# Statistical analysis plan

Identifiers: NCT02759380
Brief Title: Can Dietary Phytoestrogens Slow Down Prostate Tumor Proliferation? (PRODICA)
Official Title: Can Dietary Phytoestrogens Slow Down Prostate Tumor Proliferation? A Randomized Study (PRODICA)
Principal Investigator: Maria Hedelin
Date: 2022-05-31

**AIM**

The overall aim of the PRODICA-study is to increase the knowledge, for the impact of diet on prostate tumor proliferation.

HYPOTHESIS

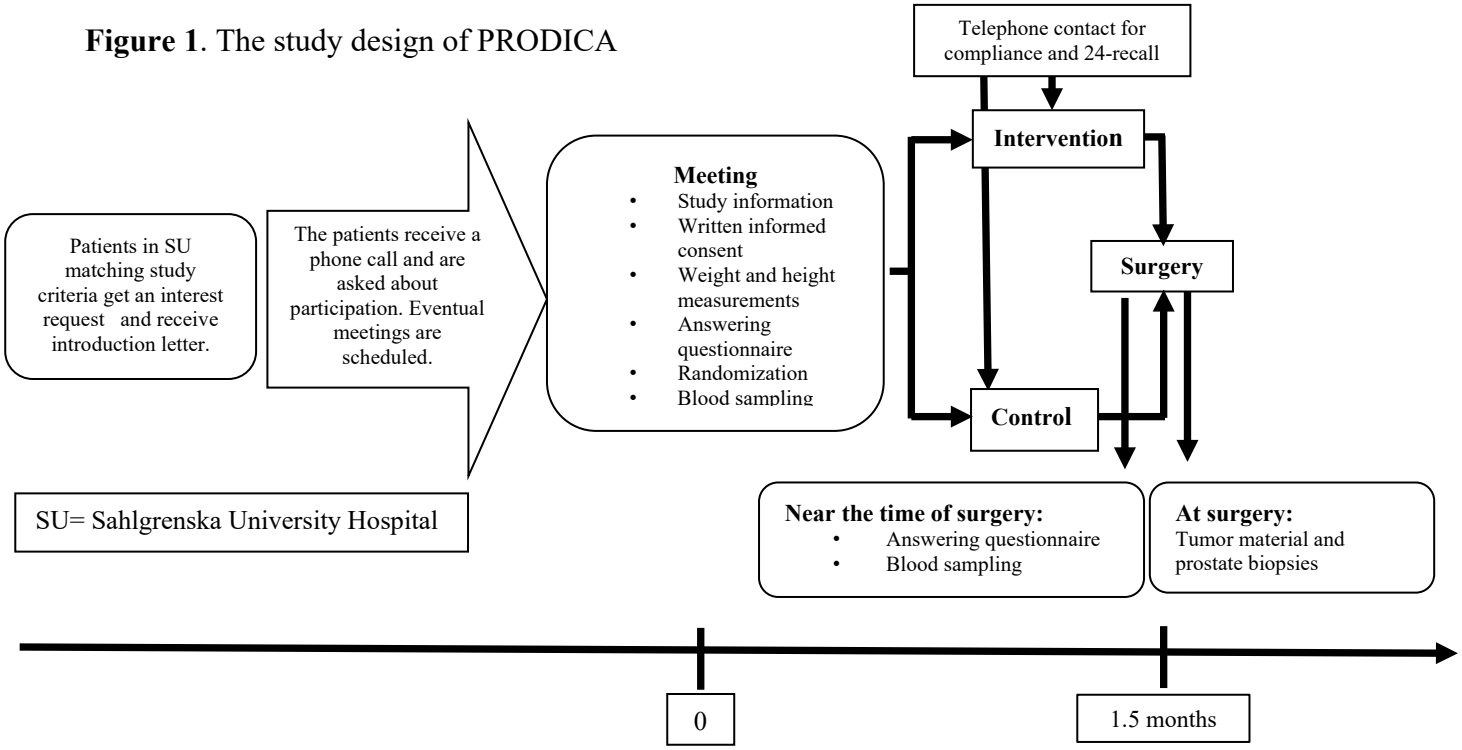
1) In men diagnosed with intermediate risk-level prostate cancer, the addition of phytoestrogen-rich foods containing 200 mg phytoestrogens per day in six weeks, gives a reduced prostate tumor proliferation compared to an absence of the addition of phytoestrogen-rich foods during the same period.
2) If the effect of phytoestrogens on prostate tumor proliferation exists, it is modified by a SNP in the promoter region of the estrogen-beta-gene (rs 2987983-13950).

STUDY DESIGN

The study design is shown in Figure 1. Men diagnosed with intermediate-risk prostate cancer T1–T2 (Gleason score <8, PSA <20), scheduled for radical prostatectomy, are identified at the Sahlgrenska University Hospital. Interested men are contacted by phone and a first meeting is scheduled. At the meeting the participants receive verbal and written information and sign an informed consent. A food frequency questionnaire including dietary and living habits and supplements are filled in and baseline blood samples are taken. The men are randomized to a diet or a control group through a closed envelop. All participants receive general dietary advice according to the Swedish National Food Agency (1). Participants are instructed to avoid supplements but no other dietary restrictions are given. The diet group receive a food package including food items high in phytoestrogens. Recipes and serving suggestions for the study diet are also given. The intake of the food items gradually steps up according to a schedule and the items are eaten in specific amounts until the surgery.

The participants are contacted by phone once during the study and a 24-hour recall is performed. The interview is made to control compliance in the study group and to estimate the intake of phytoestrogens in the control group. Near to surgery a similar food frequency questionnaire is filled in and endpoint blood samples are taken. Prostate biopsies are taken at the surgery. Hormones and PSA are analyzed from baseline and endpoint blood samples. The expression of androgen receptor and estrogen receptor alfa and beta are analyzed in tumor material. Tumor proliferation is calculated by ki-67 from prostatectomy specimens at surgery.

**Figure 1**. The study design of PRODICA




NUMBER OF SUBJECTS

The person time studied will be 203 person moments. Interim analyzes will be made when there is data to analyze 100 person moments.

INCLUSION CRITERIA

- Patients with prostate cancer T1–T2, Gleason score <8, prostate-specific antigen (PSA) <20
- scheduled for radical prostatectomy.

EXCLUSION CRITERIA

Ongoing hormone therapy
Difficult physical or psychological conditions or diminished cognitive function
Allergy or intolerance of the intervention foods

## DATA MANAGEMENT

### Definition of primary outcome

***Ki-67***

Tumor proliferation rate (Ki-67) is measured in prostatectomy specimens from the surgery. The ratio of stained prostate cancer nuclei is calculated in five different areas and the median, mean, and max

values are calculated

Analyze of prostatectomy specimens from the surgery, variables:
**ki67_median** (median value of ki-67)
**ki67_mean** (mean value of ki-67)
**ki67_max** (max value of ki-67)
Above variables will be dichotomized according to the median value in the study population:ki67_median_cat (0/1)
ki67_mean_cat (0/1)
ki67_max_cat (0/1)

**Definition of secondary outcomes**

Levels of PSA are analyzed from blood samples taken at the inclusion meeting (baseline) and in connection with another visit within one week before surgery (endpoint). The difference is calculated between endpoint and baseline.

***PSA***
**PSA_total_BL** (total PSA levels at baseline)
**PSA_total_EP** (total PSA levels at endpoint)
**PSA_total_diff** (difference in total PSA levels, endpoint-baseline)
**PSA_total_diff_cat** (total PSA difference endpoint-baseline, decreased/unchanged=0, increased=1)

**PSA_ratio_BL** (PSA ratio free/total at baseline)
**PSA_ratio_EP** (PSA ratio free/total at endpoint)
**PSA_ratio_diff** (difference in total PSA levels, endpoint-baseline)
**PSA_ratio_diff_cat** (PSA ratio diff endpoint-baseline, decreased/unchanged=0, increased=1))

***Hormones***
Concentrations of testosterone, estradiol, sex hormone binding globulin
(SHBG), and IGF-1 will be analyzed in serum at endpoint and baseline. The difference is calculated between endpoint and baseline.

**Estradiol_BL** (Estradiol levels at baseline)
**Estradiol_EP** (Estradiol levels at endpoint)
**Estradiol_diff** (difference in Estradiol levels, endpoint-baseline)
**estradiol_diff_cat** (Estradiol difference endpoint-baseline, decreased/unchanged=0, increased=1)

**Testosterone_BL** (Testosterone levels at baseline)
**Testosterone_EP** (Testosterone levels at endpoint)
**Testosteron_diff** (difference in Testosteron levels, endpoint-baseline)
**testo_diff_cat** (Testosteron difference endpoint-baseline, decreased/unchanged=0, increased=1)

**IGF1_BL** (IGF1 levels at baseline)
**IGF1_EP** (IGF1 levels at endpoint)

**IGF1_difference** (difference in IGF1 levels, endpoint-baseline)
**SHBG_BL** (SHBG levels at baseline)
**SHBG_EP** (SHBG levels at endpoint)
**SHBG_difference** (difference in SHBG levels, endpoint-baseline)

**testo_estrad_ratio_diff** (ratio testosterone/estradiol difference endpoint-baseline)
**testo_SHBG_ratio_diff** (ratio testosterone/SHBG difference endpoint-baseline)

***Gene expression***
The radical prostatectomy tissues are collected at surgery and formalin fixed paraffin embedded. Expression of genes involved in proliferation and ER signal pathways and cell-cycle progression (CCP) gene expression will be analyzed to identify the expression of genes involved in tumorigenesis. Global transcriptome-wide expression array Clariom D® will be used. For analysis of the CCP score, gene expression of the genes involved in CCP based on the Prolaris® gene panel and the Decipher Score. Thereafter, guided and unguided analyses will be performed.

***Receptors expression***
One benign- and one cancer biopsies are collected at surgery. ERα, ERβ, and AR mRNA expression will be determined.
**ER_a** (mRNA levels of estrogen receptor α)
**ER_b** (mRNA levels of estrogen receptor β)
**AR** (mRNA levels of androgen receptor)

## Definition of exposure

The participants in the intervention group are defined as exposed and the participants in the control group are defined as not exposed.

**Group_ intervention** (0=control, 1=intervention)

Intake of different phytoestrogens at endpoint is based on information from the Food Frequency Questionnaire (FFQ) and the 24-h dietary recall.

**Phytoestogens_EP_3** (intake of phytoestrogens at endpoint divided in tertiles, calculated from the FFQ)

**Lignans_EP_3** (intake of lignans at endpoint divided in tertiles, calculated from the FFQ)

**Isoflavones_EP_3** (intake of isoflavones at endpoint divided in tertiles, calculated from the FFQ)

**Coumestrol_N58_EP_2** (dichotomized intake of coumesterol at endpoint, calculated from the FFQ)

**Phyto_24_recall_3** (intake of phytoestrogens at the 24-h dietary recall divided in tertiles)

The concentrations of phytoestrogens (daidzein, enterodiol, enterolactone, equol, genistein, glycetein,

lariciresinol, and secoisolariciresinol) in plasma for a sub-group of participants:

**lignans_EP_nmol** (plasma levels of lignans at endpoint=enterolactone+ enterodiol+ lariciresinol+ seco_EP_nmol)

**lignans_BL_nmol** ((plasma levels of lignans at baseline=enterolactone+ enterodiol+ lariciresinol+ seco_EP_nmol)

**isoflavones_EP_nmol** (plasma levels of isoflavones at endpoint = daidzein_EP_nmol + genistein_EP_nmol + glycitein_EP_nmol + equol_EP_nmol)

**isoflavones_BL_nmol** (plasma levels of isoflavones at baseline = daidzein_EP_nmol + genistein_EP_nmol + glycitein_EP_nmol + equol_EP_nmol)

**phytoestrogens_EP_nmol** (plasma levels of total phytoestrogens at baseline = enterolactone_EP_nmol + enterodiol_EP_nmol + lariciresinol_EP_nmol + seco_EP_nmol + daidzein_EP_nmol + genistein_EP_nmol + glycitein_EP_nmol + equol_EP_nmol)

**phytoestrogens_BL_nmol** (plasma levels of total phytoestrogens at endpoint = enterolactone_BL_nmol + enterodiol_BL_nmol + lariciresinol_BL_nmol + seco_BL_nmol + daidzein_BL_nmol + genistein_BL_nmol + glycitein_BL_nmol + equol_BL_nmol)

**lignans_diff_nmol** = lignans_EP_nmol - lignans_BL_nmol

**isoflav_diff_nmol**= isoflavones_EP_nmol - isoflavones_BL_nmol

**phyto_diff_nmol**= phytoestrogens_EP_nmol - phytoestrogens_BL_nmol

***Definition of effect modifying factor***

The individual genotype (alleles TT, TC or CC) of the estrogen-beta-gene (ERß) are identified in the blood sample at baseline.

**cat_genotype** (created from variable Genotype, 0=TT, 1= TT/CC)

## PLANS FOR STATISTICAL ANALYZES

### *VARIABLES TO ANALYZE*

Created variables (for more details see attached variable list):

BMI (kg/m$^2$):
BMI_stata = Weight_inclusion / ((Lenght_inclusion/100) * (Lenght_inclusion/100)
BMI_group = 1= under weight (BMI <18.5); 2= normal weight (BMI 18.5-24.9); 3= Overweight (BMI 25-29.9); 4= Obese class (BMI 30-34.9) 1; 5= Obese class ≥2 (BMI ≥35)

Weight change (kg):
Weight_diff = Weight_endpoint - Weight_inclusion

Compliance (1= compliance ≥80; 0= compliance <80 %):
Compliance_24_recall (based on reported intake of phytoestrogens at the 24-h dietary recall)
Compliance_intervention_24_r (based on eaten interventions foods reported from at the 24-h dietary recall)
Compliance_intervention (based on eaten interventions foods reported from at the end of the intervention)

Intake of isoflavones (µg):
Isoflavones_ BL = Genistein_N56_BL + Daidzein_N57_BL + Formononetin_N59_BL + Biochanin_N60_BL + Equol_N75_BL

Isoflavones_EP = GenisteinN56_EP + Daidzein_N57_EP + Formononetin_N59_EP + Biochanin_N60_EP + Equol_N75_EP.

Intake of lignans (µg):
Lignans_BL = SECOtot_N61_BL + Matairesinol_N62_BL Lariciresinol_N71_BL + Pinoresinol_N72_BL + Syringaresinol_N73_BL + Medioresinol_N74_BL + Enterodiol_N76_BL + Enterolactone_N77_BL

Lignans_EP=SECOtot_N61_EP + Matairesinol_N62_EP + Lariciresinol_N71_EP + Pinoresinol_N72_EP + Syringaresinol_N73_EP + Medioresinol_N74_EP + Enterodiol_N76_EP + Enterolactone_N77_EP

Intake of phytoestrogens (ug):
Phytoestogens_BL = (Isoflavones_BL + Lignans_BL + Coumestrol_N58_BL)
Phytoestogens_EP =(Isoflavones_EP + Lignans_EP + Coumestrol_N58_EP)

Level of physical activity:
Physical_activity_BL_group = low, moderate or high
1 = Low physical activity (101, 102, 103 or 201 points)
2 = Moderate physical activity (104, 202, 203, 301 or 302 points)
3 = High physical activity (204, 303, 304, 401, 402, 403 or 404 points)

Intake of fatty fish:

Fat_fish_group_BL: 0= <1 times/month; 1= 1-3 times/month; 2= ≥1 times/week
Fat_fish_group_EP: 0= <1 times/month; 1= 1-3 times/month; 2= ≥1 times/week
Fat_fish_diff: Fat_fish_month_EP - Fat_fish_month_BL

Use of antibiotics:
antiobotics_1y_dic (0= no intake last year or don't know; 1= ≥1 time last year)
antiobotics_2_5y_dic (0= no intake last 2-5 years or don't know; 1= ≥1 time last 2-5 years)
f53_1_EP: Intake of antibiotics in recent weeks at endpoint (0= No; 1=Yes)
antibiotics_all (0= no intake during intervention and last 5 years or don't know; 1= ≥1 time during intervention and last 5 years)

Smoking:

0= never smoked or quitted smoking >5 years ago
1= Current smoker or quitted smoking ≤5 years ago

**STATISTICAL MODELS**

Difference between groups will be tested for variables stated below. Independent T-test or Mann-Whitney U-test will be used for continuous variables depending on if the data has a normal or skewed distribution. Fishers's exact test will be used for categorical variables.

Covariates to test (possible confounding factors):

- Energy_kcal_EP_3 (total energy intake at endpoint, categorized in tertiles)
- Sum_satured_fat_EP_3 (sum of saturated fats (g) at endpoint, categorized in tertiles)
- Fat_fish_group_EP (categorized intake of fat fish at endpoint)
- Calcium_EP_3 (intake of calcium at endpoint, categorized in tertiles)
- Physical_activity_BL_group (categorized physical activity at baseline)
- f47_1_EP (changed physical activity during intervention)
- Biopsies_cancer (percent of biopsies with cancer at diagnosis)
- gleason_cat (categorized Gleason score at diagnosis)
- gleason_cat_rp (categorized Gleason score at surgery)
- rp_teritargradjanej (categorized tertiary Gleason degree at surgery)
- Age_inclusion (continuous age at baseline)
- BMI_group (BMI at baseline in groups)
- Weight_diff (difference in weight between endpoint and baseline)
- smoking (categorized smoking at baseline)
- heredity (categorized heredity at baseline)
- Zinc_EP_3 (intake of zinc at endpoint, categorized in tertiles)
- Alcohol_EP_3 (intake of alcohol at endpoint, categorized in tertiles)
- Sum_monosat_fat_EP_3 (intake of monounsaturated fatty acids at endpoint, categorized in tertiles)
- Sum_polysat_fat_EP_3 (intake of polyunsaturated fatty acids at endpoint, categorized in tertiles)
- red_meat_EP_3 (intake of red meat at endpoint, categorized in tertiles)
- vegetables_all_EP_3 (intake of vegetables at endpoint, categorized in tertiles)
- fruit_all_EP_3 (intake of fruits at endpoint, categorized in tertiles)
- Vit_E_EP_3 (intake of vitamin E at endpoint, categorized in tertiles)

- Finasteride (intake of Finasteride under intervention, no=0, yes=1)

Pearson or Spearman correlation coefficient will be used to evaluate whether dietary covariates are correlated. If the correlation coefficient between two covariates in the model or between covariates and the main exposure are higher than 0.6, multicolinearity issues is considered, eventually one of the covariates will be excluded from the model.

Model building:
The association between intake of phytoestrogens and tumor proliferation will evaluated by regression models, providing estimate of risk difference (RD) relative risk (RR) and corresponding 95% confidence intervals.

Crude models:

C0: Ki-67

C1: C0 + Group_intervention

C2: C0 + Phytoestogens_EP

C3: C0 + Isoflavones_EP

C4: C0 + Lignans_EP

C5: C0 + Coumestrol_N58_EP

$C0_2$: PSA_total_diff

$C1_2$: $C0_2$ + Group_intervention

$C2_2$: $C0_2$ + Phytoestogens_EP

$C3_2$: $C0_2$ + Isoflavones_EP

$C4_2$: $C0_2$ + Lignans_EP

$C5_2$: $C0_2$ + Coumestrol_N58_EP

$C0_3$: PSA_ratio_diff

$C1_3$: $C0_3$ + Group_intervention

$C3_3$: $C0_3$ + Phytoestogens_EP

$C4_3$: $C0_3$ + Isoflavones_EP

$C5_3$: $C0_3$ + Lignans_EP

$C6_3$: $C0_3$ + Coumestrol_N58_EP

$C0_4$: estradiol_diff_cat

$C1_4$: $C0_4$ + Group_intervention

$C3_4$: $C0_4$ + Phytoestogens_EP

$C4_4$: $C0_4$ + Isoflavones_EP

$C5_4$: $C0_4$ + Lignans_EP

$C6_4$: $C0_4$ + Coumestrol_N58_EP

$C0_5$: testo_diff_cat

$C1_5$: $C0_5$ + Group_intervention

$C3_5$: $C0_5$ + Phytoestogens_EP

$C4_5$: $C0_5$ + Isoflavones_EP

$C5_5$: $C0_5$ + Lignans_EP

$C6_5$: $C0_5$ + Coumestrol_N58_EP

$C0_6$: testo_estrad_ratio_diff

$C1_6$: $C0_6$ + Group_intervention

$C3_6$: $C0_6$ + Phytoestogens_EP

$C4_6$: $C0_6$ + Isoflavones_EP

$C5_6$: $C0_6$ + Lignans_EP

$C6_6$: $C0_6$ + Coumestrol_N58_EP

$C0_7$: testo_SHBG_ratio_diff

$C1_7$: $C0_7$ + Group_intervention

$C3_7$: $C0_7$ + Phytoestogens_EP

$C4_7$: $C0_7$ + Isoflavones_EP

$C5_7$: $C0_7$ + Lignans_EP

$C6_7$: $C0_7$ + Coumestrol_N58_EP

Final model:

Which covariates included in the model will be based on:

1) If a known risk factor for the outcome (based on previous subject matter knowledge and with the help of Directed Acyclic Graphs)
2) If differs between groups of exposure
3) proportional (≥10%) change in ß-coefficients

Stratification by ERß alleles and test of interaction:

Analysis of hypothesis above (C0- $C0_7$ and corresponding models) will be stratified by genotype of

ERß (variable: cat_genotype).

Interactions between exposure and ERß SNPs on the outcome will be evaluated on the additive effect scales. Effects will be measured by RDs and interaction will be assessed in a generalized linear model by the product term between the covariates representing phytoestrogen intake and SNP genotypes.

Product term: int1=intervention*cat_genotype

**Additional analysis:**

*Intention-to-treat-analyses*: All participants will be included.

*Per-protocol-analyzes*: All participants with ≥80 % compliance will be included.

In the intervention group, compliance will be considered as ≥80 percent of the recommended intake of 200 mg of phytoestrogens, which is equal to ≥160 mg. In the control group, intakes below 160 mg of phytoestrogens will be considered compliant.

*Analysis of hypothesis will be stratified for*:
– no use of antibiotics vs. use of antibiotics
– no use of finasteride vs. use of finasteride

*Investigate the effects of the intervention diet on plasma concentrations of phytoestrogens:* Cohen's Kappa will be used to compare the agreement of classifications in tertiles from estimated intake of phytoestrogens and plasma concentrations of phytoestrogens. (Kappa values <0 will be considered as poor, 0.00-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1.00 as almost perfect). Boxplots will be used to present the data graphically. Linear regressions will be used to investigate the relationships between estimated intake concentrations (explanatory variables) and plasma concentrations of phytoestrogens (outcomes). The regression model will be stratified according to the intervention- and control groups and genotype.

*Receptor expression*
The mRNA levels of ERα, ERβ, and AR will be compared between intervention- and control groups and groups of genotype.

**AUTHORS**

Rebecca Ahlin, rebecca.ahlin@vgregion.se, 076-618 5020

Maria Hedelin, maria.hedelin@oncology.gu.se, 070-7401858

**REFERENCES**

1. Livsmedelsverket. Kostråden - hitta ditt sätt. [2018-03-07 cited]; Available from: https://www.livsmedelsverket.se/matvanor-halsa--miljo/kostrad-och-matvanor/rad-om-bra-mat-hitta-ditt-satt/.

**APPENDICES**

Appendics 1 - Variable list

Appendics 2 - Food Frequency Questionnaire