WHITE PAPER

cool clinical
Consortium for Clinical Science and AI

Machine Learning for Precision Medicine and Drug Development in Prostate Cancer

Part 3: Bicalutamide Clinical Trial Data and Preparation of Machine Learning Datasets

Felix Beacher Ph.D & Leonardo Ancora M.Sc
January 2021

About Cool Clinical

Cool Clinical is a non-profit consortium of clinical and computational scientists. Cool Clinical publishes original articles and reports for businesses, the public sector, NGOs and the general public. Our goal is to advance the conversation on AI applications to clinical science.

To learn more about the research of Cool Clinical, visit www.coolclinical.com

Contents

Introduction	4
Sources of Data	4
Objectives of the Original Studies	6
Design of the Original Studies	6
Subject population	8
Study Assessments	8
Derivation of Target Variable	11
Overall Data Preprocessing Steps	14
Final Datasets	14
References	19
Appendix 1	20
Appendix 2	22

Introduction

The aim of the analyses presented in this series of white papers is to illustrate the use of machine learning (ML) for precision medicine and predictive enrichment for clinical trials. The aim of the kind of analyses shown here is to demonstrate the ability to select subjects who would most benefit from a drug treatment. This could lead to improved efficacy and drug safety, compared to a non-selected population and could improve the chances of regulatory approval for an experimental drug.

Predictive enrichment could be implemented in various ways, for example:

1. *Using phase II data to select subjects for a phase III trial, in the same drug pipeline.*
    - The advantage of this is that the study measurements and interventions can be identical, providing optimal predictive power
    - The disadvantage of this is that phase II studies tend to have smaller samples which could limit the usefulness of ML models
2. *Using publicly available phase III data to select subjects for a new phase III trial*
    - The advantage of this is that phase III studies have larger samples which are required for ML models
    - The disadvantage of this is that the study measurements and interventions are unlikely to be identical, limiting predictive power
3. *Using phase III data from a failed phase III trial to select subjects for a new phase III trial*

Such methodologies would need to be carried over into clinical practice. This means that if a drug is approved on the basis of selection by an algorithm, it would presumably only be approved with respect to the same population used in the trial, i.e., the drug would only be approved for people identified as appropriate for that treatment by the same algorithm.

The aims of this Part 4 white paper are to:

1. Describe the studies and data used for subsequent analyses
2. Describe the building of the datasets and the planned models presented in subsequent white papers in this series

Sources of Data

The sources of data for the analyses described in this series of white papers are the Astra-Zeneca Early Prostate Cancer (EPC) clinical program. This program consisted of three phase III clinical trials on bicalutamide or placebo in subjects with prostate cancer (codes 0023, 0024 and 0025). The data, data codes, study protocol and case report form were obtained through Project Data Sphere. The main features of these studies are summarized in Table 1.

Table 1. Summary of the three Bicalutamide prostate cancer studies in the Astra-Zeneca Early Prostate Cancer (EPC) clinical program, the sources of data for the analyses described in this series of white papers.

	Study 1	Study 2	Study 3
Title	A Randomized Double-Blind Comparative Trial of Bicalutamide (Casodex) Versus Placebo in Patients with Early Prostate Cancer	A Randomised, Double-blind, Parallel-group Trial Comparing Casodex 150 mg Once Daily With Placebo In Subjects With Non-metastatic Prostate Cancer	A Randomised, Double-blind, Parallel-group Trial Comparing Casodex 150 mg Once Daily With Placebo In Subjects With Non-metastatic Prostate Cancer
Code	0023	0024	0025
Sites	US and Canada	Mostly Europe, but also Mexico, Australia and South Africa	Sweden, Norway, Finland, Denmark
Total N (not available)	3292	3603	1218
Drug arm N (available)	1645	1805	611
First subject enrolled	1995	1995	1995
Study completed	2008	2008	2008

In 2000 the data from the Study 1 were analysed. The data monitoring and safety committee (DMSC), ruled that there were no significant safety concerns with Casodex (the most common side effects were gynaecomastia and breast pain, each occurring in approximately 70% of subjects on Casodex). The analysis also indicated that the risk of disease progression was lower in subjects on Casodex than placebo (9.0% of subjects on Casodex progressed c.f. 13.8% of subjects on placebo). The DMSC therefore allowed the optional unblinding of randomized therapy and subjects were allowed to change to open-label Casodex at the investigator's discretion.

A second analysis was conducted in 2003, which considered the pooled data from the three EPC studies. The median follow-up was 5.4 years. This analysis indicated that Casodex was associated with reduced risk of disease progression, both in the pooled data and in trials 0024 and 0025, individually. No significant differences between Casodex and placebo were observed for survival, either in the pooled analysis or the individual trials (only around 15% of patients had died, limiting the ability to detect statistical differences). Subgroup analysis indicated that subjects with locally advanced disease (who would normally be managed with watchful waiting), survival may favor subjects on Casodex (the hazard ratio was 0.81, a trend but not statistically significant). In subjects with localized disease, survival may favor subjects on placebo (the hazard ratio was 1.23, a trend but not statistically significant). Following the second analysis, all randomized therapy was ended and subjects were given the option of receiving open-label Casodex.

The datasets publicly released only contain data for participants assigned to the drug arm, i.e., none of the participants assigned to the placebo arm.

Objectives of the Original Studies

The stated objectives of these studies varied slightly between protocols but were essentially equivalent and can be summarized as follows:

1. To compare **time to clinical progression** after 2 years of adjuvant bicalutamide 150 mg monotherapy vs placebo
2. To compare **overall survival** after 2 years of adjuvant bicalutamide 150 mg monotherapy vs placebo
3. To evaluate **tolerability** of 2 years of bicalutamide 150 mg therapy versus placebo
4. To compare **treatment failure** after 2 years of adjuvant bicalutamide 150 mg monotherapy vs placebo.

Design of the Original Studies

The three studies all had a randomized, double-blind, parallel group design, comparing Casodex 150 mg once daily with placebo. Subjects were randomized 1:1 to Casodex or placebo (see Figure 1). Subjects received trial therapy for up to 2 years. After 2 years, patients were treated at the investigator's discretion. See Figure 1.

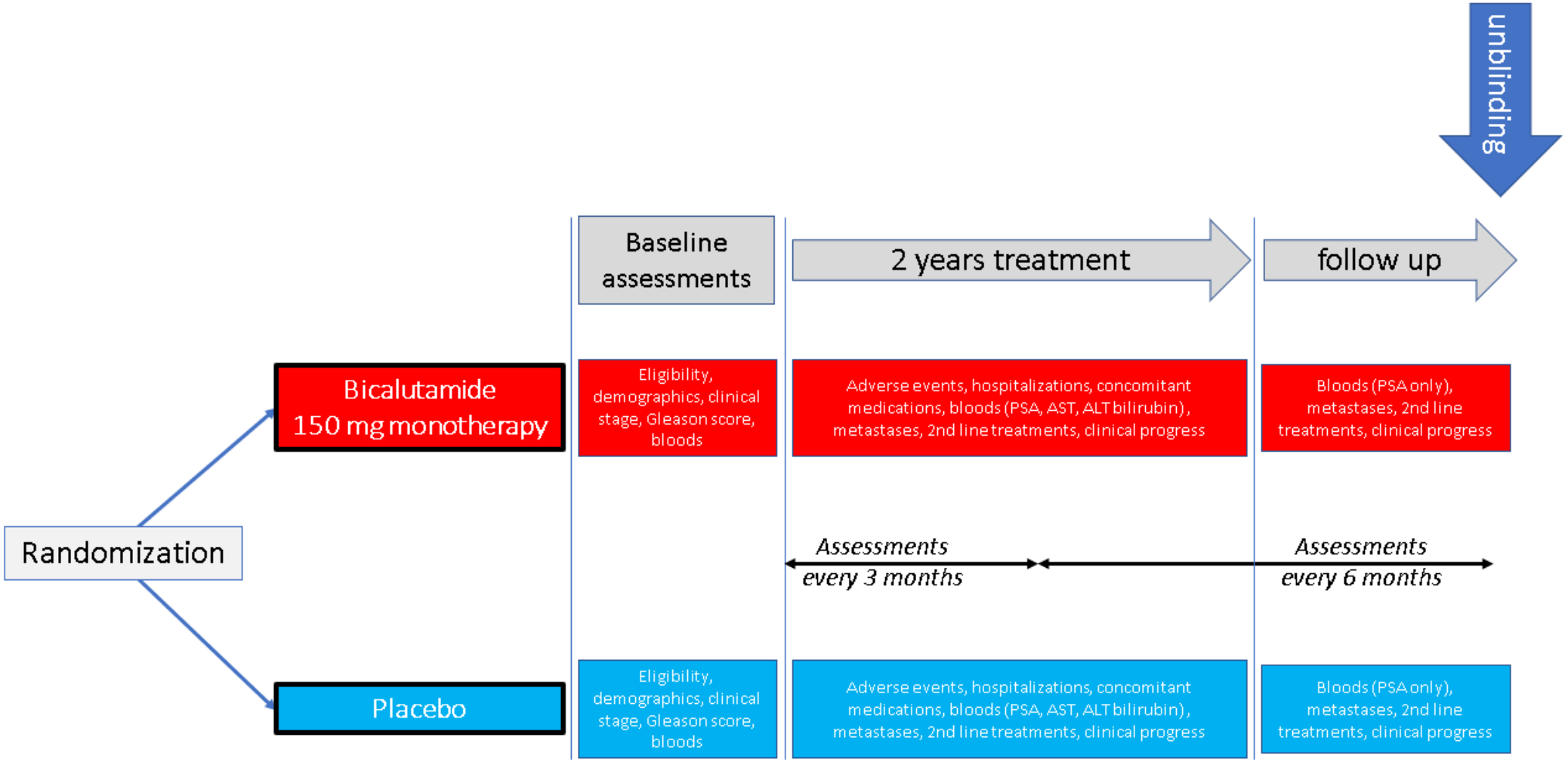

Figure 1. Outline of study design for the three studies included in the analyses presented in this paper.

Subject population

Inclusion and exclusion criteria were similar for the three studies, however there were some differences. The key inclusion and exclusion criteria are given in Table 2.

Table 2. Inclusion and exclusion criteria for the three Astra-Zeneca EPC studies.

<table>
<tr><th></th><th>Study 1</th><th>Study 2</th><th>Study 3</th></tr>
<tr><td rowspan="3">Inclusion criteria</td><td colspan="3">Adenocarcinoma of the prostate</td></tr>
<tr><td>'Absence of metastatic disease'</td><td colspan="2">'non distant metastatic disease'</td></tr>
<tr><td>Radical prostatectomy OR prostate radiation treatment within 16 weeks before randomization.</td><td></td><td></td></tr>
<tr><td rowspan="6">Exclusion criteria</td><td colspan="3">Any previous systemic therapy for prostate cancer except: 1. Neoadjuvant therapy* prior to primary therapy 2. 5-alpha-reductase inhibitors</td></tr>
<tr><td colspan="2">Other type of cancer (except treated skin carcinoma) within the last five years.</td><td>Other type of cancer (except treated skin carcinoma) within the last ten years.</td></tr>
<tr><td colspan="3">Serum bilirubin, AST or ALT >2.5x normal upper limit</td></tr>
<tr><td colspan="3">Any severe medical condition that could jeopardize trial compliance</td></tr>
<tr><td></td><td>treatment with a new drug within previous 3 months</td><td>patients for whom long term therapy is inappropriate due to low expected survival times</td></tr>
<tr><td></td><td colspan="2">Patients at risk of transmitting infection through blood, including AIDS or other STDs or hepatitis</td></tr>
</table>

* 'Neoadjuvant therapy' refers to treatment given as a first step to shrink a tumor before the main treatment, usually surgery. Examples of neoadjuvant therapy include chemotherapy, radiation therapy, and hormone therapy.

Study Assessments

The three EPC studies included a range of assessments, typical for a phase III clinical trial, including:

- Demographic features
- Physical examination
- Bone metastases via bone scans at 96 weekly intervals.
- Laboratory assessments (biochemistry, hematology, prostate–specific antigen; PSA).
- Adverse events
- Survival
- Concomitant medications

Subjects received evaluations starting at 12-week intervals. After one year for each study, the quantity of data collected was determined to be sufficient to assess all non-critical study outcome variables. Therefore, after one year from baseline, evaluations were of essential data only, namely: survival, and clinical progression status.

Concomitant Medications

Generic name and PCLA drug class of concomitant medications being taken at the trial start were included as features in all three datasets. For study 1, indications being taken at the trial start (associated with the concomitant medications) were also recorded and included as features for Study 1.

Laboratory Blood Tests

Prostate Specific Androgen (PSA)

Prostate Specific Androgen (PSA) is described in Part 2 of this series. However, it is noteworthy that PSA levels tend to decrease with androgen therapy. For example, 50mg daily bicalutamide for 24 weeks is associated with a median 56% reduction in PSA levels (Eri and Tveter, 2001).

To calculate the 2-year PSA values, a window of 600-900 days after baseline was used. In the case of multiple values in this window the value associated with the day closest to the 2-year point was used.

Aspartate aminotransferase (AST) and alanine aminotransferase (ALT)

Liver enzymes are frequently checked during inpatient care. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are major circulating enzymes in serum, both important for amino acid metabolism. AST is found in the liver, heart, skeletal muscle, kidneys, brain, and red blood cells. ALT is most common in the liver. Serum AST and ALT levels, and the AST/ALT ratio, are commonly used measures of liver health.

AST/ALT ratio is typically used to indicate various forms of liver disease:

- Alcoholic fatty liver disease
    - AST > 8 times the upper limit of normal (ULN)
    - ALT > 5 times the ULN
- Nonalcoholic fatty liver disease
    - AST and ALT > 4 times the ULN
- Acute viral hepatitis or toxin-related hepatitis with jaundice
    - AST and ALT > 25 times the ULN
- Ischemic hepatopathy (ischemic hepatitis, shock liver)
    - AST and ALT > 50 times the ULN (in addition the lactate dehydrogenase is often markedly elevated)
- Chronic hepatitis C virus infection
    - Wide variability, typically normal to less than twice the ULN, rarely more than 10 times the ULN
- Chronic hepatitis B virus infection
    - Levels fluctuate
    - the AST and ALT may be normal, though most patients have mild to moderate elevations (approximately twice the ULN)
    - with exacerbations, levels are more than 10 times the ULN

AST/ALT ratio can also predict malignant tumors such as pancreatic cancer, breast cancer and prostate cancer (Zhou et al 2020).

Bilirubin

Bilirubin is a by-product of normal hemoglobin metabolism. High blood levels of bilirubin (hyperbilirubinemia) can be caused by conditions such as jaundice, but also pharmaceutical drugs (especially antipsychotic drugs). Hyperbilirubinemia may be associated with prostate cancer (Gokcen et al 2019). Hyperbilirubinemia is also a common side effect of chemotherapy.

Comparability of the Three Studies

The studies used in our analyses, from the Astra-Zeneca Early Prostate Cancer (EPC) clinical program, were designed to yield data that could be pooled. The comparability of the studies is important to assess the reasonableness of pooling the data. The three studies had similar designs and collected similar types of data. The three studies were also similar in terms of

- Age
- Racial makeup (predominantly white, although large differences in frequencies of minority races)
- Clinical stage (with late stages predominating)
- Gleason score category (the three studies having same median value)
- Proportion of serious related AEs (although the reasons given were very different between the studies)
- Proportion of death within 2 years (this was lower in Study 1 but was less than 2% for all three studies)

However, there were differences between the three studies in some variables, for example:

- Mean PSA levels over the two-year main trial period were highest in Study 3, next highest in Study 2 and lowest in Study 1.
- There were pronounced differences in bilirubin values between studies 1 and studies 2 and 3 (means were 0.6, 9.4 and 10.1 respectively). The degree of these differences is so great that they presumably reflect some unknown measurement bias or differences in unit scaling. To address for these differences, we zero-normalized the bilirubin values on the basis of each study.
- Prostatectomy in the 16 weeks before baseline was much more common in Study 1 than Studies 2 and 3 (see Table 9), reflecting different inclusion criteria (see Table 2).
- Metastatic node disease at baseline was much more common in Study 1 than Studies 2 and 3 (see Table 9).

The differences in PSA levels requires special comment, because these were pronounced, as shown in Figure 2. The reason for the overall differences in PSA levels is unclear from the study protocols. However, given that the levels are stable for all three studies to a similar degree, it was considered appropriate to use an absolute threshold for assessing PSA increases and that this threshold should be the same size (any change greater than 0.2 ng/mL).

In study 1 there was a sharp decrease in mean PSA levels comparing baseline with the three-month values, but this decrease was not apparent for Studies 2 and 3. This PSA decrease in Study 1 is likely to relate to the inclusion criterion for Study 1 that subjects had received either radical prostatectomy or radiation treatment to the prostatic bed within 16 weeks before randomization. This inclusion criterion

was not present for Studies 2 and 3. This means that for Study 1 patients would typically have recently had had treatment which would have then caused a sharp decrease in PSA levels. This explains the sharp reduction in PSA levels in Study 1 but not Studies 2 and 3. However, average PSA levels in Study 1 were relatively stable after the initial post-baseline assessment at 3 months. Therefore, to provide a baseline comparable between the three studies, the baseline values for Study 1 were imputed from the three-month point.

Data from these trials were intended to be considered individually and also as part of a meta-analysis. However, there were some differences in data collection methods and assessments. These differences are relevant because it is only assessments in common that can be used in modelling where data across datasets was used. In some cases, differences in assessments could be resolved by converting variables from one form to another (e.g., Gleason scores could be converted into Gleason categories).

Derivation of the Target Variable

Ideal targets for drug efficacy are those with clear medical significance, such as death, time to death, progression to a more advanced clinical stage ('clinical progression'), or time to clinical progression. However, for the datasets used here, the numbers cases of death and clinical progression, even with pooling the three datasets, is relatively small and provide a challenge for ML analysis. To address this, for the primary analyses, we used PSA levels as a biomarker/proxy of disease progression. In addition to the PSA-based definition of 'good responders', we used clinical progression-based definition of 'Good responders' (Table 3).

The principal target variable was named 'Good responders' and combined both efficacy and drug toleration. A subject who was a 'good responder' was defined by the following conditions:

1. No increase in PSA levels above 0.2ng/mL (baseline vs. two year follow up)
    - PSA increase was defined in terms of a comparison of baseline with PSA values at 2 years. For Studies 2 and 3, baseline was day 0, or the day of randomization. For Study 1, the baseline was set at three months, rather than day 0. This was to address the observation that in Study 1 mean PSA values underwent a marked reduction from day 0 to 3 months (from 1.16 nm/mL to 0.66 nm/mL; a 43% reduction) and then stabilized. Thus, the use of a baseline at the three-month point was designed to create a baseline which represented the start of stable PSA values. That this sharp reduction in PSA values was observed in Study 1 but not Studies 2 or 3 is likely related to the fact that Study 1 required subjects to have undergone radical prostatectomy or radiation treatment to the prostate within 16 weeks before randomization. Thus, in many subjects the initial PSA decrease would have reflected these recent interventions.
    - PSA values for the 2-year point after randomization was operationalized by taking the mean of PSA values within a window of 600-900 days after day 0 (see Figure 2). The width of this window was designed to balance the requirement to include a sufficient number of data points, with the need to confine the window in such a way that the data could be considered to be consistent in terms of time.
    - PSA increases were only considered if they exceeded a threshold 0.2 ng/mL, similar to the method used in the literature (Okubo et al 2018).

2. Tolerated drug (absence of related serious adverse events; AEs)
    - A related AE was considered serious if it:
        1. caused withdrawal from the study OR
        2. caused disability or incapacity OR

3. required or prolonged hospitalization OR
4. required intervention to prevent impairment OR
5. was life-threatening

- Only AEs within the two-year main trial period were considered.

3. Stayed in trial for at least 600 days (i.e., was not a 'dropout')

4. Survival for two years from the start of the trial

Thus, any subject who met these criteria was considered to be a 'good responder'. Conversely, any subject who failed to meet these criteria was a 'bad responder'. Subjects were excluded from the analysis if there were insufficient data to evaluate these conditions. Usually, these subjects were excluded due to the lack of data on PSA levels.

The variable 'good responders' was relatively well balanced in terms of class frequencies (i.e., similar numbers of 'good responders' and 'bad responders').

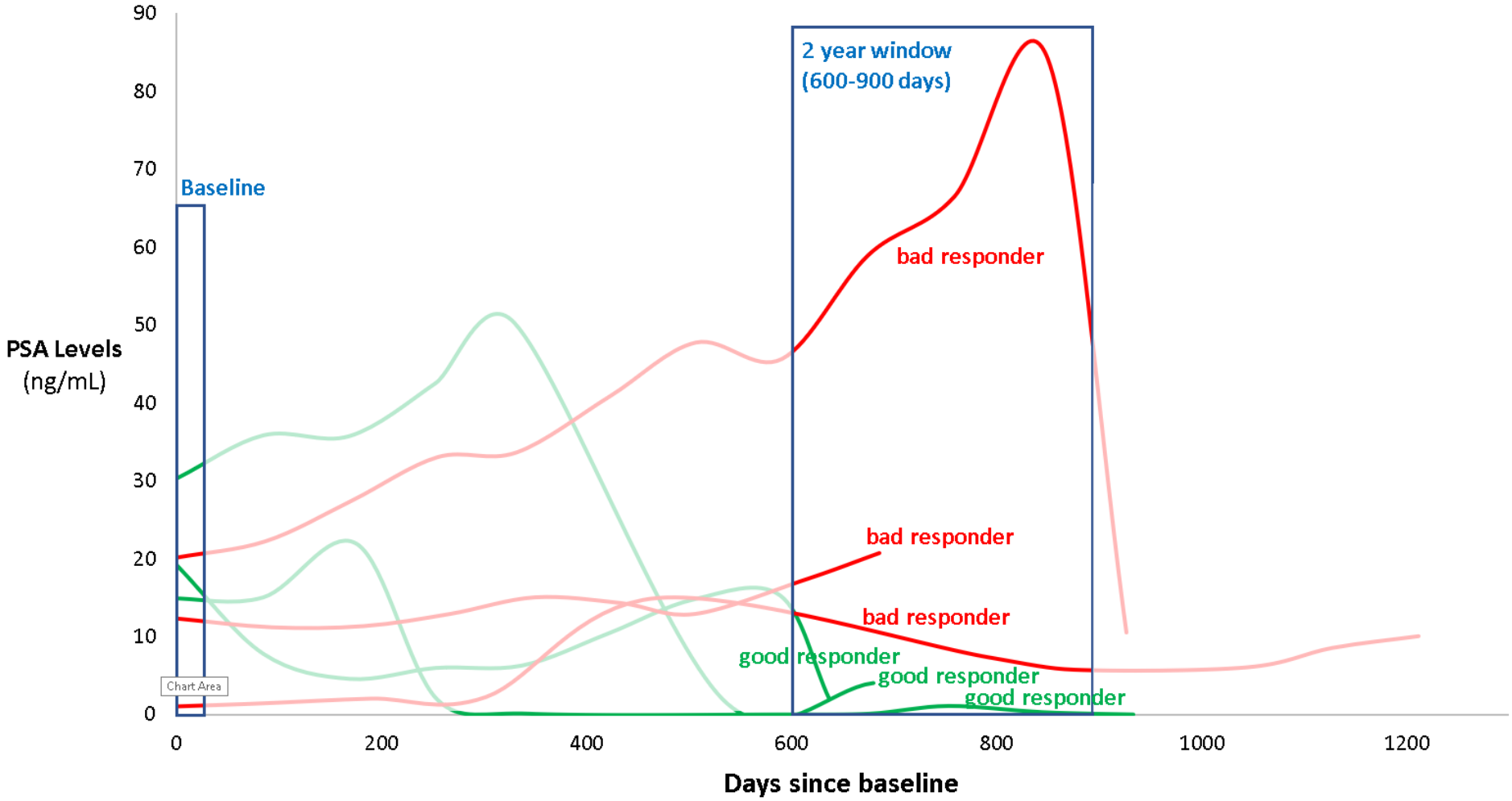

Figure 2. Illustration of good and bad responders in terms of the trajectories of PSA levels over time, for a sample of 6 subjects from Study 2. Good and bad responders were defined with respect to differences in PSA levels, comparing baseline and the mean of values within the time window 600-900 days after baseline. NB, the computation of the target variable 'good responders' also involved the consideration of additional factors; see section 'Target Variables'

Overall Data Preprocessing Steps

The main preprocessing steps were the following:

1. Conversion of sas files to csv format.
2. Selection of features of likely interest: features were excluded if they contained no variation, had less than 10 positive cases, were blank, uninformative and or had descriptions which were unclear.
3. Conversion of categorical variables to numerical form (e.g. race, age group and sex).
4. Feature extraction (e.g. baseline levels of laboratory bloods) and feature engineering (e.g. the target variable 'good responders').
5. Feature reduction (see below).
6. Identification of subjects who received less than 600 days of treatment ('dropouts'), to be excluded or included in different models.
7. Subjects were removed if the target variable could not be computed for them.
8. Imputation of missing values (in general, median as used for imputation, except for blood biochemistry values, for which mean values were used).
9. Compilation of transformed variables into a single csv dataset with one row per subject.
10. Quality control of data via spot checks of subjects for all variables.
11. Outlier inspection (this revealed some extreme values, but none which were implausible, so no extreme values were removed).
12. Zero-normalization of the data.

Final Datasets

Preprocessing resulted in the creation of three principal datasets, one for each of the studies, from which modelling was conducted. After removal of subjects for whom 'good responders' could be computed, the datasets had the following numbers of subjects:

Study 1 - 1625 subjects
Study 2 - 1440 subjects
Study 3 - 588 subjects

From these principal datasets, new datasets were created for the different models.

Table 8: Dataset target variables and features for Studies 1, 2 and 3, excluding indications (Study 1) and concomitant medications (generic drug name and PCLA drug class). Concomitant medications and indications are presented separately in Appendix 1.

Category	Feature	Frequencies/means					
		Study 1		Study 2		Study 3	
		Values	*Missing*	*Values*	*Missing*	*Values*	*Missing*
Target variable	Good responders	1155 of 1625	n/a	434 of 1440	n/a	94 of 588	n/a
Assessments used to compute 'good responders'	Dropouts (<600 days of baseline)	325	0	428	0	155	0
	PSA baseline (ng/mL)	Mean 0.65	62	Mean 11.5	24	Mean 24.8	0
	PSA 2 years (ng/mL)	Mean 0.60	139	Mean 22.5	305	Mean 32.6	112
	Prostate cancer death in 2y	3	0	25	0	10	0
	Related AE led to withdrawal 2y	73	0	76	0	17	0
	Related AE led to interventn 2y	1	0	18	0	10	0
	Related AE led to hospitalizn 2y	2	0	19	0	6	0
	Related AE led to disability 2y	0	0	21	0	10	0
	Related AE life threatening 2y	0	0	12	0	3	0
	Any Serious Related AE in 2y	74	0	98	0	21	0
Demographics	Age (years)	Mean 64	0	Mean 69	0	Mean 68	0
	Race	Cauc 1372	0	Cauc 1365	0	Cauc 584	0
		Black 188	0	Black 13	0	Black 0	0
		Asian 12	0	Asian 7	0	Asian 0	0
		Hispanic 46	0	Hispanic 25	0	Hispanic 2	0
		Other 7	0	Other 31	0	Other 2	0

Clinical features	Clinical stage category	Cat 1 0	0	Cat 1 1	0	Cat 1 0	0
		Cat 2 23	0	Cat 2 153	0	Cat 2 44	0
		Cat 3 131	0	Cat 3 245	0	Cat 3 89	0
		Cat 4 1028	0	Cat 4 550	0	Cat 4 229	0
		Cat 5 440	0	Cat 5 446	0	Cat 5 213	0
		Cat 6 3	0	Cat 6 45	0	Cat 6 13	0
	Metastatic node disease	462	0	41	0	24	0
	Neoadjuvant therapy	359	0	No data	n/a	No data	n/a
	Gleason score category	Cat 1 78	0	Cat 1 477	0	Cat 1 256	0
		Cat 2 788	0	Cat 2 574	0	Cat 2 266	0
		Cat 3 759	0	Cat 3 363	0	Cat 3 63	0
		Cat 4 0	0	Cat 4 26	0	Cat 4 3	0
Previous therapies for prostate cancer	Prostatectomy	1314	0	534	0	71	0
	Radiotherapy	319	0	295	0	30	0
	Brachytherapy	No data	n/a	17	0	0	0
	Watchful waiting	No data	n/a	630	0	491	0
Biochemistry	PSA before baseline (ng/mL)	Mean 9.8	1068	Mean 17.3	650	Mean 23.8	499
	Baseline AST (U/L)	Mean 18.8	14	Mean 20.8	22	Mean 22.4	7
	Baseline ALT (U/L)	Mean 20.1	14	Mean 22.1	21	Mean 21.5	7
	Baseline Bilirubin (UMOL/L)	Mean 0.6	14	Mean 9.4	22	Mean 10.1	7
	Baseline Creatinine (UMOL/L)	No data	n/a	No data	n/a	Mean 153	12
	Baseline Alk Phosphatase (U/L)	No data	n/a	No data	n/a	Mean 232	12

Medical abnormalities	Head eyes ears nose throat	No data	n/a	364	0	116	0
	CNS	No data	n/a	94	0	75	0
	Blood, lymph	No data	n/a	49	0	16	0
	Skin, hair	No data	n/a	148	0	51	0
	Cardiovasc	No data	n/a	430	0	152	0
	Respiratory	No data	n/a	189	0	76	0
	Gastrointestinal	No data	n/a	551	0	203	0
	Genitourinary	No data	n/a	498	0	167	0
	Musculoskeletal	No data	n/a	475	0	196	0
	Endocrine	No data	n/a	68	0	26	0
	Allergic	No data	n/a	148	0	22	0
	Psychiatric	No data	n/a	38	0	18	0
	Drug alcohol	No data	n/a	34	0	13	0
	Other	No data	n/a	725	0	108	0
Physical exam	Weight (kg)	Mean 85	118	Mean 79	54	Mean 81	6
	Height (cm)	No data	n/a	Mean 172	88	Mean 176	3
Physical abnormalities	Head	No data	n/a	43	0	7	0
	Eye	No data	n/a	263	0	60	0
	Ear	No data	n/a	124	0	34	0
	Nose	No data	n/a	21	0	9	0
	Throat	No data	n/a	36	0	7	0
	Neck	No data	n/a	20	0	15	0
	Lymph node	No data	n/a	8	0	5	0
	Breast	No data	n/a	35	0	4	0
	Heart	No data	n/a	585	0	206	0
	Abdomen	No data	n/a	378	0	70	0
	Lung / thorax	No data	n/a	220	0	65	0
	Genito-urinary	No data	n/a	497	0	123	0
	Extremities	No data	n/a	172	0	50	0
	Musculo-skeletal	No data	n/a	426	0	122	0
	Skin	No data	n/a	156	0	50	0
	Neurologic	No data	n/a	137	0	45	0

Sexual function (medians)	Sex more than twice a week	No data	n/a	No data	n/a	Never	56
	Easily aroused	No data	n/a	No data	n/a	Hardly ever	61
	Fail to get erection in sex	No data	n/a	No data	n/a	Hardly ever	66
	Fail to get erection in foreplay	No data	n/a	No data	n/a	Usually	75
	Go weeks without sex	No data	n/a	No data	n/a	Hardly ever	75
	Lose erection during sex	No data	n/a	No data	n/a	Nvr/Hdly ever	100

References

Eri and Tveter 2001. Effects of bicalutamide and leuprolide on prostate-specific antigen (PSA), acid phosphatase (ACP) and prostatic acid phosphatase (PAP) in men with benign prostatic hyperplasia (BPH). Prostate Cancer Prostatic Dis;4(3):173-177.

Zhou et al 2020. AST/ALT ratio as a significant predictor of the incidence risk of prostate cancer. Volume 9, Issue15 Pages 5672-5677.

Gökçen et al 2019. Paraneoplastic hyperbilirubinemia in metastatic prostate cancer and review of the current literature. Turk J Urol, 45(1): 70–72.

Okubo et al 2018. Two years of bicalutamide monotherapy in patients with biochemical relapse after radical prostatectomy. Japanese Journal of Clinical Oncology, Volume 48, Issue 6, Pages 570–575.

Appendix 1

Frequencies of Indications (Study 1) and concomitant medications (all three studies)

	Name	Frequencies		
		Study 1	Study 2	Study 3
Indication	Hypertension	586	No data	No data
	Prophylaxis	422	No data	No data
	Supplement	231	No data	No data
	High cholesterol	221	No data	No data
	Arthritis	184	No data	No data
	Heart disease	142	No data	No data
	Diabetes	114	No data	No data
	Acid reflux	71	No data	No data
	Allergies	62	No data	No data
	Glaucoma	59	No data	No data
	Impotence	56	No data	No data
	Incontinence	55	No data	No data
	Constipation	54	No data	No data
	Gout	54	No data	No data
	Headache	52	No data	No data
	Insomnia	53	No data	No data
	Anxiety	47	No data	No data
	Depression	45	No data	No data
	Asthma	40	No data	No data
	Ulcer	36	No data	No data
	Back pain	34	No data	No data
	Hypothyroidism	31	No data	No data
	Hernia	28	No data	No data
	Urinary tract infection	27	No data	No data
	Angina	26	No data	No data
	Post-op pain	20	No data	No data
	Diarrhea	21	No data	No data
	Afibrillation	20	No data	No data
	General aches and pains	20	No data	No data
	Arrhythmia	19	No data	No data
	COPD	17	No data	No data
	Urinary symptoms	18	No data	No data
	Cold	14	No data	No data
	Sinus infection	15	No data	No data
	Benign prostatic hyperplasia	14	No data	No data
	Anemia	12	No data	No data
	Pain	12	No data	No data
	Hemorrhoids	11	No data	No data
	Bladder outlet obstruction	10	No data	No data

Generic name	Amlodipine	61	46	26
	Aspirin	385	247	74
	Atenolol	76	84	31
	Diazepam	18	10	11
	Digoxin	51	40	15
	Doxazosin	47	15	16
	Enalapril	64	82	27
	Furosemide	35	37	29
	Glibenclamide	55	39	15
	Glyceryl trinitrate	36	54	28
	Lisinopril	84	21	13
	Metoprolol	36	38	36
	Nifedipine	83	102	14
	Omeprazole	32	38	8
	Paracetamol	108	34	10
	Salbutamol	40	52	14
	Simvastatin	49	27	19
	Vitamin, not otherwise specified	255	14	0
	Warfarin	42	21	21
PCLA class	Acetic acid derivs and related substances	47	41	12
	Alpha-adrenoreceptor blocking agents	114	55	18
	Analgesics, antipyretics, anilides	154	60	17
	Analgesics, antipyretics, salicylic acid and derivs	409	266	107
	ACE inhibitors, plain	230	192	56
	Anti-anginal vasodilators, organic nitrates	47	132	56
	Antidiabetic sulfonamides urea derivs	89	79	25
	Anxiolytic benzodiazepine derivs	65	74	19
	Beta blocking agents, plain, non-selective	56	68	31
	Beta blocking agents, plain, selective	114	153	74
	Digitalis glycosides	52	53	21
	Glucocorticoids	18	21	11
	High-ceiling diuretics, sulfonamides, plain	37	40	29
	Hmg coa reductase inhibitors	176	63	31
	Inhal glucocorticoids	39	74	20
	Inhal selective beta2-adrenoceptor agonists	48	80	27
	Low-ceiling diuretics and K-sparing agents	32	44	14
	Propionic acid derivs	154	38	15
	Proton pump inhibitors	36	44	12
	Ca channel blockers, dihydropyridine derivs	172	186	62
	Vitamin k antagonists	42	46	22

Appendix 2

Resolving Inconsistencies in the Coding of Indications in Study 1

In Study 1 the variable 'Indications' was unconstrained leading to many overlapping entries. These were combined as shown in the table below (final codes used are given in bold at the top of each cell).

There were other problems related to this variable: In 37 cases indication was listed as 'frequency', these were replaced with blanks. The terms 'Old', 'Intermittent' and 'occasional' were removed. Finally, in a very small number of cases, there were typos for which it was unsafe to infer the true meaning, e.g. 'prophylaxis gupp infection'. In these cases, the entries were deleted.

acne Adult acne	**angina** acute angina angina prophylaxis angina/headache	**arrhythmia** arrythmia arrythmias	**asthma** allergic asthma allergenic ashthma
allergies allergies environmental allergies seasonal allergies, environmental allergy allergy symptoms allergy, spots and itching enviornmental allergies pollen allergy	**anemia** anemia prevention pre-op anemia prevention pre-op anemia prophylaxis	**anesthesia** anesthesia hernia repair anesthesia urethral stricture anesthesia (root canal) anesthesia aaa repair anesthesia aaa surgery anesthesia adjunct anesthesia for colonoscopy anesthesia for hernia repair anesthesia for hip repair anesthesia for polypectomy anesthesia for polypectomy anesthesia for surgery anesthesia- hernia repair anesthesia pre catheterization anesthesia sphincter placement anesthesia- urethral stricture anesthesia, colonoscopy anesthesia-hernia repair anesthetic anesthetic for inguinal mass anesthetic urethral dilation anesthetize right groin area anthesthetic	**angioplasty** angioplasty for cad angioplasty/stent placement
acid reflux acid indigestion acid stomach acidic stomach acidity antacid antiacid dyspepsia esophageal reflux esophagus reflux gastric hyperacidity gastric hyperactivity gastric hypersecretion gastric reflux gastroesophageal reflux gastroesophageal reflux dis.	**anxiety** nervousness agitation anxiety attacks anxiety before md visit anxiety prophylaxis anxiety, overeating anxiety, sinus pain anxiety/insomnia panic attacks	**arthritis** arthritis/carpal tunnel rheumatoid arthritis osteoarthritis right left knee osteoarthritis left right knee osteoarthritis both knees arthritis knees arthritis & inflammation arthiritis arthitic pain arthritic pain arthritic pain & inflammation arthritic pain/headache arthritis and headache arthritis hand & shoulders arthritis in knees arthritis in neck	**ankle edema** ankle swelling ankle, finger edema

gastroesophageal reflux diseae gastroesphogeal reflux heart burn heartburn heartburn, chest pain high acidity in stomach high stomach acidity hyperacidity hyperacidity of stomach indigestion intermittant heartburn		arthritis knees arthritis of hip arthritis of knees arthritis rheumatoid arthritis shoulders arthritis, back arthrosis athritis djd djd pain	
back pain lower back pain low back pain hip & back pain intermittent back pain occasional back pain	**bronchitis** acute bronchitis	**anticoagulant** anticoagulation anticoagulation/prophylaxis	**cough** dry cough
cold common cold head cold cold symptoms other cold combinations cold and cough head and cold rib pain and cold head cold and chest congestion headcold sinus cold virus cold head cold and chest congestion flu/cold cold systoms cold virus cold, occ. Headaches flu influenza flu (antiviral) flu like symptoms flu prevention flu prophylaxis head/chest nasal congestion	**constipation** intermittant constipation	**diabetes** diabetes type ii diabetes mellitus diabetes mellitus adult onset diabetes diabetes millitis insulin dependent diabetes borderline diabetes diabetes type 11 dm	**diverticulitis** diverticulosis divertilitis
dry skin dry facial skin dry hands dry skin rash	**duodenal ulcer** duodenal bleeding ulcer peptic ulcer peptic ulcer disease peptic ulcer symptoms peptic ulcer/gi reflux peptic ulcers	**Epilepsy** anti-seizure epileptic prophylaxis	**emphysema** emphyzema
epididymitis epididymal pain epididymitis epididymitis, prostatitis epididymitis, prostatitis epidiymitis	**fever** elevated temperature	**fibromyalgia** fibromyalgia syndrome	**flatulence** flatulence prophylaxis gas gaseous stomach
fungal infection fungal infection foot fungal infection in groin fungal infection of foot fungal infection of groin fungal infection of toenail	**gastric ulcer** gastric ulcer and anemia gastric ulcer disease gastric ulcers	**glaucoma** gluacoma	**gastroenteritis** acute gastroenteritis

fungal infection on face fungal infection skin fungal infection toenails fungal otitis externa fungal rash fungus fungus in nail beds fungus infection fungus on feet fungus on toenails			
gastric pain epigastric discomfort epigastric disconfort epigastric distress epigastric pain gastric discomfort gastric distention gastric distress gastric irritation gastric problem gastric upset gastric upset prophylaxis	hemorrhoids external hemorrhoid external hemorrhoids hemhorroids hemmorhoids hemorrhiods hemorrhiods hemorrhoid hemorrhoids/scrotal varices	headache heachache head aches	heart disease cardiac disease coronary artery disease heart disease prophylaxis congestive heart failure prophylaxis- cardiac cardiac profalaxis cardiac prophylaxis prophylaxis cardiac problems cad prophylaxis arteriosclerotic cardiovascula arteriosclerotic heart disease ascd ashd atherosclerosis atherosclerotic heart disease heart condition heart disease prophy heart disease protection heart disease. heart disease/afi heart disease/hypertension hypertension/heart disease
headache occasional headache headaches	hernia hernia operation hernia repair hernia repair pain hiatal hernia hiatal hernia & reflux disease hiatal hernia discomfort	hay fever hayfever	high cholesterol hypercholesterolemia cholesterolemia hypercholesteremia elevated cholesterol elevated cholesterol levels elevated cholestrol hypercholaesterolemia hypercholersteolemia hypercholesterlemia hypercholestermia hypercholesterolimia hypercholestolemia hyprhigh cholesterol increased cholesterol
hypothyroidism hypothroidism hypothyroid	hypertension htn hypertensiom high blood pressure blood pressure elevated hypertension elevated bp hbp hpertension hypertension/sick sinus synd hypertensionn hypertention increased hypertension	hip pain hip pain , bone spurs	insomnia difficulty sleeping insominia insomnia and stress insomnia prophylaxis
ibs Irritable bowel syndrome	incontinence urinary incontinence	impotence male sexual dysfunction	infection prophylaxis infection prevention

Irritable bowel	stress incontinence incontinence and nocturia incontinence of stool incontinence surgery incontinence, mild incontinency	erectile dysfunction sexual dysfunction impotency difficulty with erection erctile dysfunction erectile dysfunction inpotence	infection prophlaxis
menieres syndrome menieres disease	**nasal congestion** sinus congestion congestion head nasal congestion	**osteoarthritis** osteoarthritis left shoulder osteoarthritis of right knee	**parkinsons** parkinson parkinsons disease
penis implant pain pain from penis prosthesis pain penis implant elective pain from penis prosthesis placement of penis prothesis	**prophylaxis** prophylactic prophylatic prophylactically prophylax preventative phrophylactically	**peri-operative** perioperative perioperative anaesthetic perioperative for back surgery perioperative infection preven peri-operative med perioperative medication peri-operative medication	**polyps** polyp polyp removal sinuses polypectomy polyps removed
post-op pain post op pain post operative pain surgical pain incisional pain post-operative medication post operative medication post op prophylaxis postop prophylaxis	**sinus infection** sinusitis episodic sinusitis	**smoking cessation** aid smoking cessation help stop smoking help to quit smoking	**supplement** general health nutritional supplement dietary supplement supplemental diet supplement dietary supplelment health health maintainance health prophylaxis health supplement herbal supplement nutrition nutrition supplement nutritional nutritional prophylaxis nutritional supplememt nutritional supplemental nutritional supplements nutritional suppliment nutritionl supplement overall well being
tooth pain abcess tooth abcessed tooth abscess tooth abscessed teeth abscessed tooth decayed teeth prosthetic tooth inflamation tooth abscess tooth abscess pain tooth absess tooth ache tooth caries tooth cavity toothache ascessed teeth	**urticaria** episode of urticaria	**urethrectomy** prophylaxisn urethrectomu	

About the Authors

Felix Beacher | felix@coolclinical.com

Felix Beacher is the founder of Cool Clinical. Felix has a PhD in neuroscience and has worked in drug development and various therapy areas including neurodegeneration and cancer.

Leonardo Ancora | leo@coolclinical.com

Leonardo Ancora is a machine learning expert with experience in developing AI for the financial sector.

Legal disclaimer

This publication has been written in general terms and is not intended to be relied on to cover specific situations. Any application of the information given in this publication will depend upon the particular circumstances involved. As such, we recommend that professional advice is sought before acting or refraining from acting on any of the contents of this publication. This publication and the information contained herein is provided “as is”. Cool Clinical makes no express or implied warranties that this publication is error-free or meet any particular criterion of performance or quality.

Cool Clinical accepts no duty of care or liability for any loss to any person acting or refraining from action as a result of any material in this publication.

