup>, <24<sup>0</sup> and <20<sup>0</sup> weeks were analyzed.

The Kaplan-Meier estimates and displays, shown in the following sections, explore the effect of treatment from the start of therapy through delivery.

An interesting wrinkle to the interpretation of the study results is the desire of clinicians to express deliveries in terms of gestational age rather than expressing time to delivery as a function of date of randomization or start of treatment. My review also reports Kaplan-Meier estimates that address this issue.

### **3.1.5.1 Time from Randomization to Delivery**

Typically, displays of the time course of outcomes from a clinical trial data start at the time of randomization. Graphs of Kaplan-Meier estimates of the time to delivery relative to randomization for this study show the following (Figure 3-1)<sup>7</sup>.

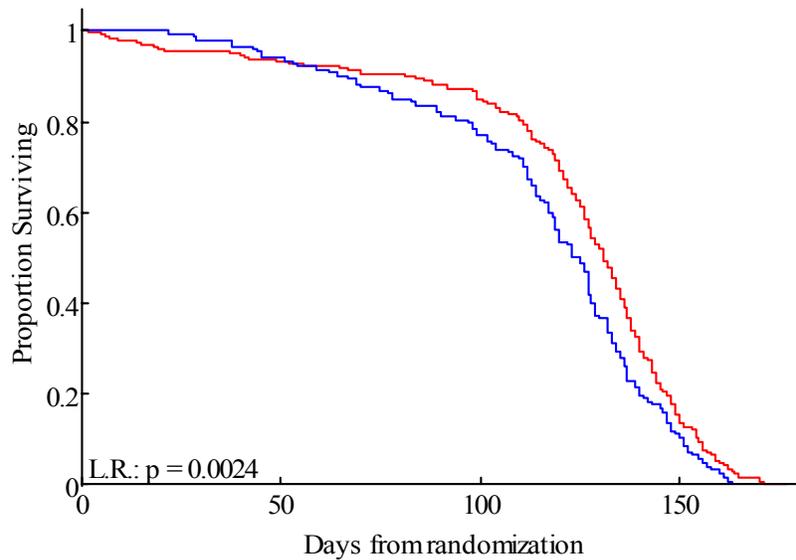
The difference in the shapes of the curves is statistically significant (p=0.0024; log-rank test).

The graphs suggest that during the first 8 weeks of treatment, women treated with 17P tend to deliver earlier than women treated with placebo. Of those remaining undelivered by 8 weeks, the pattern reverses; placebo-treated women tend to deliver earlier than women treated with 17P. The curves cross each other at around 53 days following randomization.

---

<sup>7</sup> In the clinical study report, the x-axis of Figure 11-1, Prolongation of Pregnancy, is labeled incorrectly and should read “Days from Randomization”.

Figure 3-1 Days from randomization to delivery



### 3.1.5.2 Time to delivery versus Gestational Age

While the Kaplan-Meier analyses presented in the previous section represent the traditional way of depicting survival-type data, the clinical importance of expressing the timing of deliveries as a function of gestational age requires a different approach.

To parallel the traditional Kaplan-Meier approach, the analyses need to accommodate the range of gestational ages at study entry and, therefore, the time on study drug prior to delivery. This approach differs from the analyses of the primary and secondary endpoints, which ignore the time on study drug.

To illustrate the difference consider delivery prior to 20 weeks, the WHO definition of a miscarriage. Approximately 75% of the subjects were randomized before 20 weeks gestation. Sixteen weeks gestation was the earliest age at which randomization took place. Therefore, by 20 weeks, the time on study drug ranged from 0 to 4 weeks.

Using the crude rate estimated by the secondary analyses, the rate of preterm delivery at <20 weeks among 17P-treated subjects is 5/221 or 2.3%<sup>8</sup>. The key to this analysis is the denominator -- 221 subjects. The estimation of the crude rate assumes all 221 subjects were on study drug for the same amount of time. Clearly, this is not the case. In fact, some subjects were on drug for only a single day including one subject who was a loss to follow-up. Thus, the crude rate is an underestimate of the “true” rate of preterm deliveries.

The denominator for a rate that accounts for time on study drug and that more accurately reflects the effect of study drug will be smaller than the 221 used to calculate the crude rate.

---

<sup>8</sup> Of the 310 women randomized to 17P, 221 were randomized at 20 weeks gestational age or earlier.

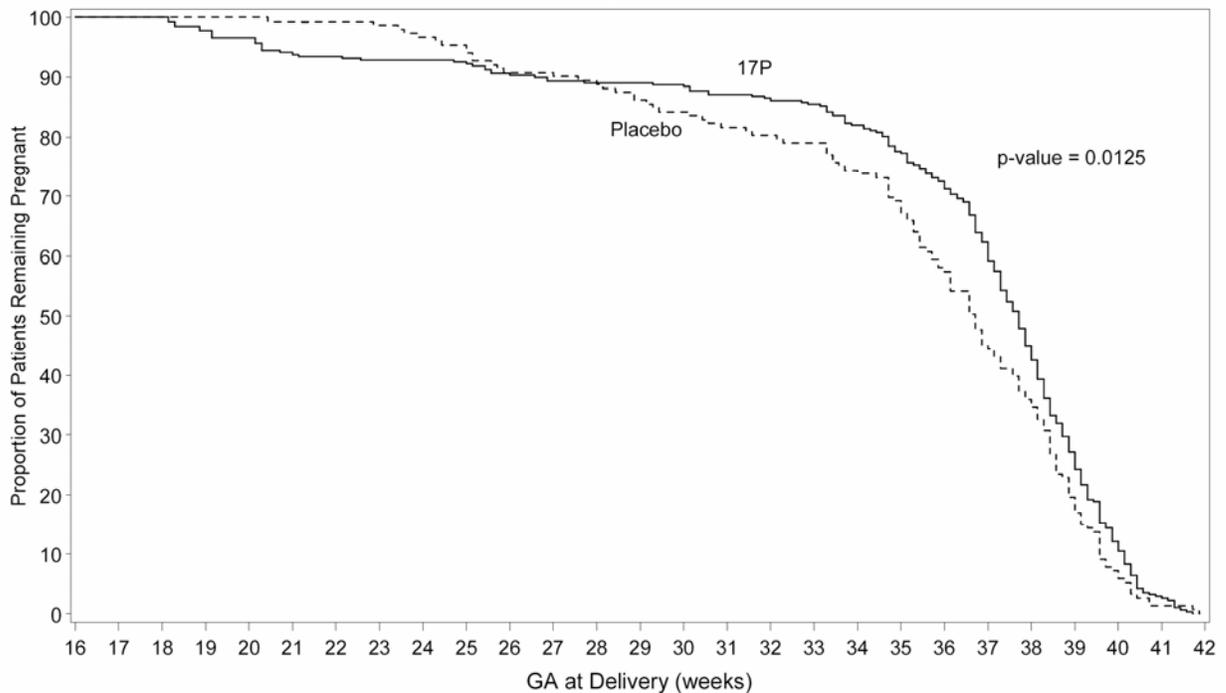
As will be shown below, when accounting for time on study drug, the estimated rate of miscarriages among 17P-treated women is 3.5%, compared with the crude rate of 2.3%.<sup>9</sup>

At our request, the Applicant provided a Kaplan-Meier analysis that takes into account the gestational age at the time a woman was randomized into the study; see Figure 3-2. The results of the log-rank test show that the difference in the shapes of the curves is statistically significant (p-value = 0.0125).

At around 28 weeks, Figure 3-2 indicates the proportions of women remaining undelivered were about the same for each treatment group. The lack of difference at 28 weeks represents the point at which the curves cross each other. This observation is consistent with the finding presented in Table 3.1.

What the table showing the results at the various time points does not disclose, however, is the following observation. Prior to 28 weeks, women randomized to 17P tended to deliver earlier than women randomized to Placebo.

**Figure 3-2. Time to delivery as a function of gestational age, using staggered entry based on the gestational age at randomization.**



Source: Adeza Response to FDA's request dated 7/20/06

Note: Four 17P patients were censored at times 18.57, 22.00, 34.39 and 36.57 weeks.

<sup>9</sup> Note that the rate of miscarriages provided by the Applicant contains an error. The rate uses a denominator of 310, the total number of women randomized to 17P. However, the denominator needs to be 221 – the number subjects who were randomized prior to 20 weeks.

### 3.1.6 Fetal Loss

Approximately 6.5% of the women in each treatment group experienced a fetal or neonatal deaths: 17P, 19/310; Placebo 11/153. Through 24 weeks gestation, all preterm deliveries were either a fetal or a neonatal death. The first viable delivery occurred just after 24 weeks gestation.

The results below show that despite the treatment groups having about the same rate of fetal and neonatal deaths, the losses occur earlier among 17P-treated women.

To assess the effect of study treatment, the rates of fetal and neonatal deaths need to account for time on study drug. The following rates come from the Kaplan-Meier estimates based on staggered entry into the study (see Appendix 5.1).

**Table 3.2 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%

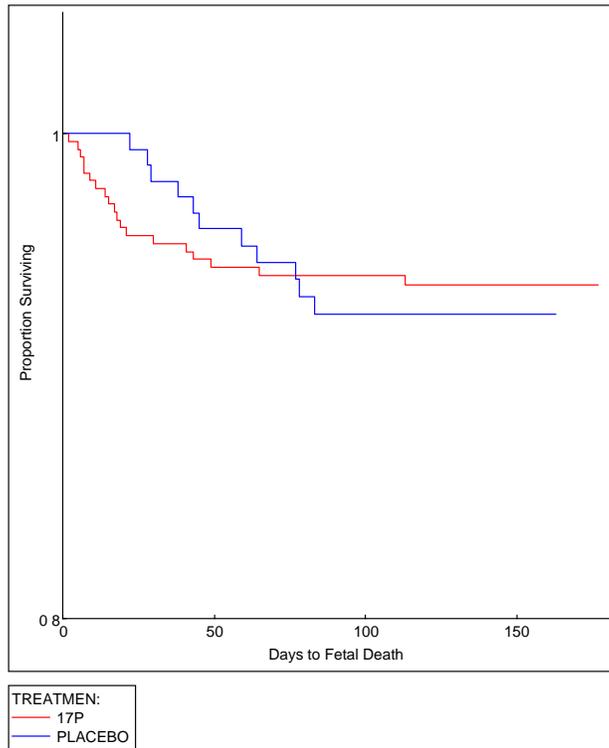
Source: Figure 3-2. Time to delivery as a function of gestational age, using staggered entry based on the gestational age at randomization., and Appendix 5.1 Listings of Kaplan-Meier Estimates of Time to Delivery, Incorporating Gestational Age at Randomization as a Left-censored Variable

This table illustrates that through 24 weeks gestation, the rates of fetal and neonatal deaths among 17P-treated women are much greater than the rates among Placebo-treated women.

To examine further the relationship between study treatment and fetal loss, I did a time-to-event analysis. Other preterm deliveries are not included.

The Kaplan-Meier estimates for fetal loss show fetal losses occur earlier among 17P-treated women when compared with Placebo-treated women (Figure 3-3). This finding reiterates the results from Table 3.2.

**Figure 3-3 Time to fetal or neonatal deaths, by treatment group.**



### 3.1.7 Summary Statistics: Mean and Median

The submission reports:

- “The mean gestational age at delivery ... was one week higher in the 17P group compared with the Placebo group (36.2 vs. 35.2 weeks;  $p=0.0024$ ).”<sup>10</sup> The median gestational ages at time of deliver are 37.5 weeks for 17P-treated women and 36.5 weeks for Placebo-treated women. The means and medians come from the subset of patients not lost to follow-up.

<sup>10</sup> See Volume 5.16: page 105 of 301, and Table 10, page 141 of 301.

- “The median prolongation of pregnancy, defined as the time from randomization until delivery or date last pregnant, was significantly higher in the 17P group (131 vs 125 days; p=0.0024).”<sup>9</sup>

In each case, the p-value comes from a log-rank test. This same log-rank test assesses whether the distributions of time from randomization to delivery for the 17P and Placebo treatment groups are different or not different. See 3.1.5.1 **Time from Randomization to Delivery** of this review.

The log-rank test is a global test. It tests whether the time-to-event curves are the same or not. It does not test specific hypotheses aimed at the median or the mean.

When analyzing time-to-event data, researchers often present the results of a log-rank test alongside summary statistics. The implication is the p-values represent the results of comparing the means and medians between treatment groups. The test, however, does not assess whether treatment groups differ in their means or medians.

However, we can compare the means and test the comparison.

**Table 3.3 Mean Gestational Age at Time of Delivery, by Treatment Group. Excludes four 17P-treated subjects who were lost to follow-up.**

Treatment Group		Treatment Difference (17P minus Placebo)	95% confidence interval adjusted for interim analyses	
17P	Placebo			
Mean	36.2 weeks	35.1 weeks	1.1 weeks	[.01, 2.0]

**Table 3.4 Mean Time to Delivery, by Treatment Group. Based on Kaplan-Meier analysis.**

Treatment Group		Treatment Difference (17P minus Placebo)	95% confidence interval adjusted for interim analyses	
17P	Placebo			
Mean	123 days	116 days	7 days	[.03, 14.03]

In each situation, with 95% confidence the true treatment difference ranges from about zero to two weeks.

### 3.1.8 A Single Study Submission

This submission contains a single study to support the claim of effectiveness for 17P. The absence of a second adequate and well-controlled clinical trial is an important review issue.

Based on my analyses, which I discuss in the following sections, this study does not appear to meet the standard of evidence needed to demonstrate efficacy of 17P.

According to “Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products,”<sup>11</sup> reasons for questioning the adequacy of a single study include the following.

- “Any clinical trial may be subject to unanticipated, undetected, systematic biases.”
- “The inherent variability in biological systems may produce a positive trial result by chance alone. ... Even if all drugs tested in such trials were ineffective, one would expect one in forty of those trials to ‘demonstrate’ efficacy by chance alone at conventional levels of statistical significance.”
- “Results obtained in a single center may be dependent on site or investigator.”

Further, “it should also be appreciated that reliance on a single study of a given use, whether alone or with substantiation from related trial data, leaves little room for study imperfections or contradictory (non-supportive) information. In all cases, it is presumed that the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post-hoc changes in analysis, or other factors is judged to be minimal.”

The guidance document identifies characteristics of a single clinical trial that could make the study adequate to support an effectiveness claim. The characteristics pertinent to this submission are:

- Large multicenter study
  - No single study site provides an unusually large fraction of the patients, and
  - No single investigator or site is disproportionately responsible for the favorable result seen
  - If analysis shows that a single site is largely responsible for the effect, the credibility of a multicenter study is diminished
- Consistency across study subsets
  - Large trials frequently have broad entry criteria and the study population may be diverse with regard to important covariates such as concomitant or prior therapy, disease stage, age, gender or race.
  - Analysis of the results for consistency across key patient subsets across key patient subsets addresses concerns about generalizability of finding to

---

<sup>11</sup> Available at <http://www.fda.gov/cder/guidance/1397fn1.pdf>

various populations in manner that may not be possible with smaller trials or trials with more narrow entry criteria.

- Multiple endpoints involving different events
  - A single study may include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect.
  - Where a study shows statistically persuasive evidence of an effect on more than one of such endpoint, the internal weight of evidence of the study is enhanced.

The guidance further cautions “even a strong result can represent an isolated or biased result, especially if that study is the only study suggesting efficacy among similar studies. ... It is critical that the possibility of an incorrect outcome be considered and that all the available data be examined for their potential to either support or undercut reliance on a single multicenter trial.” Even with one, strong result, the results need to be reproducible within the study itself.

### **3.1.8.1 Level of Statistical Significance**

To support the efficacy of a drug product, the FDA typically requires two independent adequate and well-controlled studies, each of which is statistically significant at 0.05. When only one study is submitted, the two-sided p-values need to be less than a nominal level of 0.00125<sup>12</sup>. When adjusting for interim analyses, this value is even smaller. For example, the p-value of 0.0345, which is applied to the final analysis of the primary endpoint, yields a nominal p-value of 0.0006 for a single study. This is a conservative estimate.

In this study, Delivery <37 weeks gestation is the only endpoint whose p-value (0.0003) is smaller than the nominal value of 0.00125. It also is smaller than the conservative value of 0.0006, which is adjusted for interim analyses.

None of the other analyses, including the time-to-delivery analyses, is statistically significant at .00125. Because the confidence intervals for the treatment effects for Delivery <35 weeks gestation and for Delivery <32 weeks gestation just exclude zero, we can conclude the rate of a false positive is about 1/40 for these two endpoints.

### **3.1.8.2 Higher than Anticipated Response Rates**

The 17P and Placebo response rates for the primary endpoint in this study are higher than what was anticipated, and are higher than those observed in other studies. Without a second trial, we cannot conclude whether these rates are higher because of characteristics

---

<sup>12</sup> The false positive rate in favor of study drug for a single study with a Type I error rate of 0.05 (2-sided) is  $.05/2 = 0.025$ . The probability that two studies will *both* yield a false positive is  $(0.025)*(0.025) = 0.000625$ . Thus, a two-sided p-value for a single study needs to be less than  $2*(0.000625) = 0.00125$  to ensure the rate of a false positive outcome for a single study is the same that would be seen for two, independent studies.

unique to this study population or whether they are higher simply because women at high risk for early deliveries were enrolled into a blinded, controlled clinical study.

### **3.1.8.3 Multi-center Study**

The guidance on clinical evidence stresses the importance of a large multi-center study to help establish the credibility of a single study submission. The guidance also notes the credibility of a single study is enhanced if no single center accounts for an unusually large proportion of the subjects and that no single center is disproportionately responsible for the observed results.

The University of Alabama, which enrolled 126 subjects, accounts for about 25% of the total enrollment. The next largest site, University of Tennessee, enrolled 45 subjects or about 8% of the total number of subjects. “Table 2.1 Distribution of Subjects by Study Center, Sorted by Size of Center” summarizes the number of subjects enrolled by each study center.

My review addresses three types of endpoints:

- Delivery prior to certain time points
- Days from randomization to delivery
- Gestational age at delivery, accounting for gestational age at study entry

To explore the impact, if any, of individual centers on the study results I consider each of these three endpoints separately.

#### ***Prevention of Delivery <37<sup>0</sup>, <35<sup>0</sup> and <32<sup>0</sup> Weeks Gestation***

The following table shows the results of the primary and secondary endpoints for the University of Alabama and all other centers combined. Appendix 5.2 summarizes the results for each center.

**Table 3.5 Delivery <37 weeks, <35 weeks, <32 weeks: University of Alabama versus All Other Centers**

Endpoint	University of Alabama				All Other Centers Combined			
	17P <sup>a</sup> (n=86) %	Placebo (n=40) %	p-value <sup>b</sup>	95% CI <sup>c</sup> around treatment difference	17P <sup>a</sup> (n=224) %	Placebo (n=113) %	p-value <sup>b</sup>	95% CI <sup>c</sup> around treatment difference
<37 weeks	26.7	45.0	.042	-37%, -0.8%	41.1	58.4	.003	-29%, -5%
<35 weeks	17.4	27.5	.194	-28%, 6%	22.8	31.9	.072	-20.0%, 1%
<32 weeks	10.5	25.0	.034	-32%, 0.04%	12.5	17.7	.197	-15%, 3%

Source: Response to FDA Question 1, 10/6/06; p-values calculated by statistical reviewer

<sup>a</sup> 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).

<sup>b</sup> Fisher's exact test, 2-sided

<sup>c</sup> The confidence intervals are adjusted for interim analyses.

Because the University of Alabama accounts for about 25% of the subjects, we would expect some of the results for all the centers to become non-significant when the analyses exclude University of Alabama. Thus, the one finding that is notable is the result for delivery <32 weeks among all other centers combined, which is non-significant (p=.197). Moreover, the results for University of Alabama are statistically significant for this endpoint (p=0.034). This may suggest that the University of Alabama may be responsible for the overall findings for this endpoint.

Among the other two endpoints for All Other Centers Combined, the p-values for <37 weeks and <35 weeks are .003 and 0.072, respectively, and are consistent with the results for all study centers (Table 3.1).

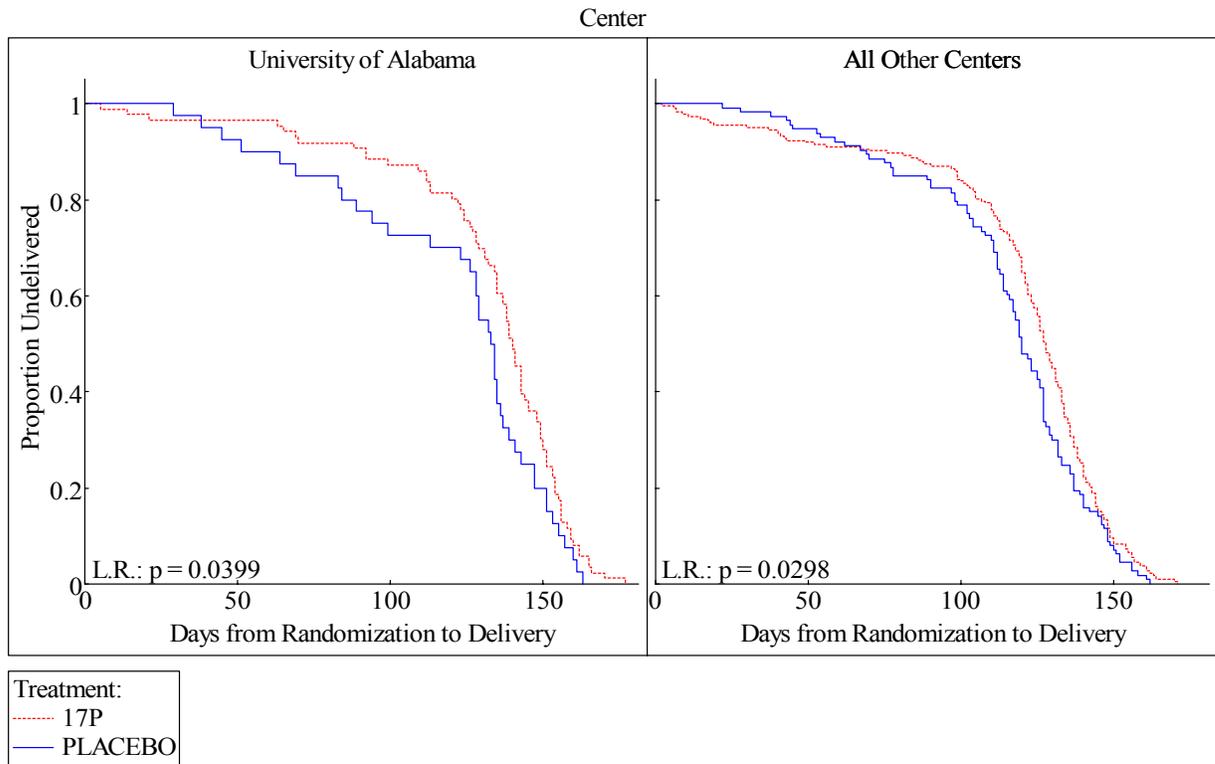
### ***Time from randomization to delivery***

For all centers combined, the difference between 17P and Placebo in the curves for time from randomization was significantly different (Figure 3-1 Days from randomization to delivery; p=.0024). To determine the impact, if any, of University of Alabama on these results, I did a stratified analysis. One stratum contains the University of Alabama and the other stratum contains all the other centers combined.

A stratified log-rank test (p=0.0036), which compares sets of time-to-event curves across strata, indicates that the results for the University of Alabama differ from the results of all the other centers combined (Figure 3-4). Inspection of the curves suggests the University

of Alabama has an attenuated treatment effect. Note also that the p-value (log-rank test) for “All Other Centers” increases from .002 to .03 when the University of Alabama is excluded, suggesting an important contribution of the University of Alabama to the overall results for this endpoint.

**Figure 3-4. Time to Delivery: University of Alabama (n=126) vs. All Other Centers Combined (n=337)**



The summary statistics for each of the curves help characterize the differences (Table 3.6). The average treatment effect at the University of Alabama is 13 days versus 4 days at all other centers combined. This difference is reinforced by the differences in medians and the 25<sup>th</sup> percentiles. The difference in the 75<sup>th</sup> percentiles is not as pronounced.

**Table 3.6 Time to Delivery: Summary statistics for University of Alabama and All Other Centers Combined. [95% Confidence Intervals Adjusted for Interim Analyses]**

Summary Statistic	University of Alabama		All Other Centers Combined	
	17P	Placebo	17P	Placebo
75 <sup>th</sup> percentile	151 [148, 156]	145 [135, 153]	140 [136, 143]	133 [127, 140]
Median	140 [135, 144]	133 [126, 137]	128 [125, 131]	120 [116, 127]
25 <sup>th</sup> percentile	126 [113, 135]	96.5 [69, 129]	113 [105, 119]	104 [90, 112]
Mean	133 [125,140]	120 [107,133]	120 [114, 125]	115 [109, 121]

Among 17P-treated women, the time-to-delivery is longer among women enrolled at the University of Alabama than those enrolled at the other centers. This difference ranges from 11 to 13 days, depending on the summary statistic used. Note that the confidence intervals for 17P from the University of Alabama do not overlap the confidence intervals from all other centers combined, indicating these differences in the summary statistics are statistically significant.

The confidence intervals for the point estimates among Placebo-treated women are essentially identical. The point estimates for the difference between the University of Alabama and the other centers combined are not consistent. The differences in the 75<sup>th</sup> percentile and the median are about 13 days longer among women enrolled at the University of Alabama. The difference in the 25<sup>th</sup> percentile is 8 days shorter among women treated at the University of Alabama.

The confidence intervals for the percentiles suggest the apparent difference in treatment effect could be driven by the 17P treatment arm. To explore this possibility, I compared the 17P treatment arms and the Placebo treatment arms (Figure 3-5). The figures also suggest the apparent difference in treatment effects appears to be driven by the 17P treatment arm.

**Figure 3-5 Time to Delivery: University of Alabama vs. All Other Centers Combined, by Treatment Group**

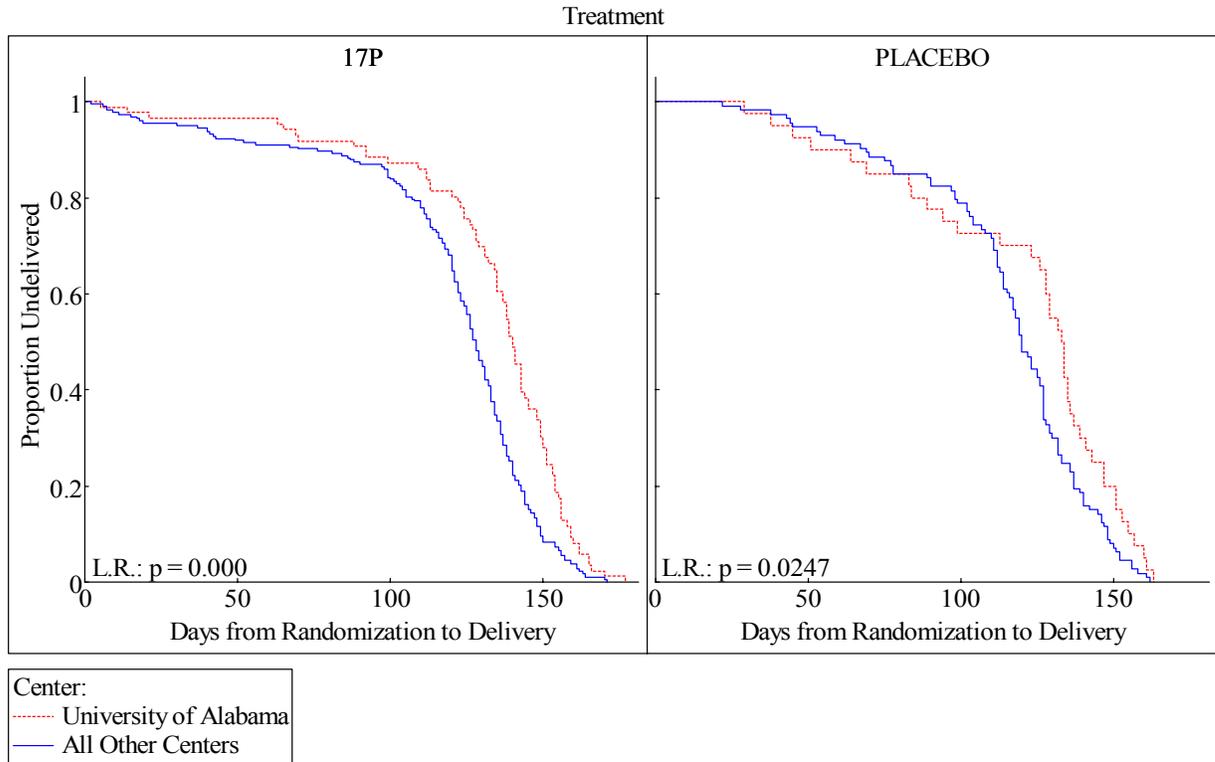
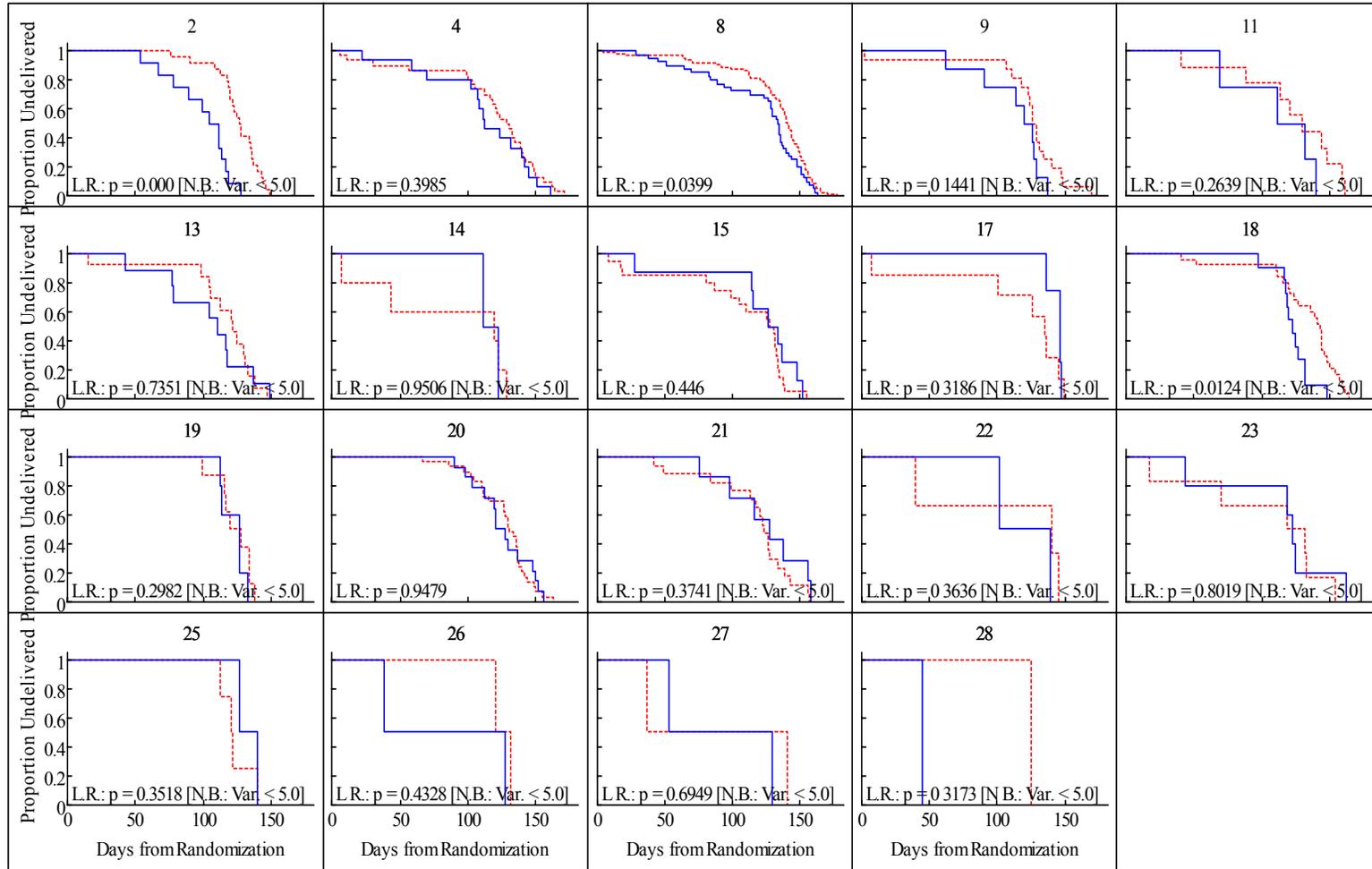


Figure 3-6 shows the time-to-delivery for each of the centers. Among the other centers with more than 30 subjects (Centers 4, 20, 18 and 2), Center 2 and possibly Center 18 appear different from the other centers.

**Figure 3-6 Time-to-delivery, by Center**

CENTER



Treatment:  
 - - - 17P  
 — PLACEBO

### ***Time to delivery versus Gestational Age***

Appendix 5.3 Time-to-delivery, Incorporating Gestational Age at Randomization as a Left-censored Variable, by Center shows the time-to-delivery versus gestational ages curves for each center. Visual inspection shows many centers did not have any early deliveries.

Generally, the pattern for University of Alabama (Center 8) mirrors what is seen for the entire study (compare with Figure 3-2. Time to delivery as a function of gestational age, using staggered entry based on the gestational age at randomization.). The difference in the shapes of the curves is not statistically significant ( $p=0.166$ ). Unfortunately, we do not have an analysis of all other centers combined.

Among the four other centers that enrolled more than 30 subjects (Centers 4, 20, 18 and 2, by decreasing size of center – see Table 2.1), only Center 4 appears to have distributions of time-to-delivery that look like the distributions for all centers combined.

Centers 2 and 18 had few early deliveries, while the first delivery at Center 2 occurred at around 28 weeks and at 31 weeks for Center 20.

At Center 18, two deliveries, both among 17P-treated women, occurred at less than 25 weeks gestation. The first delivery among Placebo-treated women occurred at around 30 weeks. Subsequently the treatment effect of 17P is substantial ( $p < 0.0001$ , log-rank test).

#### **3.1.8.4 Consistency among Important Subgroups: Delivery Endpoints**

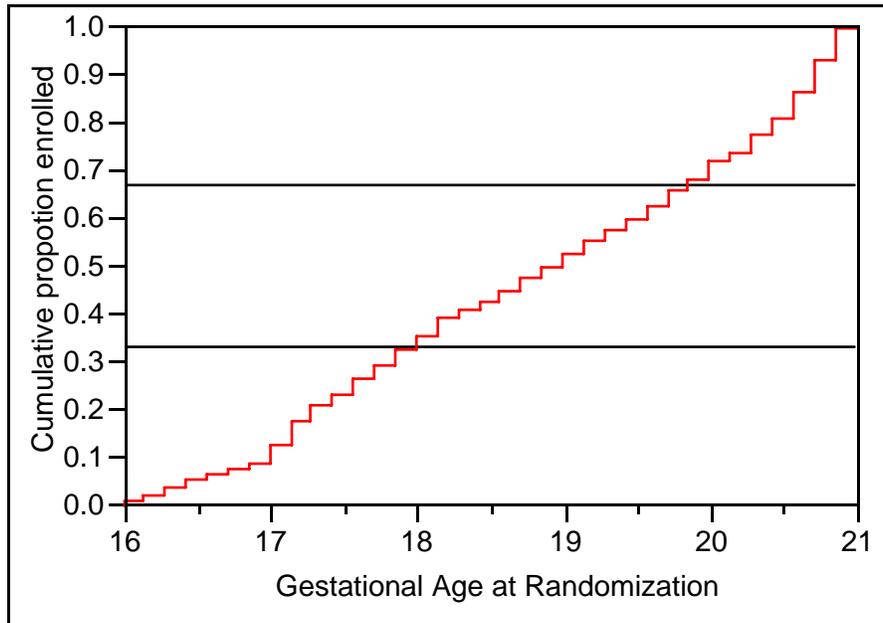
When a single study is submitted to support a claim of effectiveness, results need to show consistency among important subgroups.

17P is known to be more effective when started at earlier ages of gestation. To gain an understanding of the higher rates of delivery among 17P-treated women within the first 50 days of randomization, I explored the effect of gestational age (GA) at the time of randomization.

Based on the tertiles of the distribution of gestational age at the time of randomization, regardless of treatment (Figure 3-7), I generated analyses and graphs for three groups of gestational age:

- $\leq 18$  weeks,
- 18.1 – 20 weeks, and
- $> 20$  weeks.

**Figure 3-7 Cumulative distribution plot of gestational age at randomization. The horizontal lines denote the 33<sup>rd</sup> and 67<sup>th</sup> percentiles.**

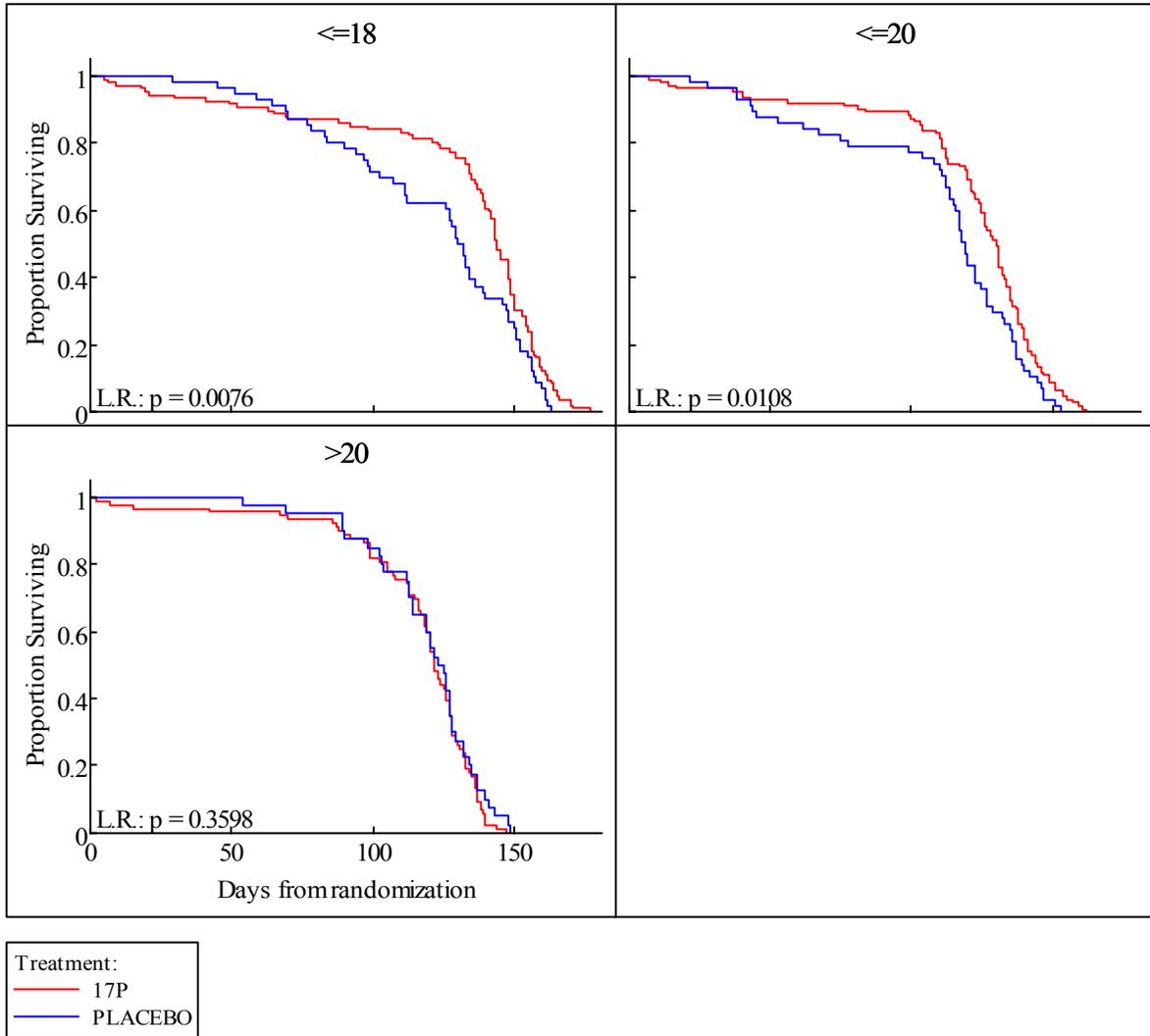


Analyses (stratified log-rank tests) show that the treatment effect varies according to when treatment started.

The following figure (Figure 3-8) shows no treatment effect among the 19% of subjects who started treatment during the 20<sup>th</sup> week of gestation ( $p=.36$ ). The treatment effect is statistically significant among the 35% of women who started treatment at 18 weeks gestation or earlier ( $p=.0076$ ) and among the 36% who started treatment between 18 and 20 weeks gestation ( $p=0.011$ ).

**Figure 3-8 Time to delivery, by gestational age at randomization**

Gestational Age at Randomization



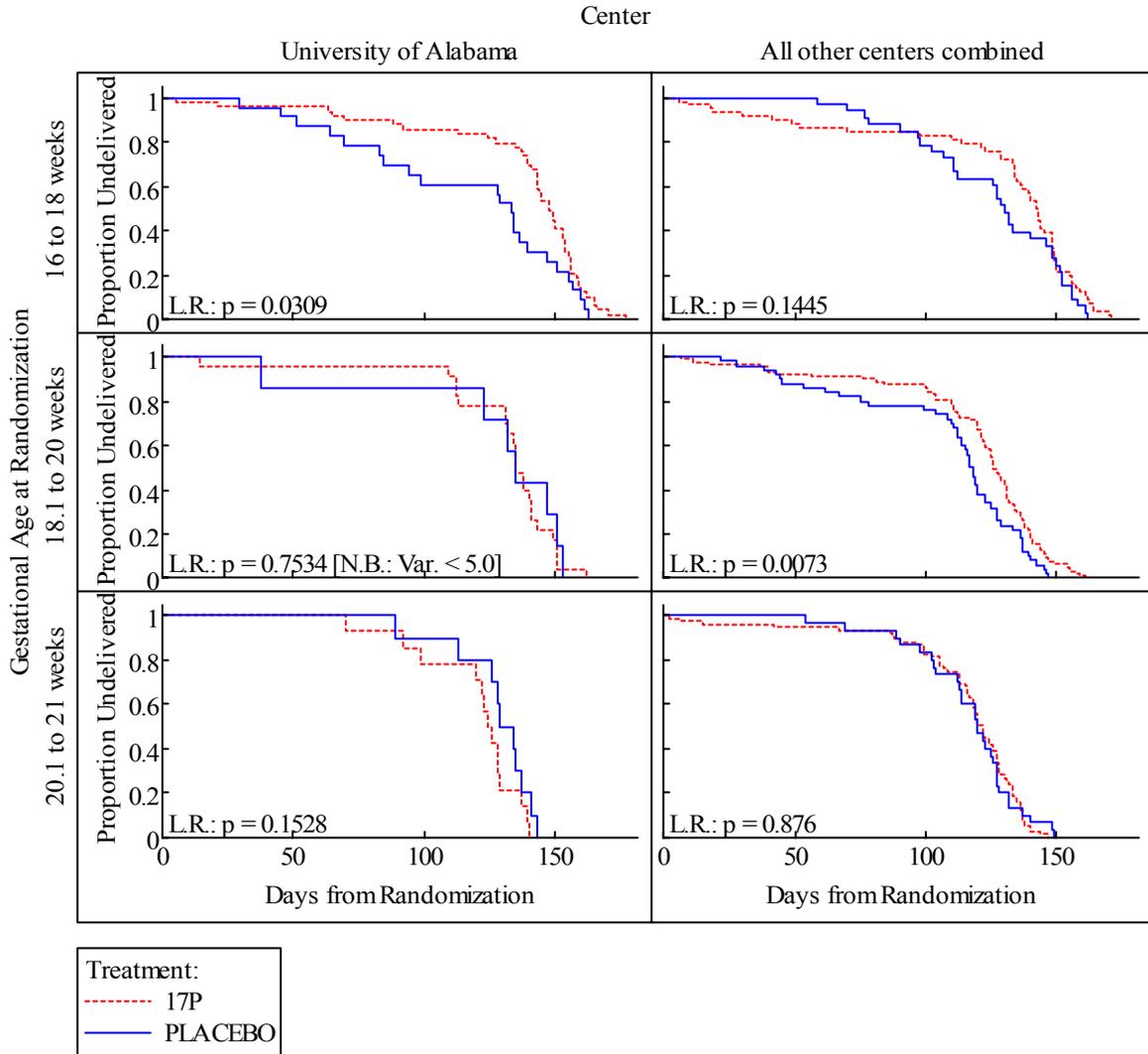
To assess the contribution to these results by the largest center, I examined University of Alabama versus all other centers combined (Figure 3-9).

Among women enrolled at 16 weeks to 18 weeks of gestational age, the treatment effect at the University of Alabama is very pronounced and statistically significant. The shapes of the curves in this subgroup of women mirror the curves for the University of Alabama when gestational age at randomization is ignored; see Figure 3-5.

For all other centers combined, the effect among women enrolled at 16 weeks to 18 weeks of gestational age is not significant.

Among all other centers combined, a statistically significant treatment effect is observed only for women enrolled between 18 and 20 weeks.

**Figure 3-9. Time to Delivery, by Gestational Age at Randomization: University of Alabama versus All Other Centers Combined**



### 3.1.8.5 Consistency among Important Subgroups: Fetal and Neonatal Deaths

On the basis of the tertiles for gestational age at randomization, the distribution of fetal and neonatal deaths shows a decreasing trend. The percentage of deaths ranges from 10% among those randomized prior to 18 weeks gestation to 2% among those randomized during the 20<sup>th</sup> week of gestation.

The University of Alabama accounts for about half of the deaths that occurred among women who were randomized at 18 weeks of gestation or earlier.

**Table 3.7 Distribution of Fetal and Neonatal Deaths, by Gestational Age at Randomization**

	Gestational Age at Randomization		
	≤18 weeks	18.1 – 20 weeks	>20 weeks
N	150	163	146
% deaths	10%	7%	2%

**Table 3.8 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
<i>Number of deaths:</i>	15	12	3
<i>Center #</i>			
2	1 ( 6.7%)	-	-
4	3 (16.7%)	2 (16.7%)	-
8	7 (46.7%)	3 (25.0%)	-
9	-	-	1 (33.3%)
13	-	2 (16.7%)	1 (33.3%)
14	-	1 ( 8.3%)	1 (33.3%)
15	1 (6.7%)	2 (16.7%)	-
17	-	1 ( 8.3%)	-
18	1 (6.7%)	-	-
21	1 (6.7%)	-	-
23	-	1 ( 8.3%)	-

### 3.1.8.6 Confounding of Center and Gestational Age at Randomization

The fewer number of early deliveries at the University of Alabama among women randomized later than 18 weeks gestation, suggests the possibility that study center is confounded with gestational age at randomization.

The following breakdown (Table 3.9) reinforces this possibility. Clearly, the distribution of gestational ages at randomization for University of Alabama differs from the distribution for all other centers.

The University of Alabama accounts for a disproportionate number of patients randomized at 18 weeks or earlier. The center enrolled 44% (72/164) of all subjects

enrolled during the first two weeks, compared with 18% and 23% in the subsequent weeks.

**Table 3.9 Distribution of Gestational Age at Randomization: University of Alabama versus All Other Centers Combined.**

	N	16 weeks - 18 weeks	18.1 weeks – 20 weeks	20.1 weeks – 21 weeks
University of Alabama	126	57% (n=72)	24% (n=30)	19% (n=24)
All Other Centers	337	27% (n=92)	41% (n=140)	31% (n=105)
Total	463	35% (n=164)	37% (n=170)	28% (n=129)

**Table 3.10 Delivery <37<sup>0</sup> Weeks, <35<sup>0</sup> Weeks, <32<sup>0</sup> Weeks Gestation, by Gestational Age at Randomization: University of Alabama versus All Other Centers**

	16 weeks - 18 weeks		18.1 weeks – 20 weeks		20.1 weeks – 21 weeks	
	<b>Delivery &lt;37 weeks gestation</b>					
	17P	Placebo	17P	Placebo	17P	Placebo
University of Alabama	29%	61%	26%	29%	21%	20%
All Other Centers	44%	61%	46%	68%	33%	40%
All	37%	61%	42%	63%	31%	35%
	<b>Delivery &lt;35 weeks gestation</b>					
University of Alabama	18%	39%	13%	14%	21%	10%
All Other Centers	27%	45%	24%	32%	19%	17%
All	23%	43%	22%	30%	19%	15%
	<b>Delivery &lt;32 weeks gestation</b>					
University of Alabama	14%	39%	4%	14%	7%	0%
All Other Centers	20%	21%	14%	22%	7%	7%
All	18%	29%	12%	21%	7%	5%

Compared with all the other centers, Center 2 appears to have a deficit of subjects enrolled at 18 weeks of gestation or earlier (Table 3.11). About half of its subjects enrolled during the 20<sup>th</sup> week of gestation. This skewed distribution likely explains the absence of early deliveries (see Figure 3-6).

**Table 3.11 Distribution of Gestational Age at Randomization, by Center**

Center #	N	16 weeks - 18 weeks	18.1 weeks – 20 weeks	20.1 weeks – 21 weeks
All	463	35%	37%	28%
2	36	8%	39%	53%
4	45	44%	44%	11%
8	126	57%	24%	19%
9	24	29%	33%	37%
11	13	38%	38%	23%
13	22	14%	55%	32%
14	7	0%	57%	43%
15	28	32%	36%	32%
17	11	18%	73%	9%
18	39	41%	41%	18%
19	13	15%	38%	46%
20	43	30%	28%	42%
21	24	29%	42%	29%
22	5	20%	60%	20%
23	11	18%	27%	55%
25	6	17%	33%	50%
26	4	0%	75%	25%
27	4	25%	75%	0%
28	2	0%	100%	0%

### 3.1.8.7 Multiple Endpoints within a Single Study

The guidance on clinical evidence notes the importance of statistically persuasive evidence of an effect on more than one endpoint. If such evidence exists, the internal weight of the study is enhanced.

In the previous sections I discussed results for various endpoints, including endpoints related to time-to-delivery from the time of randomization, gestational age at delivery using staggered entry, and for fetal and neonatal losses. The results for each endpoints appears to be related to the gestational age at the time of randomization, which appears to be confounded with study center.

## 3.2 Surrogate Outcome

The endpoint of deliveries prior to 35<sup>0</sup> weeks gestation is considered a surrogate outcome for fetal loss, and for neonatal mortality and morbidity.

Because of the single study-related issues I have raised and their impact on the interpretation of the results of the analyses, the surrogate outcomes should be used cautiously.

### **3.3 Evaluation of Safety**

Fetal and neonatal losses occurred earlier among 17P-treated women than among Placebo-treated women. Based on the patterns of gestational ages at delivery, the early losses apparently are related to the gestational age at which study drug was started. However, this pattern may also be related to study center.

For instance, the first delivery at Center 2 took place at Week 28. The distribution of gestational ages at randomization is skewed towards older ages. Relative to other centers, Center 2 is under-represented among gestational ages of 18 weeks or earlier.

By contrast, Center 8 (University of Alabama) is over-represented in this age group. Its earliest delivery – a fetal loss – occurred around 18 weeks gestation. In addition, the University of Alabama accounts for around 50% of the losses which occurred at 18 weeks or earlier.

## **4. SUMMARY AND CONCLUSIONS**

### **4.1 Statistical Issues and Collective Evidence**

The statistical issues include the following:

- The submission contains a single study to support the claim of effectiveness.
- Study site and gestational age at randomization appear to be confounded.
- Deaths occur earlier among women randomized to 17P.
- The study was not designed for drug approval. Consequently, the study is not powered for the endpoints of interest.

Deliveries prior to 35 weeks gestation is a surrogate outcome for fetal loss and for neonatal morbidity and mortality. Because this endpoint does not appear to reach the level of evidence needed for a single study, its results should be interpreted cautiously, especially when extrapolating to mortality and morbidity.

### **4.2 Conclusions and Recommendations**

From a statistical perspective, the level of evidence from Study 17P-CT002 is not sufficient to support the effectiveness of 17P. The primary reason is the absence of a second, confirmatory study. Without a second study, the generalizability of the study results to a larger population cannot be assessed.

This submission contains a single study to support the claim of effectiveness of 17P. Prior to Study 17P-CT002 another study was initiated but was halted due to drug product

manufacturing issues. Because of its small size and issues regarding drug potency, I did not review that study.

Study 17P-CT002 was stopped after the second interim analysis, which showed that Delivery <37 weeks gestation had met the stopping rules in favor of 17P. Subsequently, analyses showed that Delivery < 35 weeks gestation and Delivery <32 weeks gestation were statistically significant when accounting for the interim analyses.

Study 17P-CT002 was not designed for drug approval. FDA and the applicant did not have the usual meetings and discussions regarding the choice of endpoint needed to establish efficacy in a regulatory environment. As a result, the primary endpoint for the study – Delivery <37 weeks gestation – is not what the FDA would have advised.

After the results of the study were published, the FDA and the applicant discussed the analyses that would be submitted as part of an NDA. The FDA requested analyses of Delivery <35 weeks gestation as the primary basis of approval. The Advisory Committee reiterated the clinical importance of this endpoint as a preferred surrogate for neonatal morbidity and mortality.

Although the results are statistically significant for Delivery < 35 weeks gestation and Delivery <32 weeks gestation when accounting for interim analyses, the confidence intervals for the treatment effects are not convincing when considering that only one study was submitted to support the claim of effectiveness for 17P.

“Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products” sets forth guidance needed for the FDA to accept results from a single, clinical study. Using the guidance document, I focused my review on whether the results could be generalized to a larger population, or not.

The guidance on clinical evidence stresses the importance of a large multi-center study to help establish the credibility of a single study submission and that the credibility is enhanced if no single center accounts for an unusually large proportion of the subjects.

When compared with all other centers, one center, the University of Alabama, is disproportionately represented in the study. The University of Alabama accounts for about 25% of all subjects enrolled (126/463) and is about three times the size of the next largest center, the University of Tennessee (45/463 = 9.7%).

The effect of 17P is most pronounced when started at 18 weeks gestation or earlier and does not appear effective when started at 20 weeks of gestation or later. The rate of fetal and neonatal deaths is also most pronounced among women who started study drug at 18 weeks gestation or earlier (10%). The rate decreases to 2% when study drug is started at 20 weeks of gestation or later.

These results need to be interpreted in the larger context of confounding with study

center. The results of my analyses suggest the presence of confounding between center and gestational age at randomization. For example, the University of Alabama accounts for 44% of subjects enrolled at 18 weeks gestation or earlier and had relatively few patients at later ages. At other centers, the gestational age at randomization is skewed towards later gestational ages at the time of randomization.

Moreover, the University of Alabama accounts for about 50% of the fetal and neonatal deaths that occurred among women who started study drug at 18 weeks of gestation or earlier.

Thus, the apparent age trends in treatment effect, and fetal and neonatal deaths simply could be unique to the patient population enrolled at the University of Alabama.

When two studies are submitted, the chance of both studies yielding a false positive result is 1/1600. In the case of a single study, the results must be less than a nominal p-value of 0.00125 to ensure the same false positive rate. In Study 17P-CT002, the only endpoint that meets this criterion is Delivery <37 weeks gestation. Deliveries at times earlier than 37 weeks gestation were not statistically significant at 0.001. The results of the analyses of the 32 and 35 week endpoints suggest their false positive rates could be as great as 1/40.

Because of the public health need for a drug product to prevent preterm deliveries, we might be willing to accept a false positive rate that is somewhat greater than 1/1600 if the results appear to be generalizable. However, because of the issues introduced by the size of the University of Alabama and its findings, together with the 1/40 false positive rates for the 32 and 35 week endpoints, I do not believe the study results can be generalized to a larger population.

Therefore, from a statistical perspective, I do not believe this study meets the level of evidence needed to support the efficacy of 17P.

**5. APPENDICES**

**5.1 Listings of Kaplan-Meier Estimates of Time to Delivery,  
Incorporating Gestational Age at Randomization as a Left-censored Variable**

Adeza Biomedical  
 Study No.: 17P-CT-002  
 Final Analysis  
 Listing 1 - Kaplan-Meier Estimates of Time to Delivery Using Staggered Entry

Treatment=17P

GA at Delivery	KM Survival	Number at Risk	Number Delivered
16.00	1.0000	3	0
18.14	0.9917	120	1
18.29	0.9837	125	1
18.86	0.9771	148	1
19.14	0.9652	164	2
20.14	0.9564	220	2
20.29	0.9439	229	3
20.71	0.9405	277	1
21.00	0.9373	298	1
21.14	0.9341	297	1
22.14	0.9310	295	1
22.57	0.9278	294	1
24.71	0.9246	293	1
25.00	0.9215	292	1
25.14	0.9183	291	1
25.43	0.9120	290	2
25.57	0.9056	288	2
26.00	0.9025	286	1
26.57	0.8993	285	1
26.86	0.8930	284	2
27.71	0.8898	282	1
29.29	0.8866	281	1
30.00	0.8835	280	1
30.14	0.8771	279	2
30.57	0.8708	277	2
31.57	0.8676	275	1
31.86	0.8645	274	1
32.00	0.8613	273	1
32.71	0.8581	272	1
32.86	0.8550	271	1
33.14	0.8518	270	1
33.29	0.8423	269	3
33.43	0.8360	266	2
33.71	0.8233	264	4
33.86	0.8202	260	1
34.14	0.8138	259	2
34.29	0.8107	257	1
34.43	0.8075	256	1
34.57	0.8011	254	2
34.71	0.7852	252	5
34.86	0.7757	247	3
35.00	0.7725	244	1
35.14	0.7566	243	5
35.29	0.7534	238	1
35.43	0.7471	237	2
35.57	0.7375	235	3
35.71	0.7312	232	2

NOTE: 4 17P patients are censored at times 18.57, 22.00, 34.43, and 36.57

Adeza Biomedical  
 Study No.: 17P-CT-002  
 Final Analysis  
 Listing 1 - Kaplan-Meier Estimates of Time to Delivery Using Staggered Entry

Treatment=17P

GA at Delivery	KM Survival	Number at Risk	Number Delivered
35.86	0.7248	230	2
36.00	0.7121	228	4
36.14	0.7026	224	3
36.29	0.6962	221	2
36.43	0.6899	219	2
36.57	0.6676	217	7
36.71	0.6389	209	9
36.86	0.6229	200	5
37.00	0.5909	195	10
37.14	0.5750	185	5
37.29	0.5430	180	10
37.43	0.5239	170	6
37.57	0.5079	164	5
37.71	0.4791	159	9
37.86	0.4504	150	9
38.00	0.4248	141	8
38.14	0.3929	133	10
38.29	0.3610	123	10
38.43	0.3322	113	9
38.57	0.3194	104	4
38.71	0.2971	100	7
38.86	0.2715	93	8
39.00	0.2428	85	9
39.14	0.2172	76	8
39.29	0.1917	68	8
39.43	0.1885	60	1
39.57	0.1533	59	11
39.71	0.1437	48	3
39.86	0.1214	45	7
40.00	0.1054	38	5
40.14	0.0831	33	7
40.29	0.0639	26	6
40.43	0.0415	20	7
40.57	0.0351	13	2
40.71	0.0319	11	1
40.86	0.0287	10	1
41.00	0.0256	9	1
41.14	0.0224	8	1
41.29	0.0096	7	4
41.43	0.0064	3	1
41.57	0.0032	2	1
41.71	0.0000	1	1

NOTE: 4 17P patients are censored at times 18.57, 22.00, 34.43, and 36.57

Study No.: 17P-CT-002

Final Analysis

Listing 1 - Kaplan-Meier Estimates of Time to Delivery Using Staggered Entry

Treatment=PLACEBO

GA at Delivery	KM Survival	Number at Risk	Number Delivered
36.57	0.5026	83	6
36.71	0.4765	77	4
36.86	0.4504	73	4
37.00	0.4439	69	1
37.14	0.4308	68	2
37.29	0.4112	66	3
37.57	0.3982	63	2
37.71	0.3721	61	4
37.86	0.3590	57	2
38.00	0.3460	55	2
38.14	0.3264	53	3
38.29	0.3068	50	3
38.43	0.2676	47	6
38.57	0.2350	41	5
38.71	0.2285	36	1
38.86	0.1958	35	5
39.00	0.1697	30	4
39.14	0.1501	26	3
39.29	0.1436	23	1
39.43	0.1371	22	1
39.57	0.0914	21	7
39.71	0.0783	14	2
39.86	0.0718	12	1
40.00	0.0587	11	2
40.14	0.0522	9	1
40.29	0.0326	8	3
40.43	0.0261	5	1
40.71	0.0131	4	2
41.71	0.0065	2	1
41.86	0.0000	1	1

NOTE: No placebo patients are censored

Adeza Biomedical  
 Study No.: 17P-CT-002  
 Final Analysis  
 Listing 1 - Kaplan-Meier Estimates of Time to Delivery Using Stagger

Treatment=PLACEBO

GA at Delivery	KM Survival	Number at Risk	Number Delivered
36.57	0.5026	83	6
36.71	0.4765	77	4
36.86	0.4504	73	4
37.00	0.4439	69	1
37.14	0.4308	68	2
37.29	0.4112	66	3
37.57	0.3982	63	2
37.71	0.3721	61	4
37.86	0.3590	57	2
38.00	0.3460	55	2
38.14	0.3264	53	3
38.29	0.3068	50	3
38.43	0.2676	47	6
38.57	0.2350	41	5
38.71	0.2285	36	1
38.86	0.1958	35	5
39.00	0.1697	30	4
39.14	0.1501	26	3
39.29	0.1436	23	1
39.43	0.1371	22	1
39.57	0.0914	21	7
39.71	0.0783	14	2
39.86	0.0718	12	1
40.00	0.0587	11	2
40.14	0.0522	9	1
40.29	0.0326	8	3
40.43	0.0261	5	1
40.71	0.0131	4	2
41.71	0.0065	2	1
41.86	0.0000	1	1

NOTE: No placebo patients are censored

## 5.2 Prevention of Delivery <math><37^0</math>, <math><35^0</math>, <math><32^0</math> weeks, by Center

Table 1 Question 1.pdf: Response to FDA Question 1 October 6, 2006

TABLE 1 DELIVERY <37, <35, <32 BY CENTER		17 OHP			PLACEBO		
CENTER NAME	DELIVERY	n/N	%	n/N	%	n/N	%
Univ. of Pittsburgh	<37	5/24	20.8%	11/12	91.7%		
	<35	2/24	8.3%	8/12	66.7%		
	<32	1/24	4.2%	3/12	25%		
Univ. of Tennessee	<37	13/30	43.3%	9/15	60%		
	<35	8/30	26.7%	6/15	40%		
	<32	4/30	13.3%	3/15	20%		
Univ. of Alabama	<37	23/86	26.7%	18/40	45%		
	<35	15/86	17.4%	11/40	27.5%		
	<32	9/86	10.5%	10/40	25%		
Wayne State Univ.	<37	5/16	31.3%	5/8	62.5%		
	<35	2/16	12.5%	4/8	50%		
	<32	1/16	6.3%	2/8	25%		
Univ. of Cincinnati	<37	3/9	33.3%	2/4	50%		
	<35	3/9	33.3%	2/4	50%		
	<32	1/9	11.1%	1/4	25%		
Bowman Gray Sch. Of Med.	<37	7/13	53.8%	7/9	77.8%		
	<35	3/13	23.1%	3/9	33.3%		
	<32	2/13	15.4%	3/9	33.3%		
Univ. of Chicago	<37	3/5	60%	1/2	50%		
	<35	2/5	40%	0/2	0%		
	<32	2/5	40%	0/2	0%		

TABLE 1 DELIVERY <37, <35, <32 BY CENTER		17 OHP		PLACEBO	
CENTER NAME	DELIVERY	n/N	%	n/N	%
Ohio State Univ.	<37	11/20	55%	4/8	50%
	<35	6/20	30%	1/8	12.5%
	<32	4/20	20%	1/8	12.5%
Univ. of Miami	<37	2/7	28.6%	0/4	0%
	<35	2/7	28.6%	0/4	0%
	<32	1/7	14.3%	0/4	0%
Univ. of Texas SW Med Ctr.	<37	12/28	42.9%	8/11	72.7%
	<35	7/28	25%	2/11	18.2%
	<32	4/28	14.3%	1/11	9.1%
Univ. of Texas Hlth Sci Ctr	<37	3/8	37.5%	3/5	60%
	<35	1/8	12.5%	0/5	0%
	<32	0/8	0%	0/5	0%
Univ. of Utah Med Ctr.	<37	11/29	37.9%	7/14	50%
	<35	7/29	24.1%	3/14	21.4%
	<32	1/29	3.4%	0/14	0%
Thomas Jefferson Univ.	<37	10/17	58.8%	3/7	42.9%
	<35	4/17	23.5%	2/7	28.6%
	<32	3/17	17.6%	2/7	28.6%
Brown Univ.	<37	1/3	33.3%	1/2	50%
	<35	1/3	33.3%	0/2	0%
	<32	1/3	33.3%	0/2	0%

Table 1 Question 1.pdf: Response to FDA Question 1 October 6, 2006

CENTER NAME	17 OHP			PLACEBO		
	DELIVERY	n/N	%	n/N	%	%
Columbia Univ.	<37	2/6	33.3%	1/5	20%	20%
	<35	2/6	33.3%	1/5	20%	20%
	<32	2/6	33.3%	1/5	20%	20%
Case Western Univ.	<37	2/4	50%	1/2	50%	50%
	<35	0/4	0%	1/2	50%	50%
	<32	0/4	0%	0/2	0%	0%
Univ. of Texas Houston	<37	0/2	0%	1/2	50%	50%
	<35	0/2	0%	1/2	50%	50%
	<32	0/2	0%	1/2	50%	50%
Univ. of NC Chapel Hill	<37	1/2	50%	1/2	50%	50%
	<35	1/2	50%	1/2	50%	50%
	<32	1/2	50%	1/2	50%	50%
Northwestern Univ.	<37	1/1	100%	1/1	100%	100%
	<35	0/1	0%	1/1	100%	100%
	<32	0/1	0%	1/1	100%	100%
All Centers minus Alabama	<37	92/224	41.1%	66/113	58.4%	58.4%
	<35	51/224	22.8%	36/113	31.9%	31.9%
	<32	28/224	12.5%	20/113	17.7%	17.7%

		17 OHP		PLACEBO	
CENTER NAME	DELIVERY	n/N	%	n/N	%
All Centers	<37	115/310	37.1%	84/153	54.9%
	<35	66/310	21.3%	478/153	30.7%
	<32	37/310	11.9%	30/153	19.6%

### **5.3 Time-to-delivery, Incorporating Gestational Age at Randomization as a Left-censored Variable, by Center**

Figure 1a  
Time to Delivery - Center 2

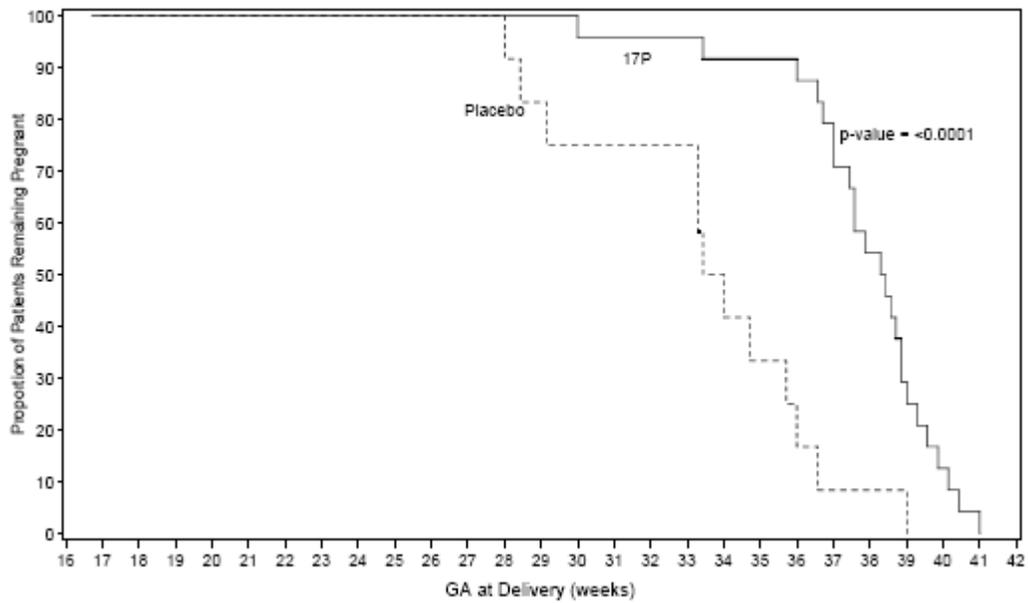


Figure 1b  
Time to Delivery - Center 4

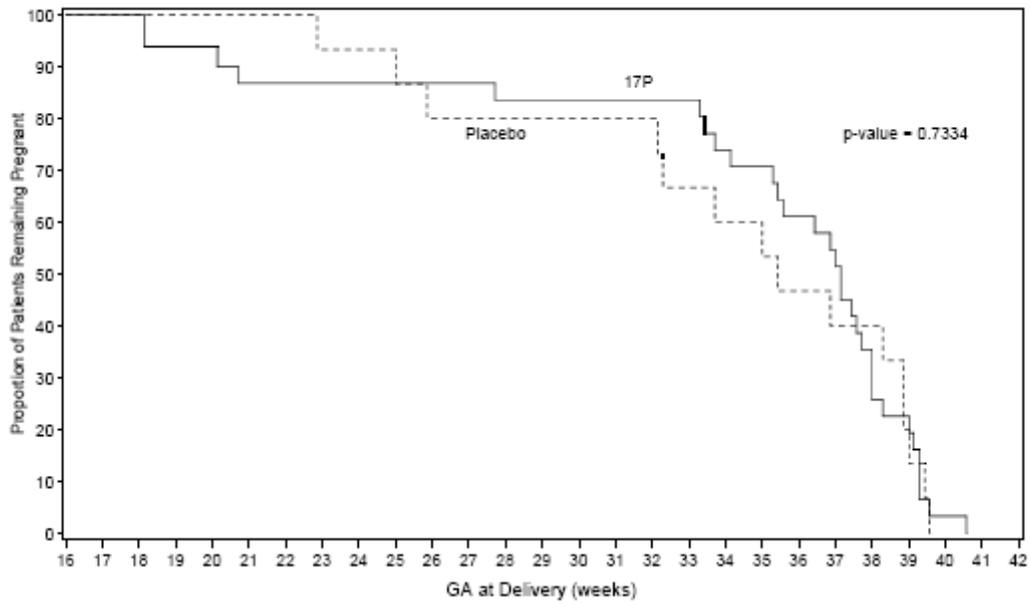


Figure 1c  
Time to Delivery - Center 8

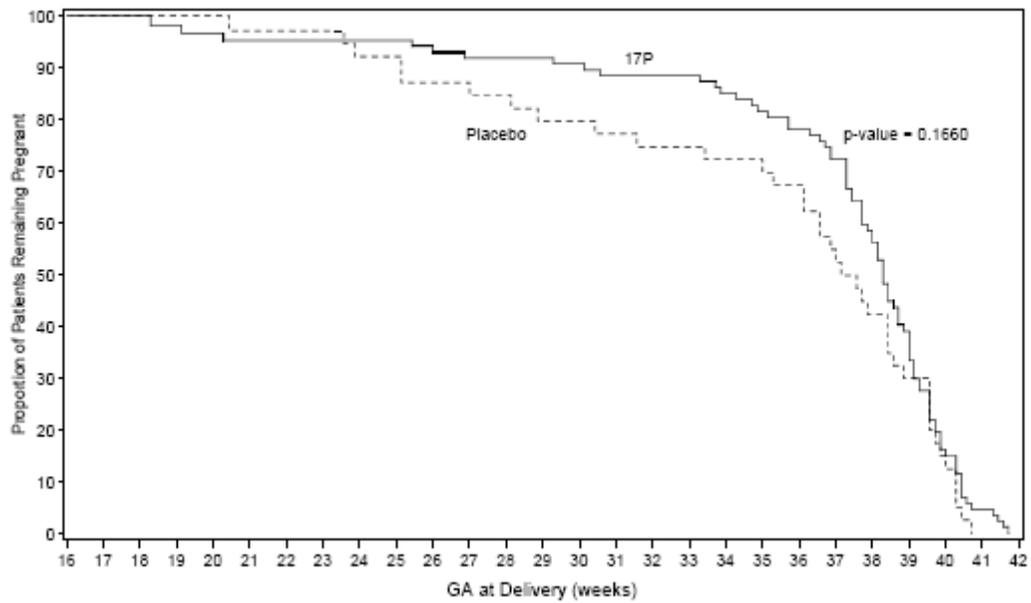


Figure 1d  
Time to Delivery - Center 9

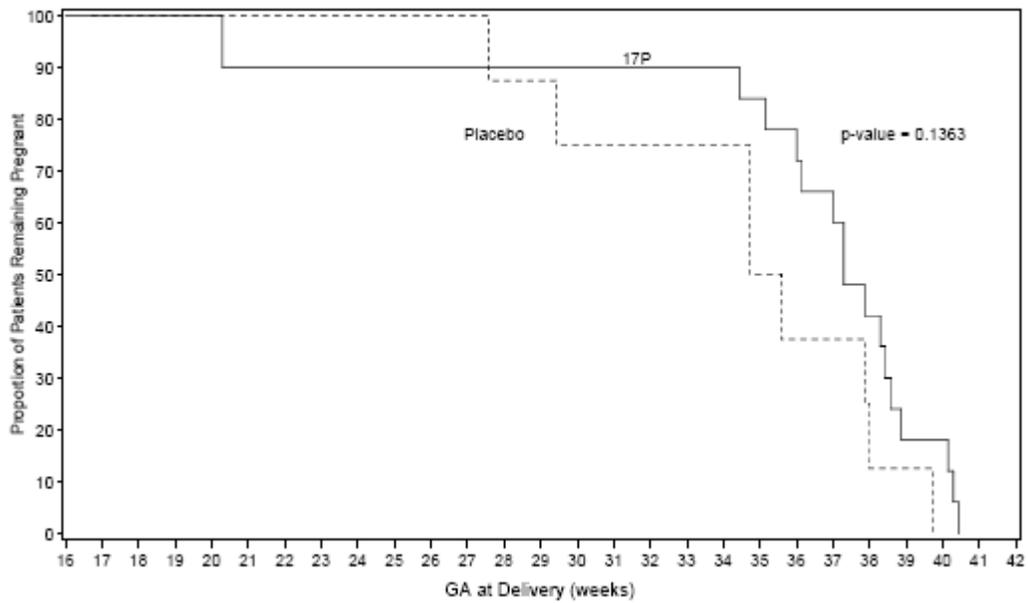


Figure 1e  
Time to Delivery - Center 11

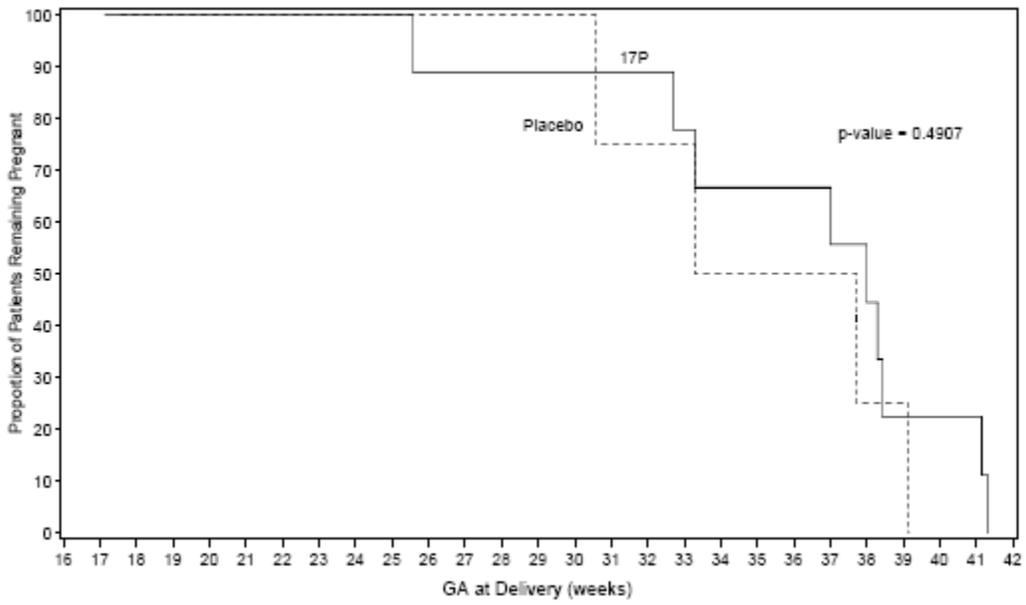


Figure 1f  
Time to Delivery - Center 13

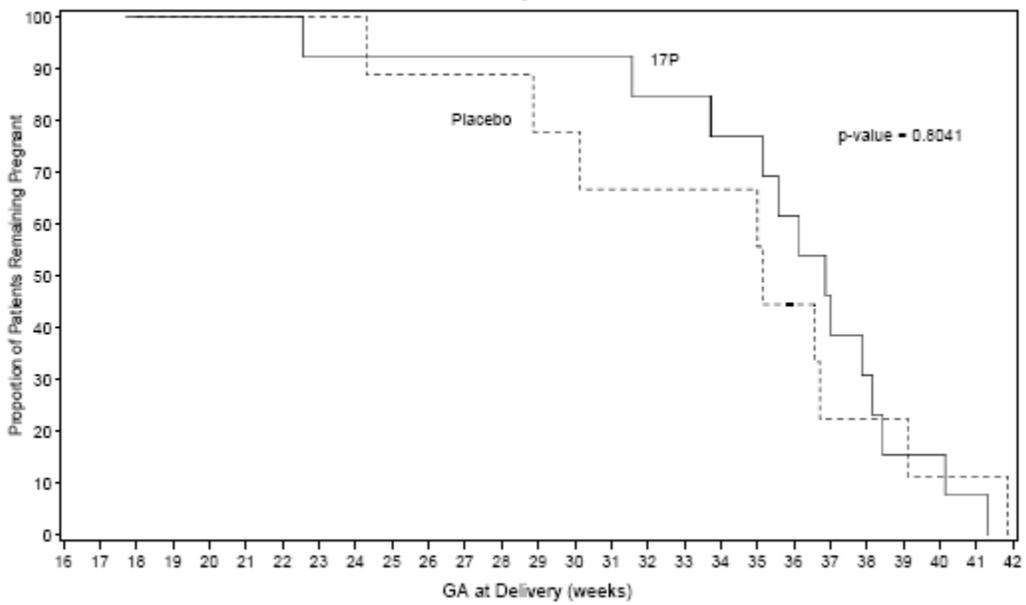


Figure 1g  
Time to Delivery - Center 14

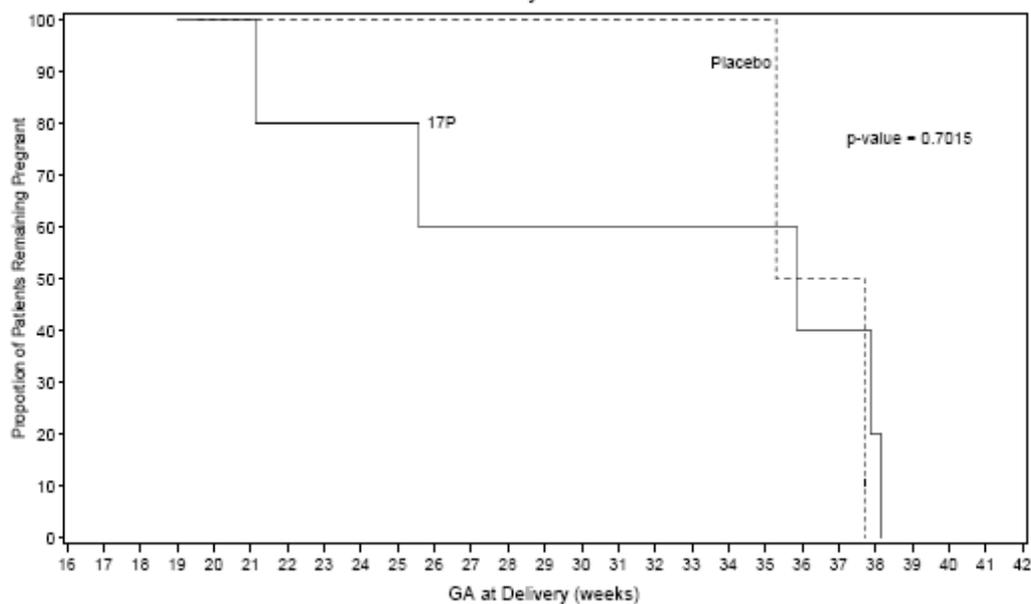


Figure 1h  
Time to Delivery - Center 15

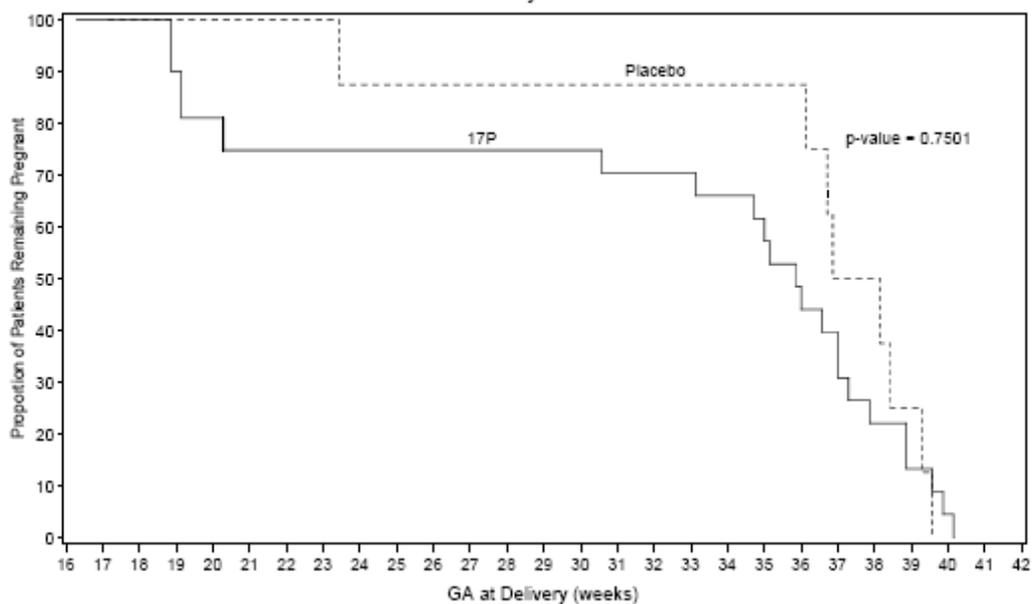


Figure 1j  
Time to Delivery - Center 17

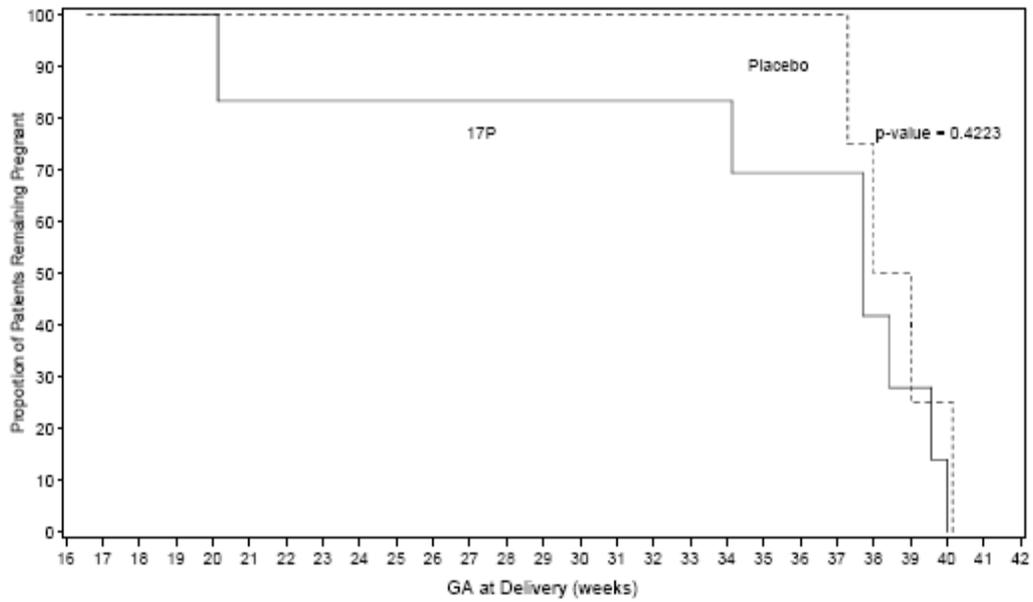


Figure 1j  
Time to Delivery - Center 18

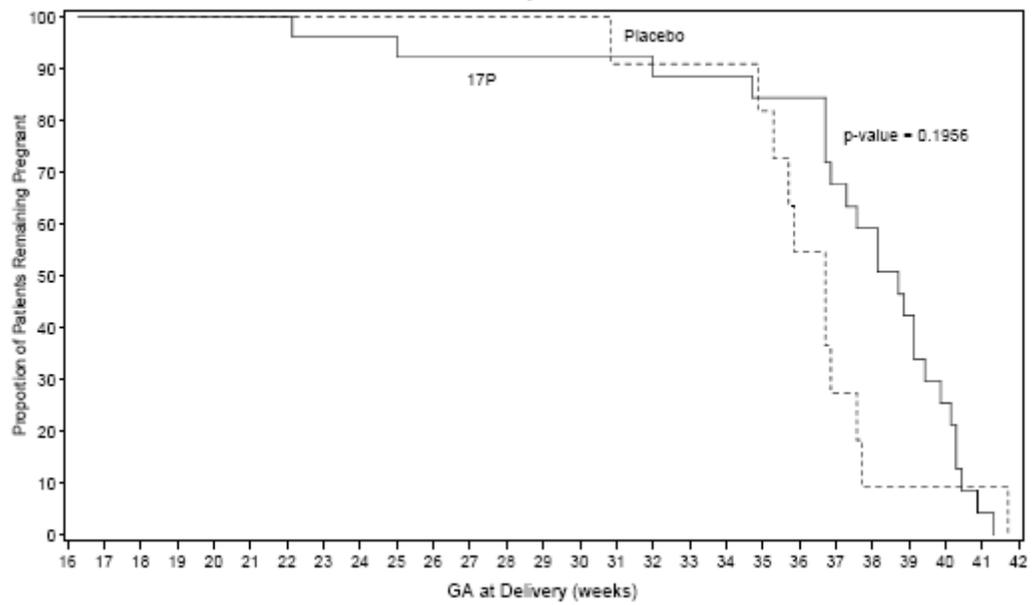


Figure 1k  
Time to Delivery - Center 19

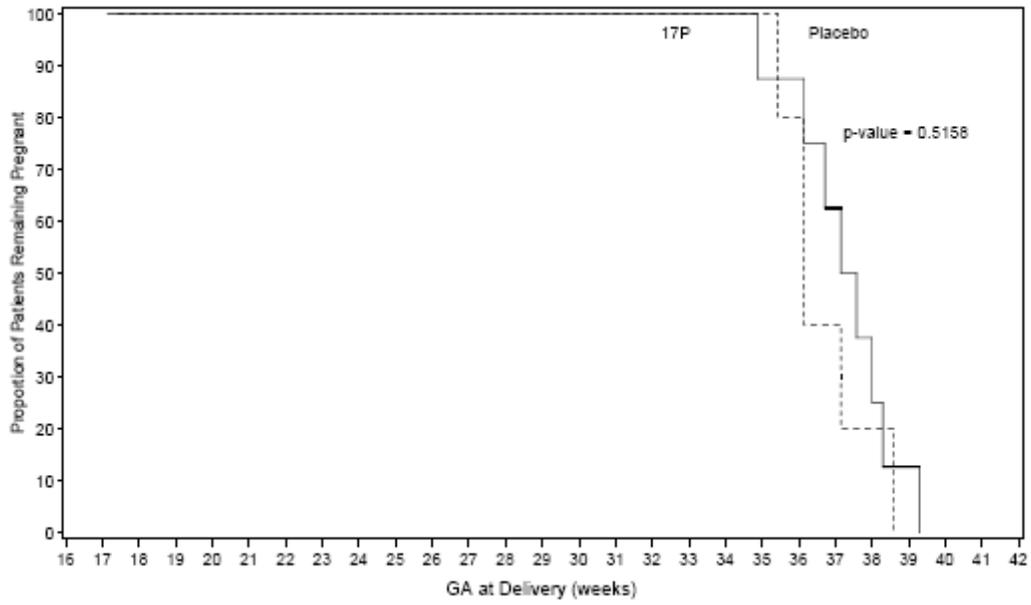
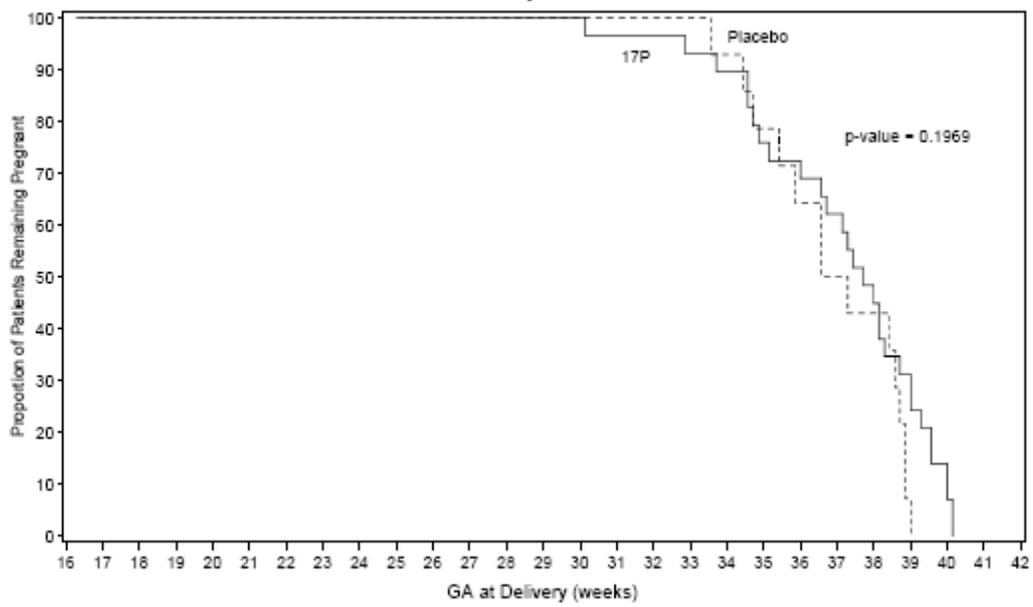


Figure 1l  
Time to Delivery - Center 20



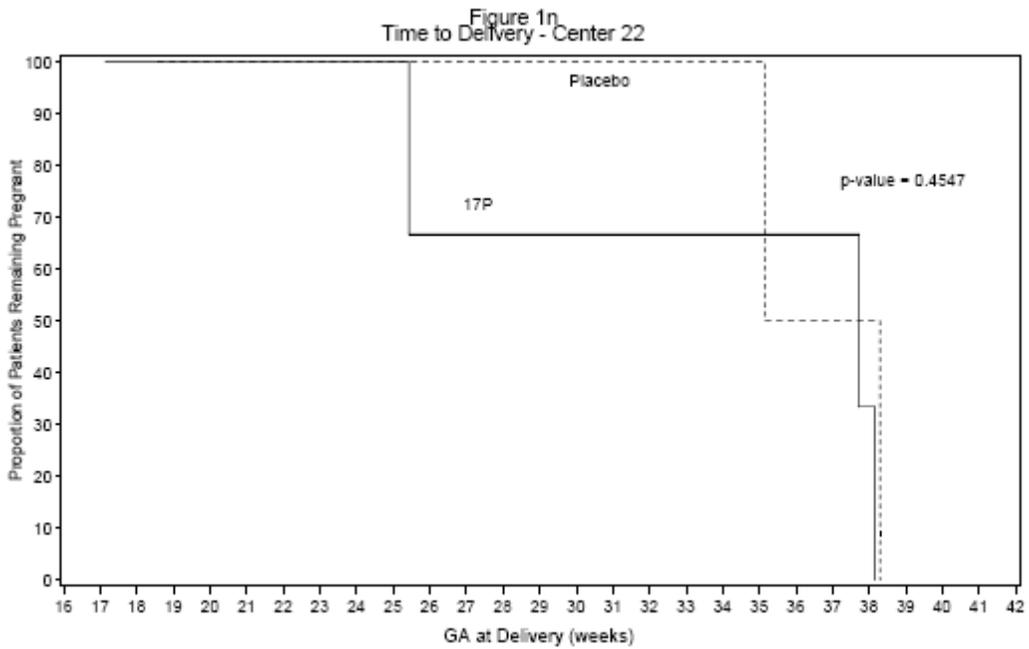
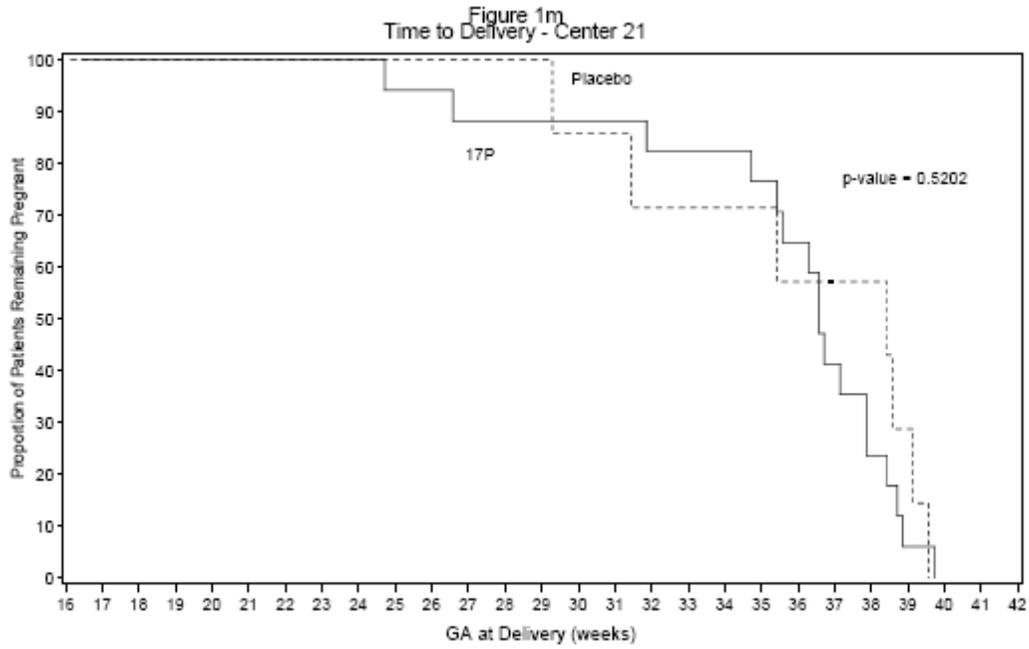


Figure 1o  
Time to Delivery - Center 23

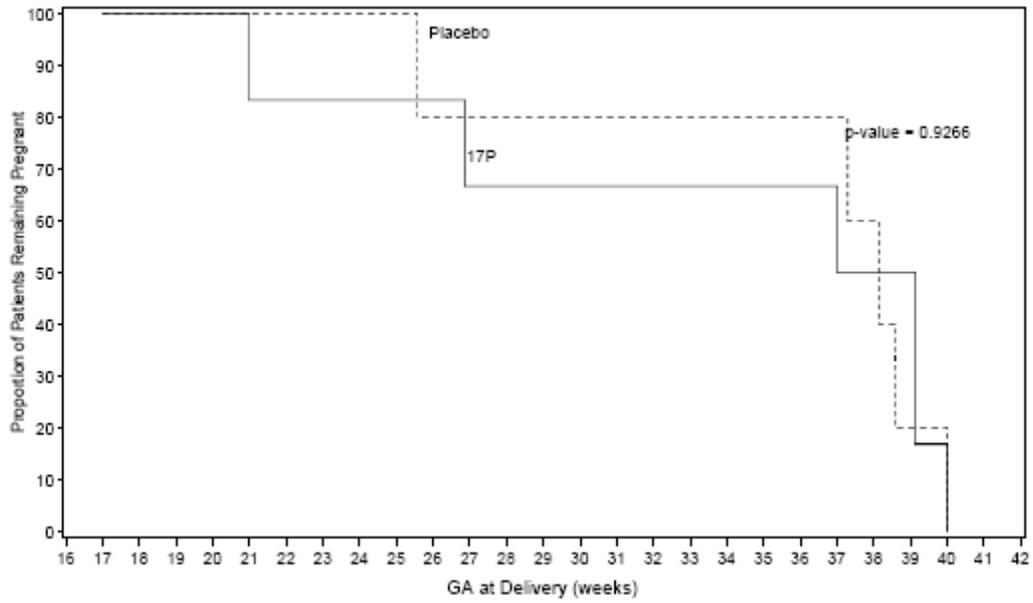


Figure 1p  
Time to Delivery - Center 25

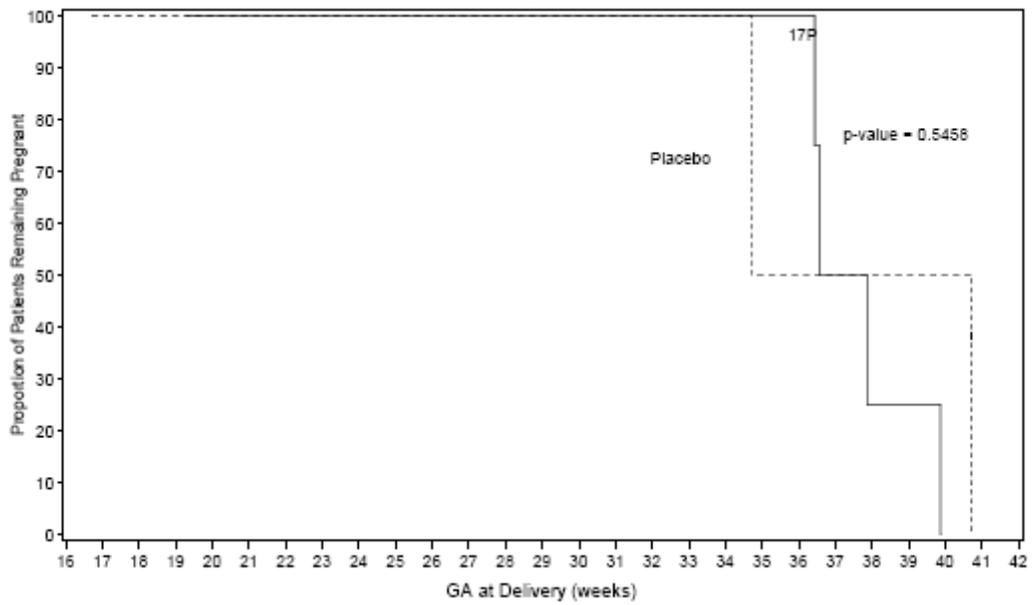


Figure 1q  
Time to Delivery - Center 26

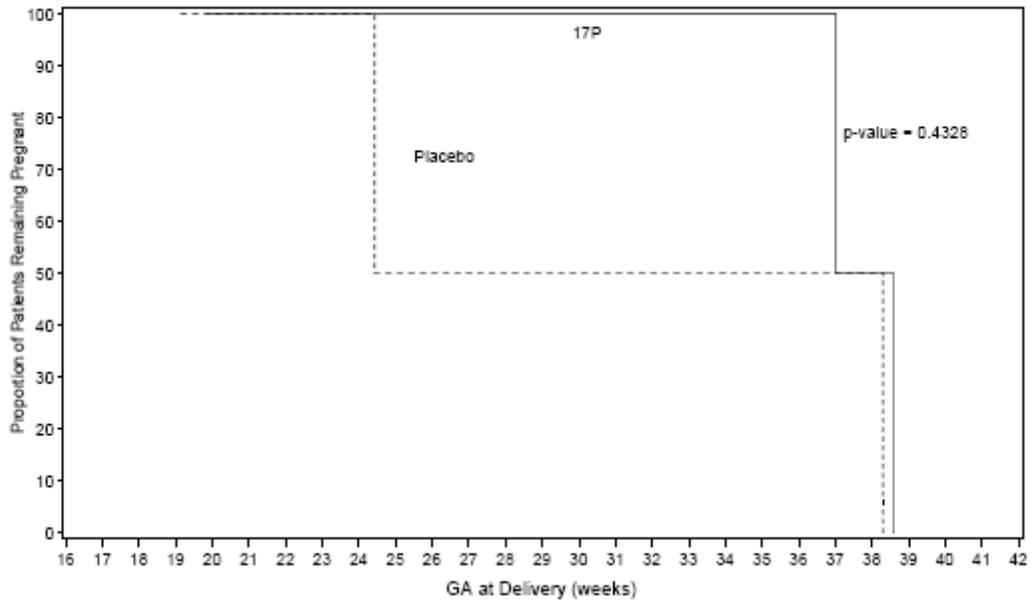
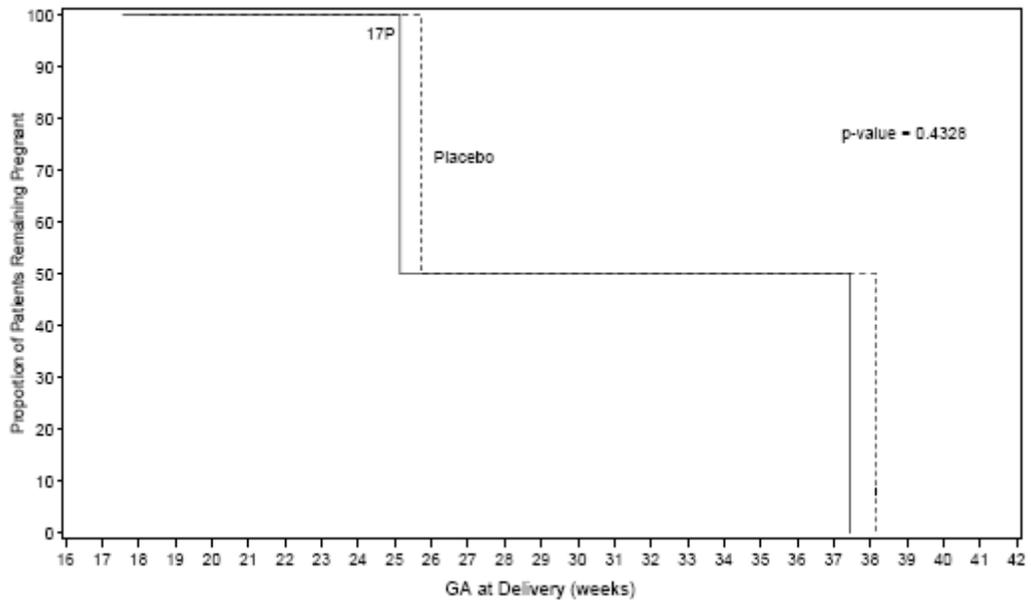
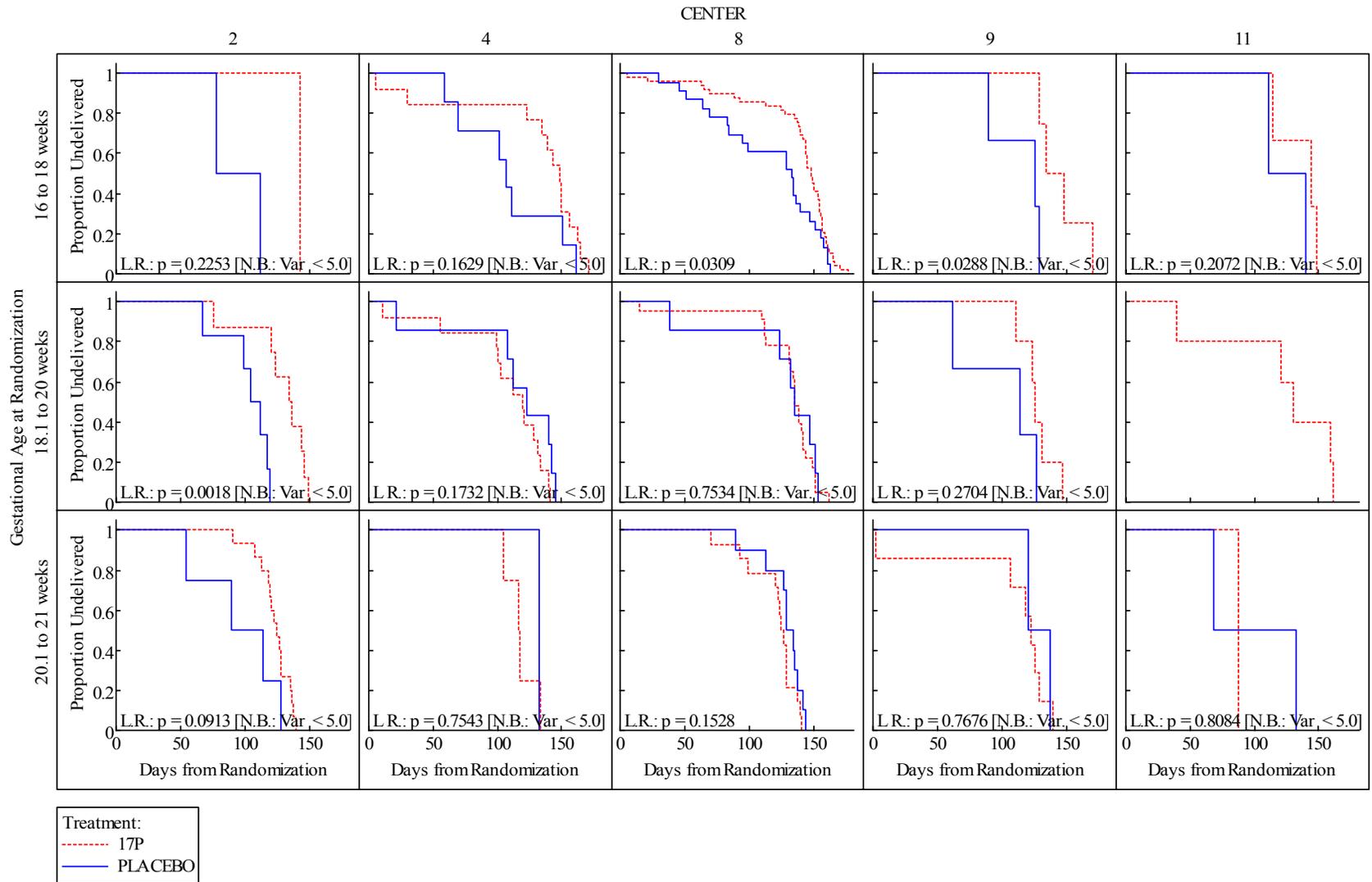
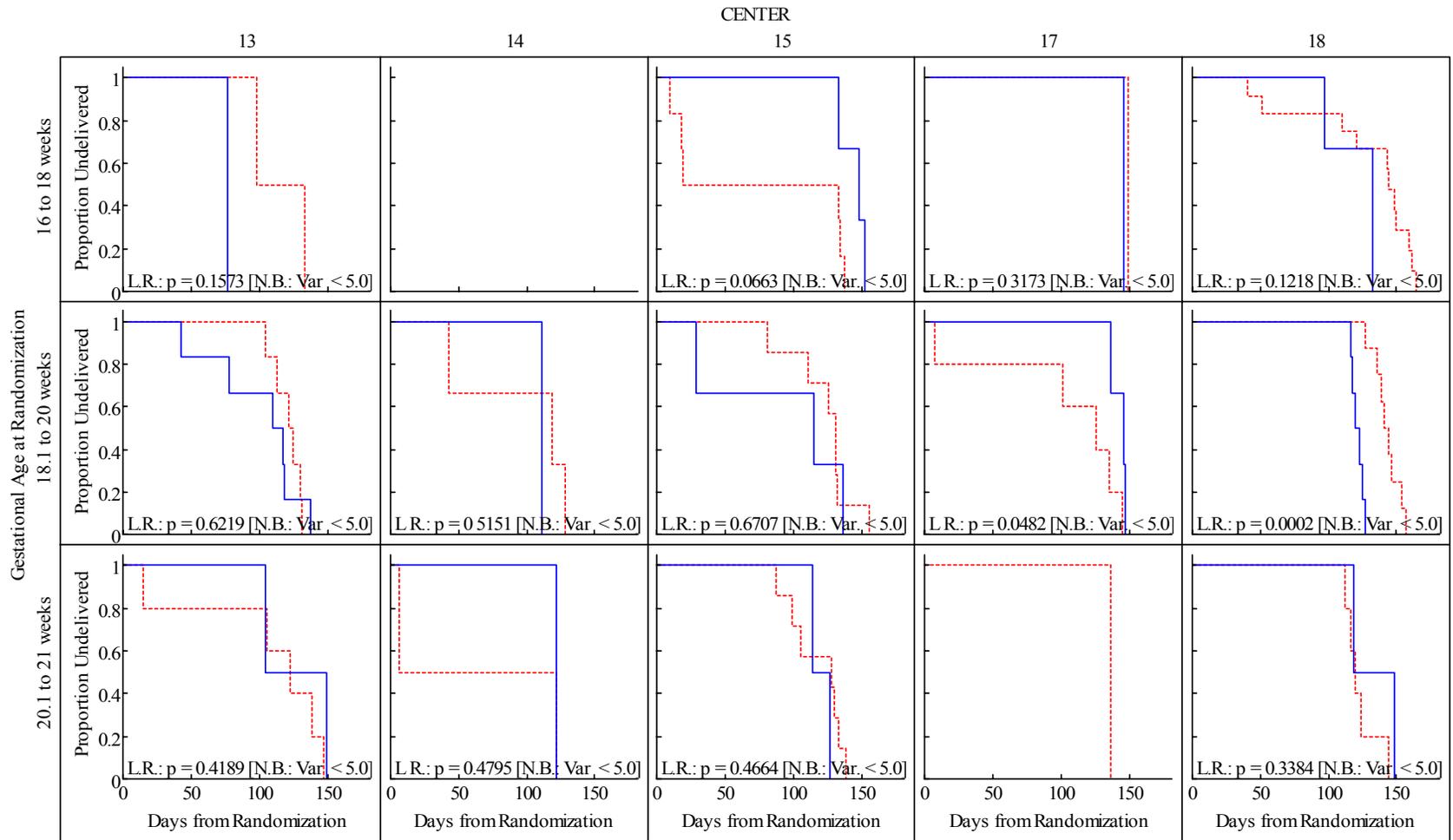


Figure 1r  
Time to Delivery - Center 27

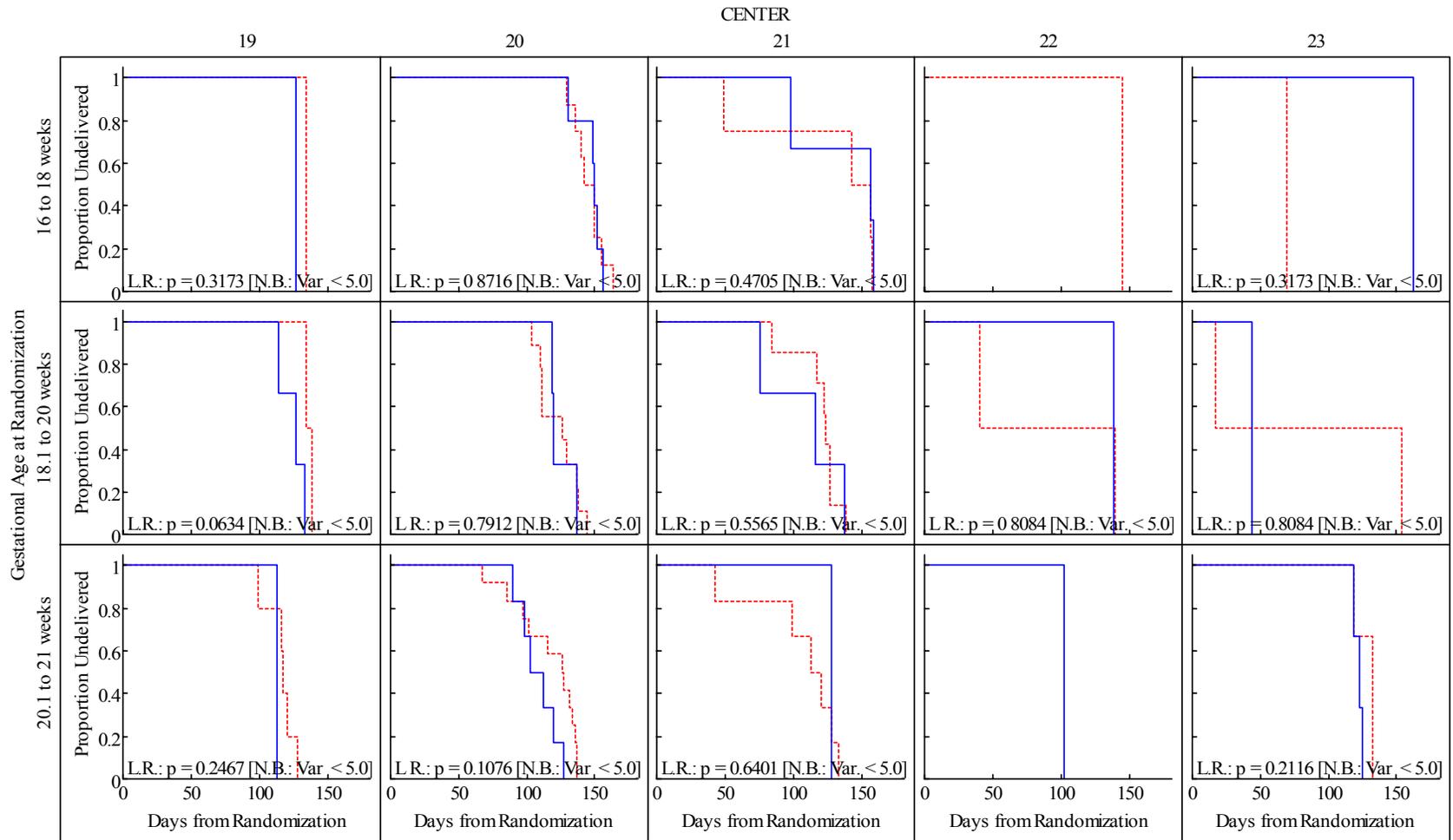


#### **5.4 Time-to-delivery from Randomization, by Center and by Gestational Age at Randomization**

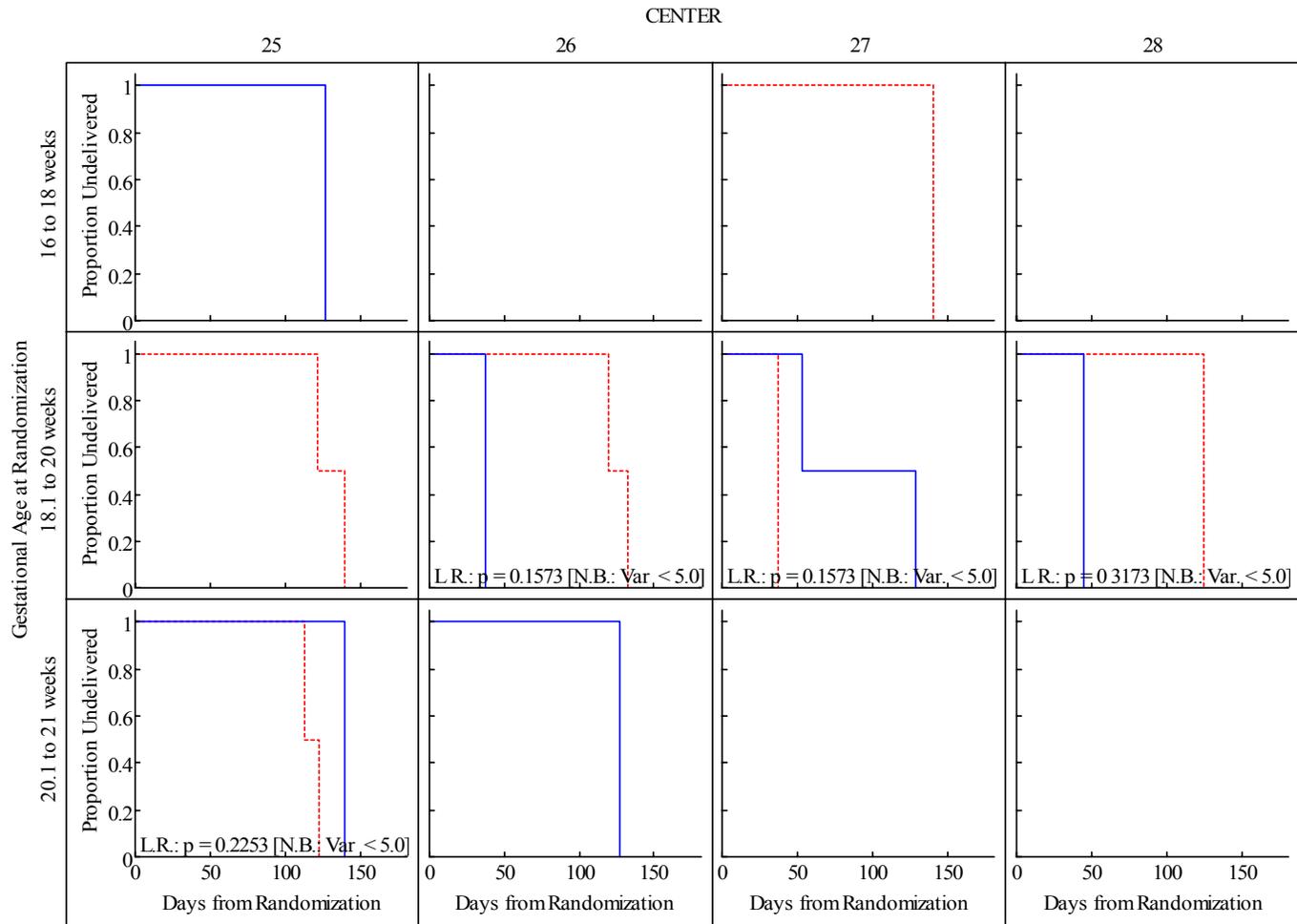




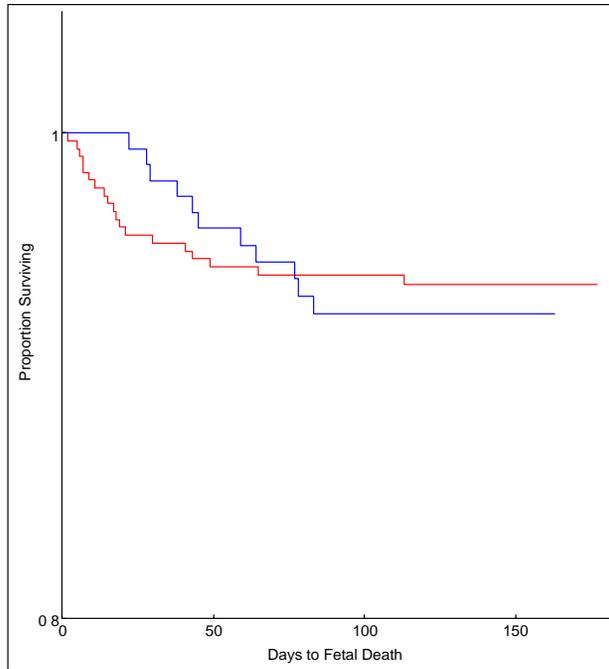
Treatment:  
 - - - 17P  
 — PLACEBO



Treatment:  
 - - - 17P  
 — PLACEBO



Treatment:  
 - - - 17P  
 — PLACEBO



TREATMEN:  
— 17P  
— PLACEBO

-----  
**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  
10/19/2006 03:40:38 PM  
BIOMETRICS