Appendix A. Supplementary Data and Information

Estetrol Co-treatment of Androgen Deprivation Therapy in Infiltrating or Metastatic, Castration-Sensitive Prostate Cancer: A Randomized, Double-blind, Phase II Clinical Trial (PCombi)

Herjan J.T. Coelingh Bennink[a,*], Jeroen A. van Moorselaar[b], E. David Crawford[c], Erik P.M. Roos[d], Diederik M. Somford[e], Ton A. Roeleveld[f], Tjard D. de Haan[g], Harm H.E. van Melick[h], Yacov Reisman[i], Yvette Zimmerman[a], Gonnie van Osta[j], Jan Krijgh[a], Neal D. Shore[k], Fred Saad[l], Andrew V. Schally[m], Frans M.J. Debruyne[n]

[a] Pantarhei Oncology, Zeist, The Netherlands;[b] Amsterdam UMC, Free University, Amsterdam, The Netherlands; [c] University of California San Diego, La Jolla, CA, USA; [d] Antonius Hospital, Sneek, The Netherlands; [e] Canisius-Wilhelmina Hospital, Nijmegen, The Netherlands; [f] North-West Hospital, Alkmaar, The Netherlands; [g] Isala Hospital, Zwolle, The Netherlands; [h] St. Antonius Hospital, Nieuwegein, The Netherlands; [i] Amstelland Hospital, Amstelveen, The Netherlands; [j] Author! et al, Hilversum, The Netherlands;[k] Carolina Urologic Research Center, Myrtle Beach, SC, USA; [l] University of Montreal Hospital Center, Montreal, QC, Canada; [m] University of Miami, FL, USA; [n] Andros Clinics, Arnhem, The Netherlands

* Corresponding author:

Herjan JT Coelingh Bennink, MD, PhD, Pantarhei Oncology BV, P.O. Box 464, 3700 AL Zeist, The Netherlands, or at hcb@pantarheibio.com.

ORCID: 0000-0003-1550-94

Table of Contents

Principal investigators participating in the PCombi study (PR3109)

Site #	Principal investigator	Institution
101*	Prof F (Frans) Debruyne f.debruyne@andros.nl	Andros Men's Health Institutes P.O. Box 30005 6803 AA Arnhem The Netherlands
102*	Dr H. (Harm) van Melick h.van.melick@antoniusziekenhuis.nl	Department of Urology Antonius Hospital Koekoekslaan 1 3435 CM Nieuwegein The Netherlands
103	Dr E. (Erik) Roos e.roos@antonius-sneek.nl	Department of Urology Antonius Hospital Bolswarderbaan 1 8601 ZK Sneek The Netherlands
104	Dr R (Rik) Somford r.somford@cwz.nl	Department of Urology Canisius Wilhelmina Hospital Weg door Jonkerbos 100 6532 SZ Nijmegen The Netherlands
105*	Dr Y. (Cobi) Reisman uro.amsterdam@gmail.com	Department of Urology Amstelland Hospital Laan van de Helende Meesters 8 1186 AM Amstelveen The Netherlands
106*	Prof Dr JA (Fred) Witjes Fred.witjes@radboudumc.nl	Department of Urology Radboud University Medical Center Geert Grooteplein Zuid 10 6525 GA Nijmegen The Netherlands
107	Dr M (Martijn) Steffens m.steffens@isala.nl Dr T (Tjard) de Haan (subinvestigator) t.d.de.haan@isala.nl	Department of Urology Isala Hospital Dokter van Heesweg 2 8025 AB Zwolle The Netherlands
108	Dr T (Ton) Roeleveld t.a.roeleveld@nwz.nl	Department of Urology North-West Hospital Wilhelminalaan 12 1815 JD Alkmaar The Netherlands

Site #	Principal investigator	Institution
	Coordinating investigator: Prof. Dr. RJA van Moorselaar RJA.vanmoorselaar@vumc.nl	Department of Urology Free University Medical Center De Boelelaan 1117 1081 HV Amsterdam The Netherlands

* no patients randomized

Statistical analysis center

Responsible statistician	Institution
G van Osta, PhD gvosta@author.nl	AUTHOR! Data analysis and reporting Julianalaan 15 1213 AP Hilversum The Netherlands

PROTOCOL PR3109

A double-blind, randomised, placebo-controlled, multi-center study to evaluate effects of estetrol on testosterone suppression and quality of life in prostate cancer patients treated with an LHRH agonist

Investigational Product:	**Estetrol**
Sponsor:	Pantarhei Oncology BV Boslaan 11 3701 CH Zeist Telephone: +31 30 6985020 Telefax: +31 30 6985021
Clinical study protocol:	PCombi
Protocol number:	PR3109
EudraCT number:	2017-003708-34
Version number:	2.0 Final
Version date:	08 November, 2018

4.2 Period of study

Total duration of the study, including screening will be 224 days for each patient (Screening from Day -28 till Day -1, Clinical part from Day 1 till Day 168, Follow up on Day 197).

4.3 Study medication

4.3.1 Formulation, dose and regimen

Estetrol (E4), formulated in 20 mg tablets, will be investigated in patients by a once daily oral administration for 168 days (24 weeks). In this study a dose of 40 mg E4 is investigated, this means that the patient will need to take two tablets of 20 mg E4 daily.

The study medication has to be taken orally, every morning (before 12:00 am) with a glass of water, without regard to food intake. In case a patient forgets to take the study medication, the missed study medication should be taken on the same day as soon as he remembers. When he remembers at the time of the next dose, the missed dose should not be taken anymore. Further intake should be done at the usual time. The patient should not throw away any missed tablets from the blister. They will be instructed to bring the blisters with missed tablets back to the clinic.

Composition of the E4 tablet:

- E4
- Lactose monohydrate
- Modified starch
- Corn starch
- Purified Water
- Povidone 30
- Magnesium Stearate

The placebo tablets will be identical in appearance and composition. Only difference being the absence of the active component E4.

Composition of the placebo tablet:

- Lactose monohydrate
- Modified starch
- Corn starch
- Purified Water
- Povidone 30
- Magnesium Stearate

The study medication (investigational and placebo tablets) are manufactured by Haupt Pharma, Münster, Germany, in accordance with applicable current Good Manufacturing Practice (GMP). The tablets will be blistered by Haupt Pharma. There will be 28 tablets in one blister. The blisters will be labeled and packaged by the pharmacist (See Section 4.3.2 and 4.3.3). Patients will be instructed to take two tablets every day.

4.3.2 Method of treatment allocation

At Visit 2 (Baseline), patients will be randomised to one of the two treatments:

Group 1: E4 group
Group 2: Placebo group

Randomisation will be performed with an allocation ratio of 2:1 between the two treatments (investigational product or placebo treatment). In the E4 group 40 patients will be included, and in the placebo group 20 patients will be included.

Study medication will be packed per medication number according to a randomisation list by the pharmacist (ACE Pharmaceuticals BV, The Netherlands). Each study medication package will have a unique medication number. The pharmacist will send the study medication to the sites. After a patient has been found eligible, a medication number will be assigned to the patient by the eCRF system. The investigator will receive a sealed envelope for each medication number specifying the treatment, which is to be opened only in case of emergency.

4.3.3 Packaging and labelling

- E4 treatment will be supplied as a tablet and will be packed in blisters of 28 tablets
- Placebo treatment will be supplied as a tablet and will be packed in blisters of 28 tablets

Study medication will be packed per patient. Packaging and labeling of all study medication will be the responsibility of ACE pharmaceuticals, Zeewolde, the Netherlands. Two boxes containing 6 blisters and one box containing 1 blister (spare blister) will be packed per medication number in one outer box. At Visit 2 (Baseline) and Visit 6 (Day 57), the patient will be supplied with one box containing 6 blisters. The spare blister remains at the site and will only be dispensed if needed. Subject numbers should be written on the study medication labels.

Four different labels will be used for the study medication:

a) Label on the blister;
b) Label on the box containing 6 blisters;
c) Label on the box containing 1 blister (spare blister);
d) Label on outer box.

The labelling of the study medication is done according to GMP. Packaging and labeling of investigational product or placebo treatment will be indistinghuishable with respect to appearance and shape.

At each visit, patients will be instructed to bring the study medication box, containing all blisters (used and unused), to the clinic. At the clinic, a check will be performed on the returned study medication to assess treatment compliance.

4.3.4 Storage and handling

The study medication will be stored in a locked storage facility at ambient temperature: below 25°C but not frozen.

4.3.5 Drug aooountability

For both groups the study medication will be dispensed at Visit 2 (Baseline) and Visit 6 (Day 57). The patients should be instructed to bring the used and unused study medication back at each study visit during the treatment period. The investigators or their staff should check at these study visits if the patient's study medication use has been compliant.

The study monitor will check the used and unused study medication that has been collected by the site during the course of the study. The study medication should not be used for any other purpose than those described in this protocol and should not be given to any other person except those randomised in the study.

At first dispensing of study medication each patient will receive a Subject Emergency Card, containing the subject number, as well as emergency contact information numbers. Dispensing of the Subject Emergency Card must be documented appropriately.

The study medication will be stored at the site in a locked storage facility under the required conditions. The investigator or other appropriate individual who is designated by the investigator must maintain complete and accurate records, showing the receipt and disposition of all supplies of study medication delivered to the site. Any discrepancy must be explained in writing. Remaining study medication will be collected by the Sponsor for destruction or will be destroyed by the Pharmacy at the site after completion of the study.

4.3.6 Prior and concomitant therapy

The medications as mentioned in the exclusion criteria are *not* allowed during the study (see Section 4.4.3). This includes medication which are a part of:

- Hormone therapy – within 12 months prior to start treatment;
- Immunotherapy – within 12 months prior to start treatment;
- Chemotherapy – within 12 months prior to start treatment;
- Other investigational agents – within 4 weeks prior to start treatment.

Allowed during the study are 14 days concomitant treatment with an anti-androgen to prevent the flare-up, radiotherapy and low dose radiation to help in the prevention of gynecomastia.

The use of any concomitant medication should be recorded in source documents and on the Medication Case Report Form (CRF) pages. The use of pre-treatment and concomitant medication should be recorded from 60 days before screening up to and including the last trial day. Pre-treatment medication is medication used until the intake of the first study medication tablet.

4.3.7 Treatment beyond the clinical trial

No treatment with the study medication after the last tablet intake will be allowed under the current protocol.

4.4 Patient selection

4.4.1 Planned number and source

The patients for this study will be recruited from the patient database of the investigator under the responsibility of the investigator. Eligible patients will be men with prostate cancer, qualifying for treatment with an LHRH agonist. In total 60 patients are expected to be enrolled in the study.

The selection of patients will be made according to the in- and exclusion criteria listed in Section 4.4. Patients can only be included in the trial if all inclusion criteria are fulfilled, while none of the exclusion criteria are present. Recruitment will be stopped when 60 patients have been randomised.

4.4.2 Inclusion criteria

In order to be eligible to participate in this study, a patient must meet all of the following criteria:

1. Male patients with prostate cancer, qualifying for treatment with a LHRH agonist;
2. Age ≥ 18 years;
3. Body mass index (BMI) between ≥ 18.0 and ≤ 35.0 kg/m^2 (inclusive);
4. Reasonable physical and mental health as judged by the Investigator determined by physical examination, clinical laboratory assessments and vital signs;
5. Eastern Cooperative Oncology Group (ECOG) Performance Status 0-1;
6. Life expectancy of at least 2 years;
7. Willing to give informed consent in writing.

4.4.3 Exclusion criteria

A potential patient who meets any of the following criteria will be excluded from participation in this study:

1. Current or prior (during the last 12 months) hormonal therapy, immunotherapy or chemotherapy for prostate cancer. Allowed are 14 days concomitant treatment with an anti-androgen to prevent the flare-up, radiotherapy and low dose radiation to prevent gynecomastia;
2. History of deep vein thrombosis, pulmonary embolism, or cerebrovascular accident. However, patients with such history using anticoagulants for ≥ 6 months are eligible for the study provided anticoagulant treatment is continued throughout the whole study;
3. History of myocardial infarction or a coronary vascular procedure (e.g. percutaneous coronary intervention, coronary artery bypass graft). However, patients with such history using anticoagulants for ≥ 6 months are eligible for the study provided anticoagulant treatment is continued throughout the whole study;
4. Patients who have unstable angina or clinical congestive heart failure;
5. A defect in the blood coagulation system, assessed at screening: deficiencies in AT-III, protein C and protein S and elevated factor VIII;
6. Mutation in coagulation factor II and/or positive for factor V Leiden, assessed at screening;
7. Diabetes mellitus with poor glycaemic control in the past 6 months (haemoglobin A1c (HbA1c) above 7.5%);
8. Known primary hyperlipidaemias (Fredrickson);
9. Disturbance of liver function: cholestatic jaundice, a history of jaundice due to previous estrogen use, Rotor syndrome and Dubin-Johnson syndrome;
10. Known porphyria;
11. Uncontrolled hypertension, i.e. systolic blood pressure >160 mmHg and/or diastolic blood pressure >100 mmHg in the last 6 months with or without medication;
12. Subjects with a history of (within 12 months) alcohol or drug abuse;
13. Administration of any other investigational drug within 4 weeks prior to start treatment;
14. Any other condition (e.g. presence of any other malignancy), which in the Investigator's opinion, would not make the patient a good candidate for the trial.

4.4.4 Informed consent

The principles of informed consent will be implemented in accordance with the ethical principles that have their origin in the Declaration of Helsinki, the International Conference on Harmonization (ICH) Guideline for Good Clinical Practice (GCP), and applicable regulatory requirements.

Q-Man Questionnaire

Questions for patients participating in the PCombi study

A) Before start of the treatment

Instruction: Would you indicate whether you are experiencing these symptoms by ticking the yes or no box?

1. I have hot flushes
   ☐ Yes ☐ No
2. I suffer from night sweats
   ☐ Yes ☐ No
3. I have sleeping problems
   ☐ Yes ☐ No
4. I have pain in my joints (for example in knee, wrist or ankle)
   ☐ Yes ☐ No
5. I am tired easily
   ☐ Yes ☐ No
6. I have sensitive of painful nipples
   ☐ Yes ☐ No
7. My breasts are swollen
   ☐ Yes ☐ No
8. I am forgetful
   ☐ Yes ☐ No
9. I am easily agitated or angry
   ☐ Yes ☐ No
10. I cry quickly
    ☐ Yes ☐ No
11. I feel down
    ☐ Yes ☐ No
12. I need intimacy with my partner
    ☐ Yes ☐ No ☐ I have no partner
13. I am sexually active
    ☐ Yes ☐ No → You may skip question 14, 15 and 16
14. I am able to have an erection
    ☐ Yes ☐ No
15. I am able to reach an orgasm
    ☐ Yes ☐ No
16. I am satisfied with my sex life
    ☐ Yes ☐ No

Subject no.: ☐☐☐☐☐☐

Visit no.: ☐

Q-Man Questionnaire

Questionnaire for patients participating in the PCombi study

B) During of the treatment

Instruction: Would you indicate whether you are experiencing these symptoms by ticking the yes or no box?

1. I have hot flushes
   ☐ Yes ☐ No
2. I suffer from night sweats
   ☐ Yes ☐ No
3. I have sleeping problems
   ☐ Yes ☐ No
4. I have pain in my joints (for example in knee, wrist or ankle)
   ☐ Yes ☐ No
5. I have less strength in arms and legs
   ☐ Yes ☐ No
6. I am tired easily
   ☐ Yes ☐ No
7. I have sensitive of painful nipples
   ☐ Yes ☐ No
8. My breasts are swollen
   ☐ Yes ☐ No
9. I am forgetful
   ☐ Yes ☐ No
10. I am easily agitated or angry
    ☐ Yes ☐ No
11. I cry quickly
    ☐ Yes ☐ No
12. I feel down
    ☐ Yes ☐ No
13. I need intimacy with my partner
    ☐ Yes ☐ No ☐ I have no partner
14. I am sexually active
    ☐ Yes ☐ No → You may skip question 15, 16, 17, 18 and 19
15. I have less desire for sex
    ☐ Yes ☐ No
16. I am able to have an erection
    ☐ Yes ☐ No
17. I am able to reach an orgasm
    ☐ Yes ☐ No
18. I produce less ejaculate
    ☐ Yes ☐ No

19. I am satisfied with my sex life
☐ Yes ☐ No

20. My relationship is suffering from the treatment
☐ Yes ☐ No ☐ I have no partner → You may skip question 21

21. My partner thinks our relation is suffering from the treatment
☐ Yes ☐ No

1 General

This Statistical Analysis Plan (SAP) describes in detail methods and presentation of the data analyses which will be conducted by AUTHOR! et al B.V. for study PR3109. This plan is written in agreement with protocol version 2.0, dated 8 November 2018, and with the Case Report Form (CRF) version 5.2.

The protocol and the (annotated) CRF are the primary source for this document, together with the relevant International Conference on Harmonization and Good Clinical Practice (ICH-GCP) guidelines. Furthermore, sponsor requirements for reporting are taken into account. Additional changes or updates of those documents or requirements will result in a new version of the reporting/statistical analysis plan. For general study information (objectives, study design, study flow chart and study endpoints), is referred to the study protocol.

2 Subjects for Analysis

2.1 Analysis Populations

- **All-Subjects-Randomised (ASR)** population will consist of all randomised patients.
- **All-Subjects-Treated (AST)** population will consist of all randomised patients who received at least one dose of study medication. Only if all dispensed study medication is returned, the patient will be considered non-treated and will not be included in the AST group. Analyses on the AST population will be done according to the 'as treated' principle.
- **Intent-to-Treat (ITT)** population will consist of all randomised patients who received at least one dose of study medication. Analyses on the ITT population will be done according to the 'as randomised' principle.
- **Per-Protocol (PP)** population will consist of all randomised patients who completed the study without major protocol deviations.

The number/percentage of patients included in each of the analysis populations will be summarized per treatment group and overall, for the ASR population. This information will also be listed including reasons for exclusion from each dataset on a per subject level. Screen failures will be listed.

2.2 Protocol Deviations:

A protocol deviation is defined as a lack of compliance with the protocol which may interfere with the main efficacy analysis. Protocol deviations will be defined at the time of the evaluability assessment, the time between the database soft lock and hard lock before unblinding, by clinical and statistics personnel during a Blinded Data Review Meeting (BDRM).

Possible protocol deviations that lead to exclusion from the PP include, but are not limited to:

1. Informed consent not obtained/signed

2. Violation of in/exclusion criteria
3. Not receiving treatment to which the patient was randomised
4. Less than 80% compliance to treatment
5. Last efficacy assessment more than 7 days after last medication intake
6. Taken forbidden concomitant medication/therapy
7. Timing of Luteinising Hormone Releasing Hormone (LHRH) treatment interfering with efficacy assessment

2.3 Blinded Data Review Meeting

The patients will be classified to the various analyses populations during the BDRM. Invitees to this meeting are the data manager, the clinical project manager, the clinical research associate, sponsor representative and the study statistician, but more roles can be invited if considered necessary. Input to this meeting will be supplied by the involved data management provider and statistician/statistical programmer at least one week in advance of the meeting: a blinded list of all protocol deviations (including missing and outlier data), with specific detailing and description regarding the deviation.

The goal of this meeting is to reach consensus on screened, randomised and treated populations and on minor and major protocol deviations. In case of a deviation impacting the primary or secondary endpoints, the specific patient will be excluded from the PP population, either completely or from a specific time point onwards.

The meeting will be held prior to unblinding of the CRF data and blinded lab data. The unblinded lab data (total Testosterone (T), free T, Dihydrotestosterone (DHT), endocrine parameters, Dehydroepiandrosterone Sulfate (adrenal androgen (DHEAS)), bone turnover and haemostasis) will not be considered when deciding on the analysis populations and protocol deviations.
The decisions taken during the meeting will be documented by the project manager (or a delegate as agreed) and sent for review to all parties involved as soon as possible after the meeting, but before database lock. If all parties involved agree, then the document is finalised, signed by all attendees, and stored before database lock by the project manager. After the document is signed and the CRF data and blinded lab data are locked, the SAP is finalised and subsequently the study is unblinded by the responsible statistician. Then, the unblinded lab data are added to the lab database and the complete database will be locked. Changes to the SAP after unblinding are described in section 6.1 of this SAP.

2.4 Study Day

Study day is defined as the number of days since first study medication intake.

3 Statistical Analysis

3.1 General Considerations

All raw data will be listed. In line with the protocol analyses will be descriptive and explorative in nature. Statistical testing will be limited to the comparison between the active treatment group (estetrol (E4)) and placebo as described in section 5.3. Further comparisons between the two treatment groups are performed using descriptive statistics. E4 and Placebo

will be used as treatment group labels in tables and listings. For all efficacy variables, the 6 months visit (Visit 9) is the primary endpoint.

Descriptive statistics presented in summary tables will be the number of non-missing observations (n), arithmetic mean, Standard Deviation (SD), 95% confidence interval limits, median, minimum, maximum and interquartile range for continuous data. For categorical data, number of non-missing observations (n) and percentages in each category will be presented. Appropriate rounding will be performed for all statistics.

Baseline is defined as the last value measured prior to the first study medication intake, including unscheduled visits.

3.2 Application of Analysis Populations:

The primary population for efficacy analysis is the PP population.

Subject enrolment, disposition, drug exposure, demographics and baseline disease characteristics will be shown for all populations. Efficacy analyses of primary and secondary endpoints will be performed on both the PP and ITT populations. All safety evaluations, medical history and medication use will be based on the AST.

When analysing the PP or ITT group, the analysis will be according to the "as randomised" principle. For the AST group subjects will be analysed "as treated".

3.3 Missing or excluded data

As all analyses applied to the primary and secondary endpoints will be explorative and descriptive in nature, no imputation of missing data will be performed. For handling any missing data of statistical analyses applied to the primary and secondary endpoints, refer to the respective analysis section.

Missing/incomplete information related to adverse events (AEs) will be handled as listed below, if applicable:

- In case of (partially) missing start/onset dates, the following rules will be followed:
  - In case only the start day is missing:
    - In case month and year of start date are equal to or after the month and year of first study medication intake, but before or equal to the month and year of last study medication intake, the AE is considered a Treatment-emergent AE (TEAE), unless it is clear from the (partial) stop date that the AE stopped before first study medication intake.
    - In case month and year of start date are before month and year of first study medication intake, the AE is considered pre-treatment.
    - In case month and year of start date are after month and year of last study medication intake, the AE is considered post-treatment.
  - In case the start day and month are missing:
    - In case the year is equal to or after the year of first study medication intake and equal to the year of last study medication intake, the AE is considered a TEAE, unless it is clear from the (partial) stop date that the AE stopped before first study medication intake

- In case the year is before the year of first study medication intake, the AE is considered pre-treatment.
- In case the year is after the year of the last study medication intake, the AE is considered post-treatment.

- In case the start date is completely missing, the AE is considered a TEAE, unless it is clear from the (partial) stop date that the AE stopped before first study medication intake.

Following these steps using programming will ensure that an AE is considered treatment emergent for partially missing start dates, using the most conservative approach.

- In case intensity is missing for a certain TEAE, this will be regarded as severe.
- In case causality is missing for a certain TEAE, this will be regarded as related.
- In case seriousness is missing for a certain TEAE, this is discussed and addressed prior to database lock and unblinding.

Regarding prior/concomitant medication, a similar approach will be followed for partially missing dates, as will be done for AEs as described above.

Based on the BDRM, certain values of efficacy endpoints can be decided to be excluded from analyses. These values will be listed, but not included in descriptive statistics, plots or statistical analyses. This may be outlier data that are a result of unambiguous measurement errors. Non-excluded outlier data will be included in all analyses.

3.4 Interim Analysis

No interim analysis is planned.

4 Baseline Characteristics

4.1 Demographics and other Baseline/Screening Data

Subjects will be described using demographic information and baseline characteristics recorded during the screening phase.

The following demographic data will be summarized per treatment group and overall: age, race, ethnicity, body weight measured at baseline (Visit 2), height measured at screening, Body Mass Index (BMI) and smoking habits. In addition, number of patients treated in each study site will be presented. BMI will be derived using body weight measured at baseline and height measured at screening (see programming conventions). Descriptive statistics for age, weight, height, BMI, race, ethnicity, and smoking habits will be given. Demographic data will be listed for the ASR population on a per subject level.

Inclusion/exclusion criteria will be listed on a per subject basis for the ASR population, if applicable.

4.2 Baseline Disease Characteristics

Other baseline/screening data include malignancy and cancer therapy history, Eastern Cooperative Oncology Group (ECOG) Performance Status, PSA, Total T and free T levels and Quality of Life by the QMAN-A questionnaire.

Malignancy and cancer therapy history including ECOG Performance Status will be summarized in a table per treatment group and overall using descriptive statistics. This table will include Tumor Nodes Metastasis (TNM) classifications and stage of disease at first cancer diagnosis and screening, Gleason score at screening in categories ($\leq$6, 7, $\geq$8), previous prostate cancer treatments and ECOG grade. In addition, duration of disease measured in days will be presented in this table derived by subtracting diagnosis date from screening date, plus one day. PSA, Total T and free T and QMAN-A will be summarized per treatment group and overall using descriptive statistics.

4.3 Medical History

Medical history will be listed on a per subject basis and summarised for the AST population including diagnosis (reported term, Preferred Term (PT) and System Organ Class (SOC)), start-, and end date (if applicable), and start-, and end day. The PT and SOC coding of the medical history diagnosis will be determined by the data management provider using the MedDRA coding system (MedDRA version will be determined at database lock).

4.4 Disposition

The disposition of patients will be summarised including the number/percentage of patients randomised, treated, and completed by treatment group and overall. Randomised will be determined/confirmed by non-missing date and time of randomization, treated will be determined/confirmed by non-missing date of first study medication intake (if date of first study medication intake is missing, it will be investigated if doses are missing during drug accountability evaluations and if so, the subject is considered treated) and completed will be determined by the question if study was completed. Disposition information will be listed including reason for discontinuation and study day of when the patients discontinued.

In addition, a table will be created summarizing number/percentage of discontinuations and reasons of discontinuation.

A complete schematic overview of the number of patients screened, randomised to both groups, treated, discontinued, and completed including reasons for exclusion will be visualized in a consort graph.

5 Statistical analysis of efficacy

5.1 Statistical Analysis Primary Endpoint: total T and free T

The first primary outcome for this study is to assess the additional suppressive effects of E4 on total T and free T (calculated and measured). Total T and free T will be assessed before dosing at baseline and at regular intervals during the study up to and including Visit 9/end of treatment.

5.1.1 Actual Value of total T, free T (calculated) and free T (measured)

The actual values of total T and free T calculated and measured at scheduled visits will be summarized in a table for each visit per treatment group using descriptive statistics. The statistics measured at Visit 9 will be the primary time-point.

The median total T and median free T values per visit will be displayed in graphs with visit number on the x-axis and median total T or median free T on the y-axis. Only scheduled visits will be included. Separate graphs will be created for median total T and median free T. In each graph, two curves will be plotted: one curve per treatment group. The two curves in each graph will be distinguished using two different line types.

5.1.2 Nadir of total T and free T

Nadir and time to nadir of total T and free T will also be assessed. Nadir is defined as the lowest point in total T and free T and will be derived by selecting the minimum value of total T and free T measurements over all visits (including unscheduled visits) for each patient. Time to nadir will be measured in study days using the date of the lowest observation. Nadir and time to nadir of total T and free T will be summarized in a table per treatment group using descriptive statistics. Furthermore, nadir and time to nadir will be listed per subject.

5.1.3 Breakthrough Response total T

Breakthrough responses will be identified for total T over the course of the treatment period (baseline excluded). A breakthrough response is defined as an absolute total T value of 20 ng/mL or higher. Patients with at least one breakthrough response in absolute total T will be counted and the number/percentage will be presented in a table per treatment group. In addition, number of study days to the first breakthrough response will be calculated and presented as well per treatment group using descriptive statistics. Furthermore, the value of the breakthrough response and the number of days till first breakthrough response will be listed per subject.

5.2 Statistical Analysis Primary Endpoint: Hot Flushes

The second primary outcome for this study is to assess the effects of E4 on hot flushes. Patients will be asked three times during the treatment period to fill out a daily diary about the number and severity (mild=1; moderate=2; severe=3; and very severe=4) of hot flushes over a one-week period (7 days), following the week after Visit 4 (Week 5), the week after Visit 7 (Week 13) and the week before Visit 9 (Week 23). The following time frames will be considered for determining which analysis week the daily diary[1] belongs to: diaries completed between day 1 (Week 1) and day 63 (Week 8) will be attributed to week 5, diaries completed between day 64 (Week 9) and day 127 (Week 18) will be attributed to week 13 and diaries completed after day 127 (week 19 onwards) will be attributed to week 23.

A mean daily hot flush score, defined as the number of hot flushes per day multiplied by their mean severity per day, averaged over the one-week period, will be calculated per subject per analysis week. The mean severity per day is derived as the total hot flush score divided by the number of hot flushes, where the total hot flush score will be derived as: *number of mild hot*

[1] First day of diary completion determines to which week the diary is attributed

flushes multiplied by 1 plus number of moderate hot flushes multiplied by 2 plus number of severe hot flushes multiplied by 3 plus number of very severe hot flushes multiplied by 4. The mean daily hot flush score will be derived using the actual number of days a diary is completed which means that in case of a missing day, the missing day will not be included when deriving the mean daily hot flush score. The mean daily hot flush score will be summarized per treatment group per week using descriptive statistics.

A graph will be created displaying the mean daily hot flush scores per week with week number on the x-axis and mean daily hot flush score +/- 95% confidence interval on the y-axis. In this graph, two curves will be plotted: one curve per treatment group. These curves will be distinguished using two different line types.

For every week the hot flush diary is filled in, the number/percentage of patients experiencing at least one hot flush will be summarized per treatment group. In addition a table with the maximum reported severity by week will be created.

The number of mild, moderate, severe and very severe hot flushes and the mean daily hot flush score will be listed per week on a per subject level.

5.3 Exploration of efficacy on primary endpoints using statistical testing

Difference in the measured free T in week before visit 9 between the 2 treatment groups will be explored for the PP population using a t-test with the following hypothesis:

μ_1 = measured free T at visit 9 in active treatment group
μ_2 = measured free T at visit 9 in placebo group

The following hypothesis is set for $\mu 1$ and $\mu 2$:

H_0 : $\mu_1=\mu_2$
H_A : $\mu_1\neq\mu_2$

Free T data will only become available after unblinding. As a first step, the normality assumption of measured free T will be investigated on the blinded visit 9 data set. If there is excessive skewness a non-parametric alternative to the t-test (Wilcoxon rank sum test) will be considered for the visit 9 treatment comparison.

Difference in the occurrence of hot flushes (patient experiencing at least one hot flush) in the week before visit 9 between the 2 treatment groups will be explored for the PP population using Fisher's exact test with the following hypothesis:

$\mu 1$ = proportion of patients experiencing hot flushes in the week prior to visit 9 in active treatment group

$\mu 2$ = proportion of patients experiencing hot flushes in the week prior to visit 9 in placebo group

The following hypothesis is set for $\mu 1$ and $\mu 2$:

H_0 : $\mu_1=\mu_2$

$$H_A: \quad \mu_1 \neq \mu_2$$

Statistical testing will be conducted two-sided with a significance level of 5% for both endpoints. All confidence intervals will be presented two-sided with a confidence level of 95%. A resultant probability value of $p<0.05$ will be judged as being of statistical significance. There will be no efforts to control for multiplicity.

5.4 Statistical Analysis Secondary Endpoints

5.4.1 Endocrine Parameters, Adrenal Androgens, DHT, SHBG, IGF-1

The first series of secondary outcomes for this study is to assess the effects of E4 on endocrine parameters Luteinising Hormone (LH), Follicle Stimulating Hormone (FSH) and Estradiol (E2); adrenal androgen parameter DHEAS; DHT; Sex Hormone Binding Globulin (SHBG) and IGF-1 (Insulin-like Growth Factor-1). The actual values measured at baseline and at scheduled visits, as well as the change from baseline at post-baseline visits will be summarized per treatment group per visit using descriptive statistics.

For each parameter separately, the actual values (including baseline) at scheduled visits will be displayed in a graph with visit number on the x-axis and the actual values on the y-axis. In each graph, two curves will be plotted: one curve per treatment group. The two curves in each graph will be distinguished using two different line types.

5.4.2 PSA Response

The second secondary outcome for this study is to assess the effects of E4 on the Prostate-Specific Antigen (PSA) response. The actual values measured at baseline and at scheduled visits, as well as the change from baseline and the percentage change from baseline at post-baseline visits will be summarized per treatment group using descriptive statistics.

Nadir and time to nadir will be assessed and presented for PSA response in the same way as for the primary outcomes total T and free T.

The median of the percentage change from baseline in PSA over time will be displayed in a graph with visit number on the x-axis and percentage change from baseline on the y-axis. Only scheduled visits will be included. In this graph, two curves will be plotted: one curve per treatment group. The two curves will be distinguished using two different line types.

5.4.3 Health Related Quality of Life

The third secondary outcome for this study is to assess the effects of E4 on quality of life scales. Quality of life was assessed using two questionnaires: the Functional Assessment of Cancer Therapy-Prostate (FACT-P) and the Q-sum questionnaire.

Facr-P

The FACT-P questionnaire was assessed at Visit 6 and Visit 9. The FACT-P questionnaire will be summarized using the following (sub)scales consisting of the sum of a subset of items:

(Sub)scale	Items
FACT-P Total	All items
FACT-G** Physical well-being	GP1*, GP2*, GP3*, GP4*, GP5*, GP6*, GP7*
FACT-G Social/Family well-being	GS1, GS2, GS3, GS4, GS5, GS6, GS7
FACT-G Emotional well-being	GE1*, GE2, GE3*, GE4*, GE5*, GE6*
FACT-G Function well-being	GF1, GF2, GF3, GF4, GF5, GF6, GF7
FACT-G Total	All FACT-G subscales
FACT-P Prostate Cancer Subscales (PCS)	C2*, C6, P1*, P2*, P3*, P4, P5, P6*, P7*, BL2*, P8*, BL5

*Items followed by an asterisk need to be reversed by subtracting the score from 4 to ensure large scores mean better quality of life.
**G = General.

In case of missing items, subscales can be prorated by multiplying the sum of the subscale by the number of items in the subscale divided by the number of items answered. This prorating is only acceptable as long as more than 50% (e.g. 4 out of 7 items, 4 out of 6 items) of the items of a subscale are answered. When less than 50% of the items of a subscale is completed, the subscale is considered missing. For the FACT-P total score and the FACT-G total score, at least 80% of the items need to be answered (e.g. 22 out of 27 FACT-G items completed). In case less than 80% is completed, the total score is considered missing.

The FACT-P (*sub*)scales scores will be summarized in separate tables for each visit per treatment group using descriptive statistics. In addition, FACT-P subscales will be listed on a per subject level.

Q-MAN

The Q-man questionnaire was assessed at baseline, Visit 4, Visit 7, Visit 9 and Visit 10. The Q-man questionnaire measured at baseline consists of 16 symptoms (version A) and the Q-man questionnaire measured at post-baseline visits consists of 21 symptoms (version B).

The number/percentage of patients reporting Yes/No for the various symptoms assessed by the Q-man questionnaire at baseline and post-baseline visits will be summarized using frequencies. In addition, per Q-man item, the most positive value across all post-baseline values will be selected and differences and similarities between the treatment groups on these most positive values will be explored.

A total Q-Man B post-baseline score will be determined by visit by adding all scores of a patient at that visit, with the answers to questions 13, 14, 16, 17, 19 and 20 reversed, resulting in a total Q-Man B score measuring burden. In case of missing items, the total score will be prorated by multiplying the sum of the score by the number of symptoms in the score (i.e. 16 or 21) divided by the number of symptoms scores answered for the score. This prorating will be applied as long as more than 50% (e.g. 8 out of 16 Q-Man A items, 11 out of 21 Q-Man B items) of the symptoms questions of the questionnaire are answered. If this 50% is not reached for a patient, the Q-Man score will be missing for that visit. Withdrawal effects of

active treatment group will be studied by calculating the difference between the Visit 9 and Visit 10 total Q-Man B score and exploring this difference for the two treatment groups.

5.4.4 Lipids

The fourth secondary outcome for this study is to assess the effects of E4 on lipids parameters total cholesterol, triglycerides, High Density Lipoprotein (HDL)-cholesterol and Low Density Lipoprotein (LDL)-cholesterol. The actual values measured at baseline and at scheduled visits, as well as the change from baseline at post-baseline visits will be summarized by treatment group per visit using descriptive statistics. Differences and similarities between the two treatment groups on the change from baseline will be explored.

For each parameter separately, the actual values (including baseline) at scheduled visits will be displayed in a graph with visit number on the x-axis and the actual values on the y-axis. In each graph, two curves will be plotted: one curve per treatment group. The two curves in each graph will be distinguished using two different line types.

5.4.5 Bone Turnover

The fifth secondary outcome for this study is to assess the effects of E4 on bone turnover parameters osteocalcin and Type I Collagen Telopeptide (CTX-1). The actual values measured at baseline and at scheduled visits, as well as the change from baseline and the percentage change from baseline at post-baseline visits using appropriate descriptive statistics. Differences between the two treatment groups on the percentage change from baseline will be explored.

For each parameter separately, the actual values (including baseline) at scheduled visits will be displayed in a graph with visit number on the x-axis and the actual values on the y-axis. In each graph, two curves will be plotted: one curve per treatment group. The two curves in each graph will be distinguished using two different line types.

5.5 Safety Evaluation

5.5.1 Adverse Events

An Adverse Event (AE) is defined as any untoward medical occurrence in a subject during the study, whether or not considered to have a causal relationship with the investigational product.

Treatment-emergent AEs (TEAEs) are those AEs occurring from time point of first study medication intake until last visit.

Adverse Events (AEs) will be coded by SOC and PT using MedDRA (MedDRA version will be determined at database lock) by the Data Management provider. In tables, SOC and PT will be presented in descending order of frequency. If several SOCs/PTs have the same number of frequencies, the SOCs/PTs will be presented in alphabetical order.

An AE overview table will be created displaying the number/percentage of patients experiencing any TEAE, deaths, any SAE, any TEAEs causing discontinuation, any severe TEAE and any drug-related TEAE per treatment group.

Incidence tables will be created displaying incidence of patients with at least one TEAEs by SOC and PT per treatment group. Separate incidence tables will be created for related TEAEs and TEAEs by severity.

An AE is considered related if the relationship to study drug is classified as either 'Possible', 'Probable' or 'Definite'.

The summary tables will be accompanied by individual subject listings of all AEs including information on AE number, actual AE description, date of start and end of AE (or ongoing), start-, and end day (or ongoing), PT, SOC, severity, relationship, seriousness, action taken and outcome. AEs appearing during pre-treatment phase (i.e. occurring after signing of Informed Consent but before first intake of study medication) and follow-up phase will only be included in the AE listing. In this listing, a clear distinction will be made between pre-treatment and treatment..

Separate listings will be created for SAEs, deaths and AEs leading to discontinuation, if applicable.

5.5.2 Laboratory Safety Measurements

The following routine laboratory safety data are collected for this study:

Haematology	Biochemistry	Haemostasis
Leucocytes	Blood Urea Nitrogen	D-dimer
Lymphocytes	Glucose (blood)	Fibrinogen
Monocytes	Albumin	Anti-thrombin III
Neutrophils (total)	Alkaline phosphatase	Protein S
Basophils	ALT	Factor VIII
Eosinophils	AST	
Platelets	Total bilirubin	
Haemoglobin	Total protein	
Haematocrit	Creatinine	
MCV	Sodium	
MCH	Potassium	
MCHC	Chloride	
Red Blood Cell count	Calcium	
	Gamma GT	
	Phosphate	
	LDH	
	HbA1c	

Laboratory safety data for haematology, biochemistry and haemostasis will be summarized using descriptive statistics per visit and per treatment group. Change from baseline and percentage change from baseline will be calculated and presented for continuous data. In addition, abnormal values outside reference ranges will be summarized in a listing per visit and per subject. An out-of-range listing will also be created for the efficacy lab parameters total T, free T, endocrine parameters, adrenal androgen parameter DHEAS, DHT, SHBG,

PSA response, lipids parameters and bone turnover parameters (of which the descriptive statistics are part of the efficacy evaluation for the PP/ITT population). The out-of-range summary listing for these efficacy lab parameters will be presented for the AST population.

If applicable, laboratory safety data collected as additional/unscheduled assessments (i.e. apart from those per protocol) will only be listed and will not be used in summary statistics.

Safety laboratory parameters will be presented in the tables and listings in the same units as supplied, which will be standardized or SI units.

5.5.3 Vital Signs

Vital sign data consist of measurements of body weight, heart rate, systolic blood pressure and diastolic blood pressure. Height and BMI will not be included, as those are only baseline measurements and already included in the Demographics table. Vital signs will be summarized per visit and per treatment group using appropriate statistics. Change from baseline will be calculated and presented as well. Incidence of vital signs data outside normal ranges will be summarized in table per parameter, per visit and per treatment group using number/percentage of patients. If applicable, vital sign measurements collected as additional/unscheduled assessments (i.e. apart from those per protocol) will only be listed and will not be used in summary statistics.

5.5.4 Concomitant Medication

The use of concomitant medication will be listed for all patients: included will be the reported medication generic name, dose, route of administration, start and stop date, start and stop day (relative to first study medication intake), frequency and reason for administration, as well as information if given for an AE. Differentiation will be made between prior and concomitant medication. Prior concomitant medication is defined as any medication taken before start date of first study medication intake. If a concomitant medication is taken before start date of first study medication intake and still ongoing during the in-treatment phase, the concomitant medication will also be considered prior medication.

5.5.5 Physical Examination

General physical examination data will be listed, and any clinically significant abnormal findings will be summarized per treatment group.

5.6 Scheduled visits, Randomization, Extent of Exposure and Treatment Compliance

5.6.1 Randomization

A listing of randomization information per subject will be created for the AST population.

5.6.2 Treatment Duration, Extent of Exposure and Treatment Compliance

Treatment duration, extent of exposure and treatment compliance will be presented per treatment group for the ASR, AST, ITT and PP population. Treatment duration will be derived as the difference in days between date of first study medication intake and date of last study medication intake. Extent of exposure will be derived as the number of doses (2 tablets

per day) taken which will equal the number of days exposed to study treatment. Treatment compliance will be defined as the ratio of the number of doses taken and number of planned doses to be taken during the study where number of planned doses to be taken is based on the treatment duration. In addition, number of missed doses will be presented as well.

Treatment duration, extent of exposure, treatment compliance and number of missed doses \will also be listed on a per subject level for the AST population. In addition, a drug accountability listing will be created per visit including the total number of doses taken up to that visit (cumulative number) for the AST population.

6 Changes from protocol

Explorative statistical testing of the primary efficacy parameters was added to the statistical analysis plan.

6.1 Changes to the SAP after breaking the blind

Bicalutamide use at baseline:

- For each patient it was determined if patient was using bicalutamide at the moment of their baseline efficacy laboratory assessment. An indicator of that use was added to the demographics table.

Efficacy Lab-parameters:

- LLOQ values were replaced by 0.5*LLOQ value in descriptive statistics, inference and graphs
- ULOQ values were replaced by 1.5*ULOQ value in descriptive statistics, inference and graphs
- Because there were problems with the DHT-assay, DHT results were excluded from the analysis
- To support publication of the study results various statistical analyses were added to the SAP:
    - Baseline testing of differences in demographics and baseline disease characteristics using a t-test for continuous variables and a Chi^2-test for categorical variables.
    - Differences between treatment groups over time for measured free T, total T and PSA, measured at baseline, Day 15, Day 29, Day 43, Day 57, Day 85, Day 128 and at Day 168/Week 24 were analysed using an unstructured repeated measures mixed model on log-transformed values investigating treatment, day and treatment*day interaction.
    - Secondary efficacy parameter (i.e LH, FSH, E2, DHEAS, SHBG, IGF-1, Osteocalcin and CTX1) differences between treatment groups at Day 168/Week 24 were analysed using a Kruskall-Wallis test.

Q-Man B total score:
The Q-Man B total score as described in section 5.4.3 of the blinded SAP was removed from the analysis.

Hard updates:

- Hot flush date of Subject 107-005 has been updated from '2019-08-21' to '2019-09-21', as the study nurse incorrectly entered the first date in the database, while the patient had used the last date in the (paper) diary.
- Drug accountability was corrected ('hard update') for Subject 107-001 on Visit 8 (140 used tablets instead of 84) and Visit 9 (222 used tablets instead of 166)
- For patient 107-002 an initial SAE form with narrative events: (1) Decompensation cordis and anemia with persistent rectal bleeding after radiotherapy for prostate cancer. (2) Kidney function loss was reported. The SAE form was updated by the investigator with narrative event: (1) Symptomatic anaemia with rectal bleeding after radiotherapy for prostate cancer. The investigator responded in a query: "Decompensation cordis is secundair due to anemia and event in now classified as AE". The SAE database was updated accordingly, but the AE CRF (Viedoc) was not corrected for seriousness. After database lock, the field SAE for the AE Decompensation Cordis of patient 107-002 was changed from Yes to No.

7 Technical details

7.1 Programming conventions

BMI was calculated as follows: weight measured at baseline/height measured at screening^2, with weight in kg and height in meter (unit kg/m^2).

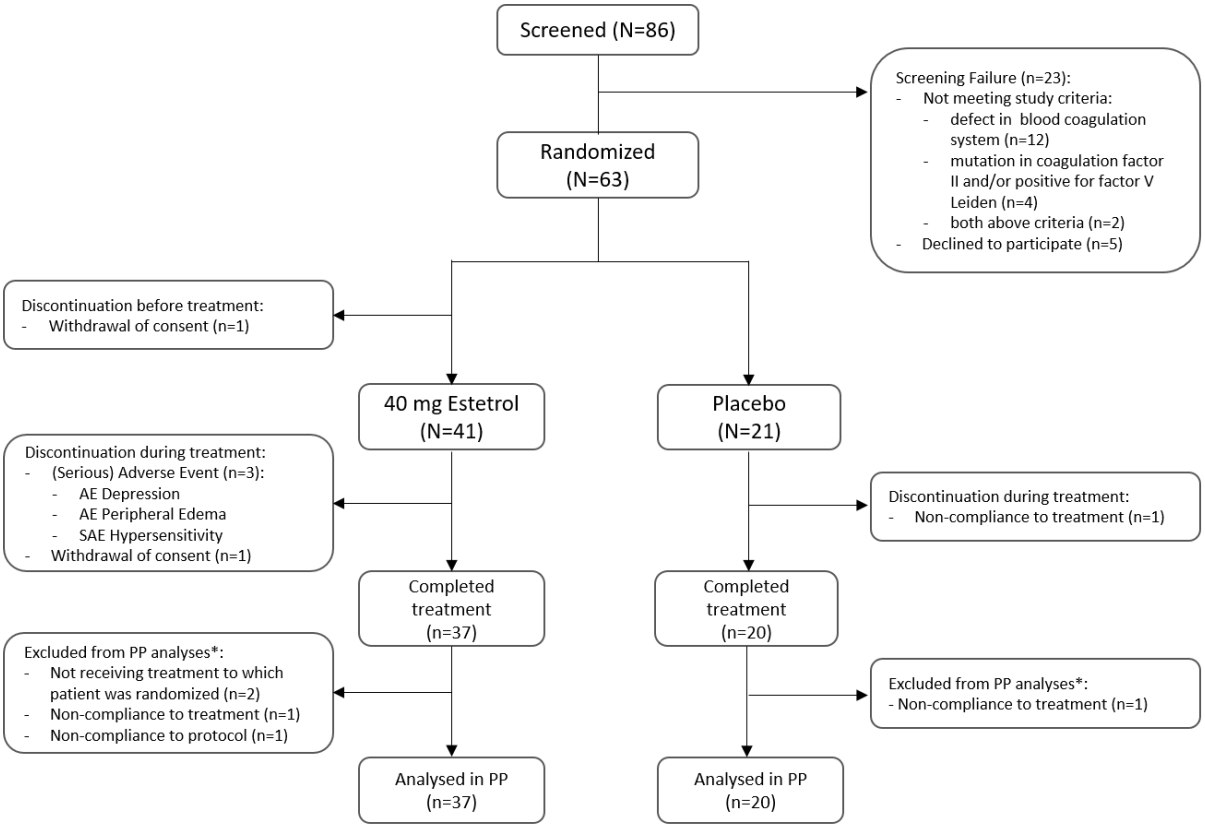
7.2 Analysis software

The statistical analysis and reporting will be done using SAS® for Windows™ version 9.4.

7.3 Presentation of tables, listings, graphs

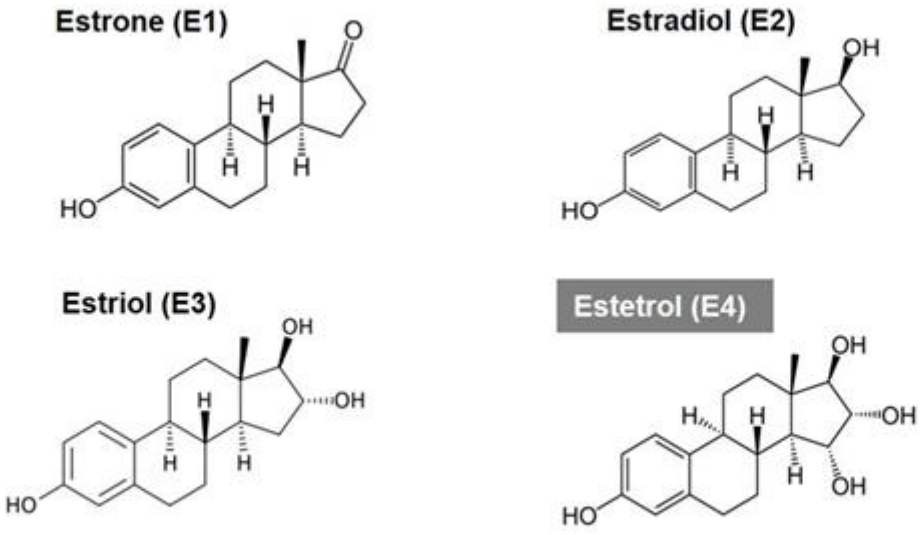
All output will be generated as tables, graphs and listings following ICH-E3 standards for Section 14 and Section 16 numbering.

Supplementary Figure 1



* Based on all patients who started treatment; patients not completing study treatment could still be included in PP analyses if all study assessments were performed (e.g. end of study visit). This was applicable for the patients of the estetrol group, but not for the patient of the placebo group. All patients who received at least one dose of study medication were included in the analyses of the safety parameters

Supplementary Figure 2 The four estrogens



Supplementary Table 1 Lipid and hemostasis parameters at baseline and after 24 weeks of treatment with 40 mg estetrol or placebo daily co-administration in patients with prostate cancer treated with an LHRH agonist (All-subjects-treated group)

Laboratory Parameter	40 mg Estetrol (N=41)			Placebo (N=21)		
	Baseline	End of Treatment (week 24)	Percentage change from baseline	Baseline	End of Treatment (week 24)	Percentage change from baseline
Total Cholesterol (mmol/L)	4.888 (0.922)	5.020 (0.894)	3.9 (15.7)	4.935 (1.095)	5.096 (1.185)	3.5 (12.7)
Triglycerides (mmol/L)	1.764 (0.916)	1.821 (0.715)	15.0 (45.5)	1.927 (0.749)	2.129 (1.057)	15.7 (48.7)
HDL Cholesterol (mmol/L)	1.283 (0.332)	1.550 (0.334)	22.6 (15.1)	1.206 (0.336)	1.278 (0.346)	6.5 (18.3)
LDL Cholesterol (mmol/L)	2.805 (0.906)	2.633 (0.802)	-2.9 (28.6)	2.849 (0.971)	2.883 (1.080)	3.8 (33.1)
Antithrombin (%)	101.4 (10.7)	88.6 (9.3)	-12.1 (6.7)	100.3 (11.5)	96.4 (12.7)	-3.4 (12.1)
Protein S activity (%)	109.9 (23.5)	80.1 (18.4)	-25.5 (14.9)	105.1 (24.1)	99.5 (29.4)	-7.5 (14.7)
Factor VIII (%)	144.8 (36.6)	153.2 (44.1)	9.2 (31.2)	150.0 (40.5)	162.3 (44.5)	10.4 (28.2)
Fibrinogen (g/L)	3.493 (0.551)	3.302 (0.562)	-3.8 (18.8)	3.593 (0.694)	3.708 (0.710)	4.1 (14.7)
D-dimer (mg/L FEU)	0.734 (0.577)	0.565 (0.337)	-4.4 (49.9)	1.291 (1.878)	1.261 (2.248)	-8.8 (39.0)

Results reported as mean (standard deviation)
HDL, high density lipoprotein; LDL, low density lipoprotein