PIONEER

The European Network of Excellence for Big Data in Prostate Cancer

Studyathon

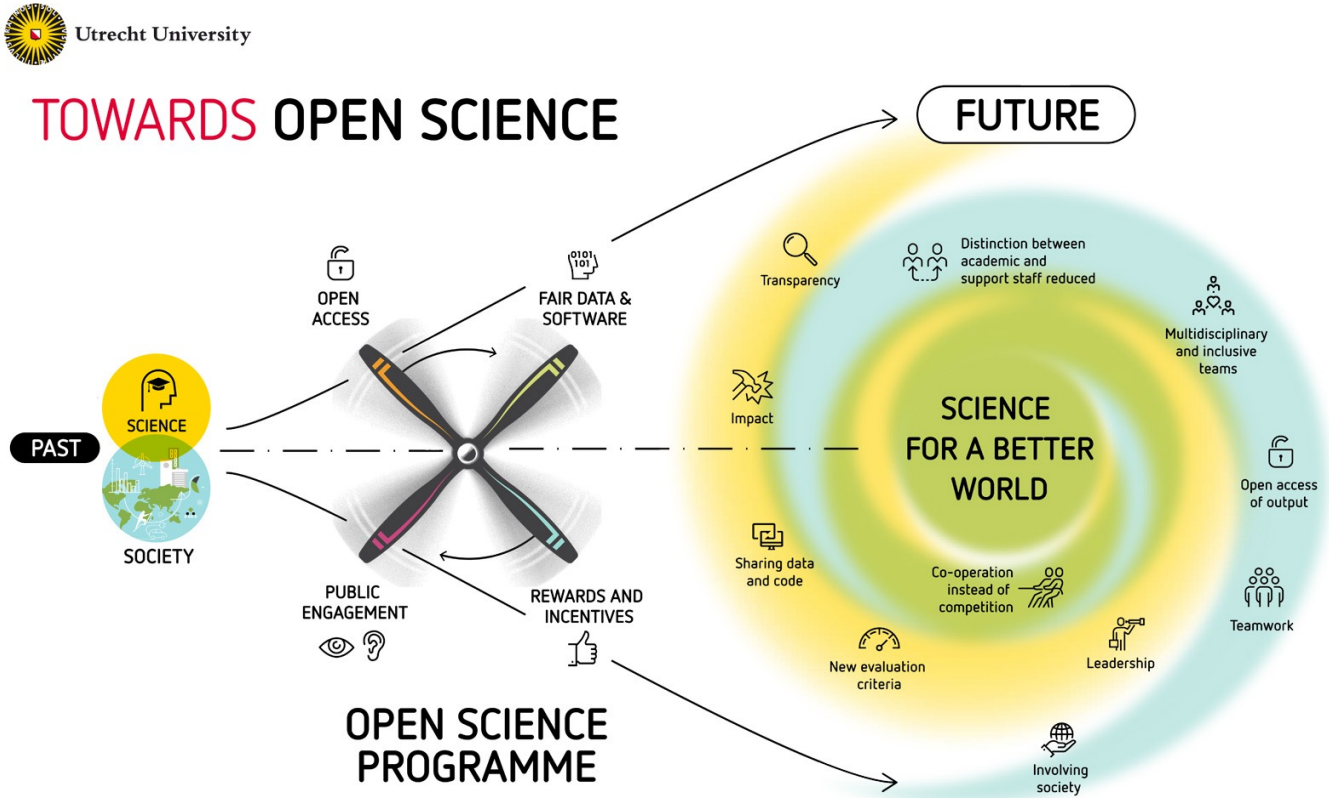
8-12 March 2021

Co-organized by:

EHDEN EUROPEAN HEALTH DATA & EVIDENCE NETWORK

OHDSI OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

What are we doing here?



Big Data for Better Outcomes

PIONEER

BD4BO

imi innovative medicines initiative

THEMES/ENABLERS:

- Design sets of standard outcomes and demonstrate value
- Increase access to high quality outcomes data
- Use data to improve value of HC delivery
- Increase patient engagement through digital solutions

DISEASE-SPECIFIC PROJECTS:

- ROADMAP: Alzheimer's disease
- HARMONY: Haematologic malignancies
- BigData@Heart: Cardiovascular diseases
- **PIONEER: Prostate cancer**

CO-ORDINATING PROJECTS:

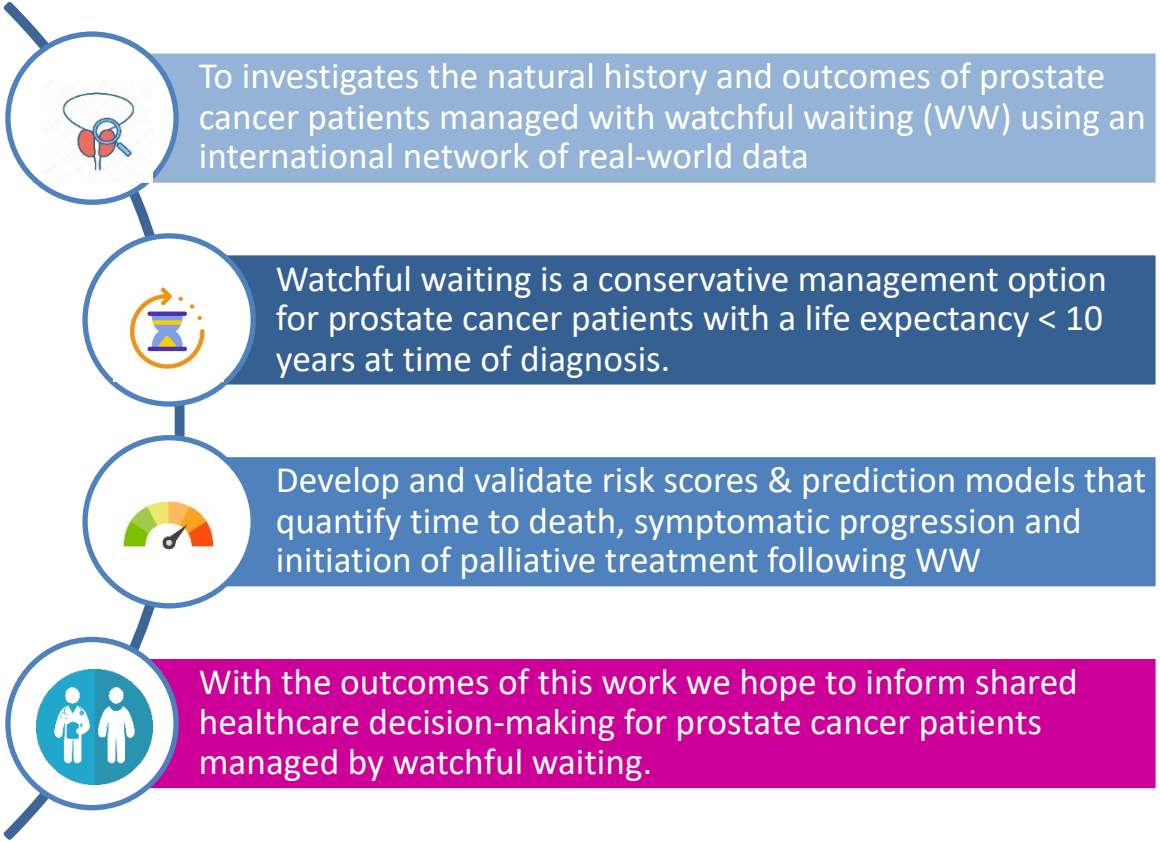
- DO->IT: Coordination & support actions

OVERARCHING:

- European Health Data Network (EHDEN)

PIONEER

Studyathon Objectives



Organising committee

Management team

- Ariel Achtman, Movember
- Kees van Bochove, The Hyve
- Bertrand De Meulder, EISBM
- Christian Reich, IQVIA
- Robert Snijder, Astellas
- Carl Steinbeisser, collaborate.eu/Bayer

Sub-team Leads	
Literature review	Katharina Beyer, King's College London
Clinical characterisation	Giorgio Gandaglia, Vita-Salute San Raffaele, Milan, Italy
Phenotyping	Asieh Golozar, Regeneron Pharmaceuticals, Johns Hopkins Bloomberg School of Public Health
Prediction	Ronald Herrera, Bayer
Data sources	Susan Evans Axelsson, Lund University, Sweden

Supporting Leadership	
PIONEER	James N'Dow, EAU, Alex Asiimwe, Bayer
EHDEN	Nigel Hughes, J&J, Peter Rijnbeek, ErasmusMC
OHDSI	Patrick Ryan, J&J

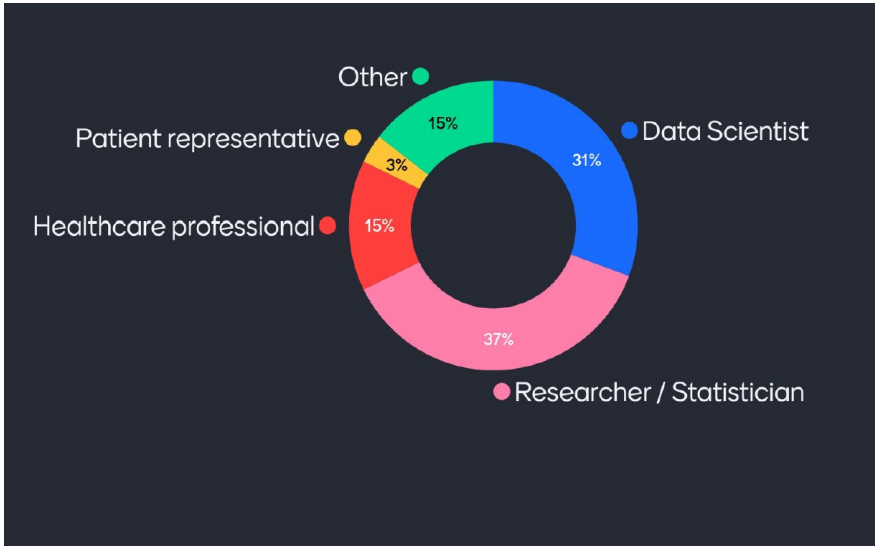
PIONEER

Sub-teams

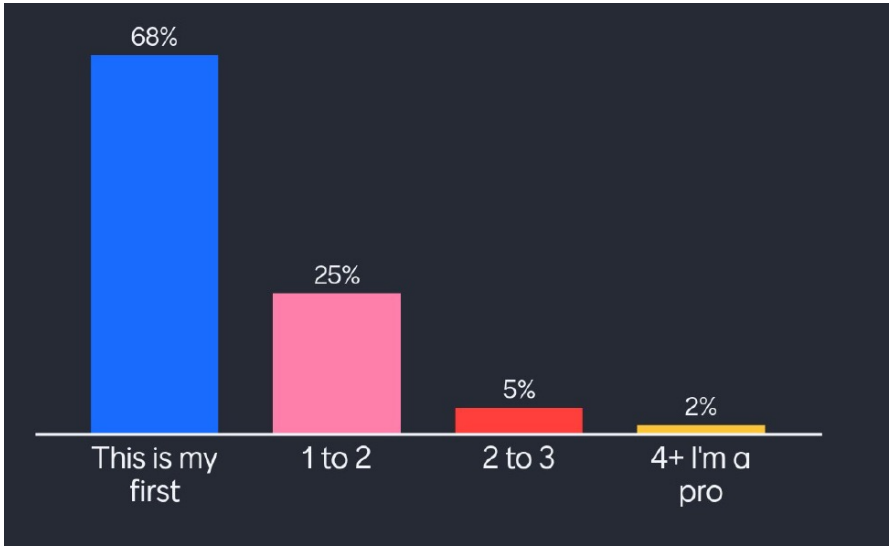
Sub-teams	Objectives
Clinical characterisation	Describe the demographic and clinical characteristics of patients with prostate cancer under watchful waiting (WW) & estimated clinical outcomes of these patients including those who initiated treatment.
Phenotyping	Define the study phenotypes clearly, unambiguously and accurately to generate meaningfully evidence considering differences/nuances of the databases
Prediction	Develop a prediction model, in the context of WW, that predicts an outcome (symptomatic progression, death, death without symptoms) at a specific moment in time (6, 12, 24 months) based on a combination of patient characteristics.
Data sources & study execution	Identify & recruit appropriate databases to the study; developed the code to run analyses for clinical characterisation and to compile results in an easy-to-install R package

Group dynamics

What is your background?



How many studyathons have you attended?

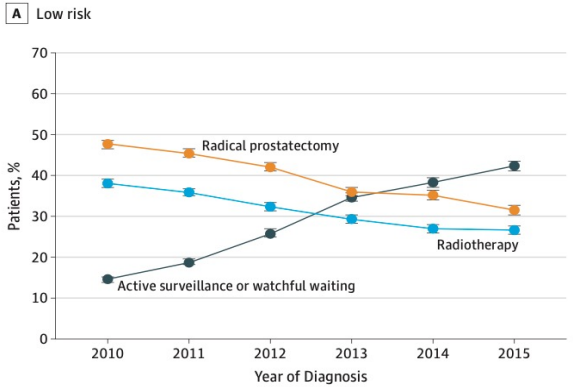


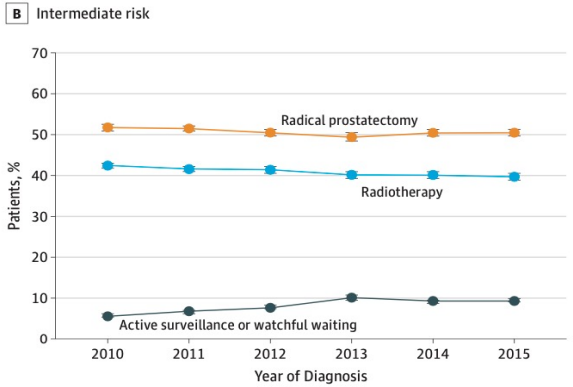
What countries were represented

60 responses

Prostate Cancer Treatment: Options

- Curative-intent treatments:
  - Radical Prostatectomy
  - Radiotherapy
- Conservative management:
  - Active Surveillance
  - Watchful Waiting
- Palliative treatments





Watchful Waiting – The Clinical Perspective

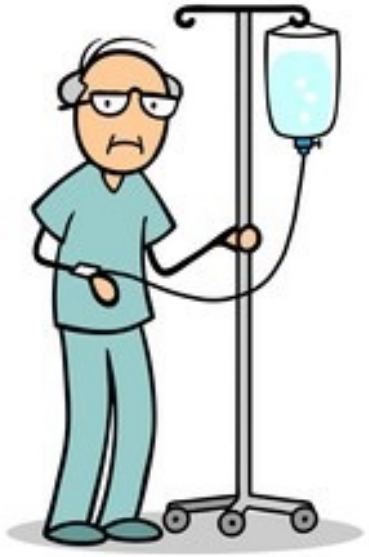
	Active surveillance	Watchful waiting
Treatment intent	Curative	Palliative
Follow-up	Predefined schedule	Patient-specific
Assessment/markers used	DRE, PSA, re-biopsy, mpMRI	Not predefined
Life expectancy	> 10 years	< 10 years
Aim	Minimise treatment-related toxicity without compromising survival	Minimise treatment-related toxicity
Comments	Mainly low-risk patients	Can apply to patients with all stages

Watchful waiting refers to conservative management for patients deemed unsuitable for curative treatment right from the outset, and patients are 'watched' for the development of local or systemic progression with (imminent) disease-related complaints, at which stage they are then treated palliatively according to their symptoms, in order to maintain QoL

EAU Guidelines on PCa, 2021

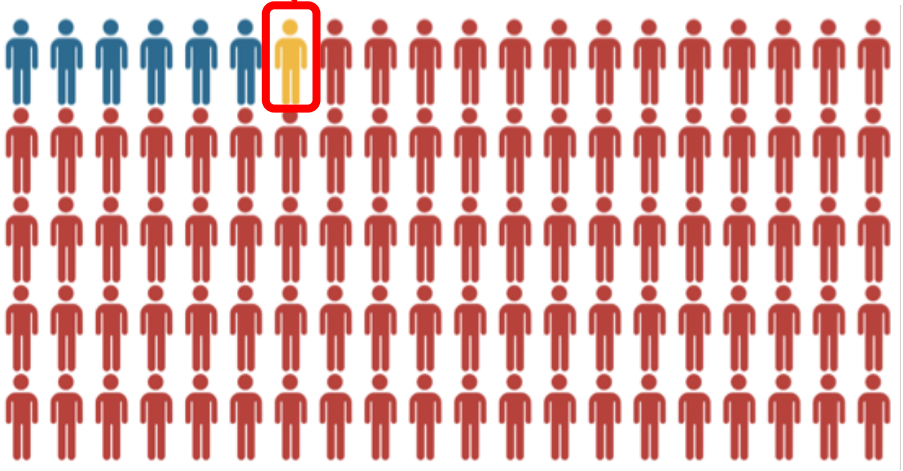
Watchful Waiting – The Clinical Perspective

Patient profile
Male 82 years old
Smoker
High blood pressure & cholesterol
Type 2 DM
History of CAD & CVD
PSA 10 ng/mL
DRE: cT2b
Prostate biopsy: Gleason score 4+3
CT & Bone Scan: negative

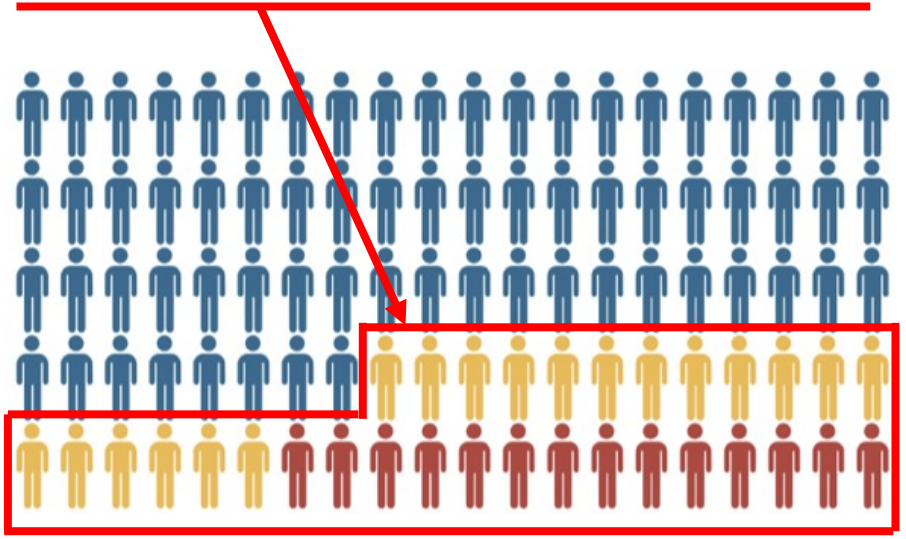


Watchful Waiting – Avoiding Overtreatment

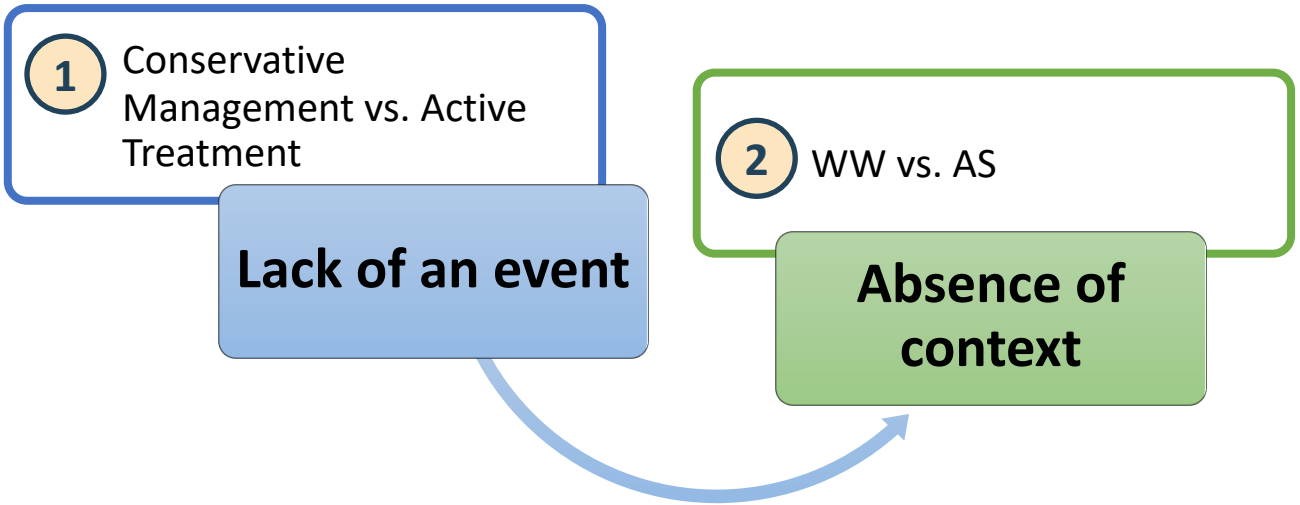
Untreated: **One** prostate cancer death



Radical prostatectomy or radiotherapy: **32** with varying degree of urinary incontinence

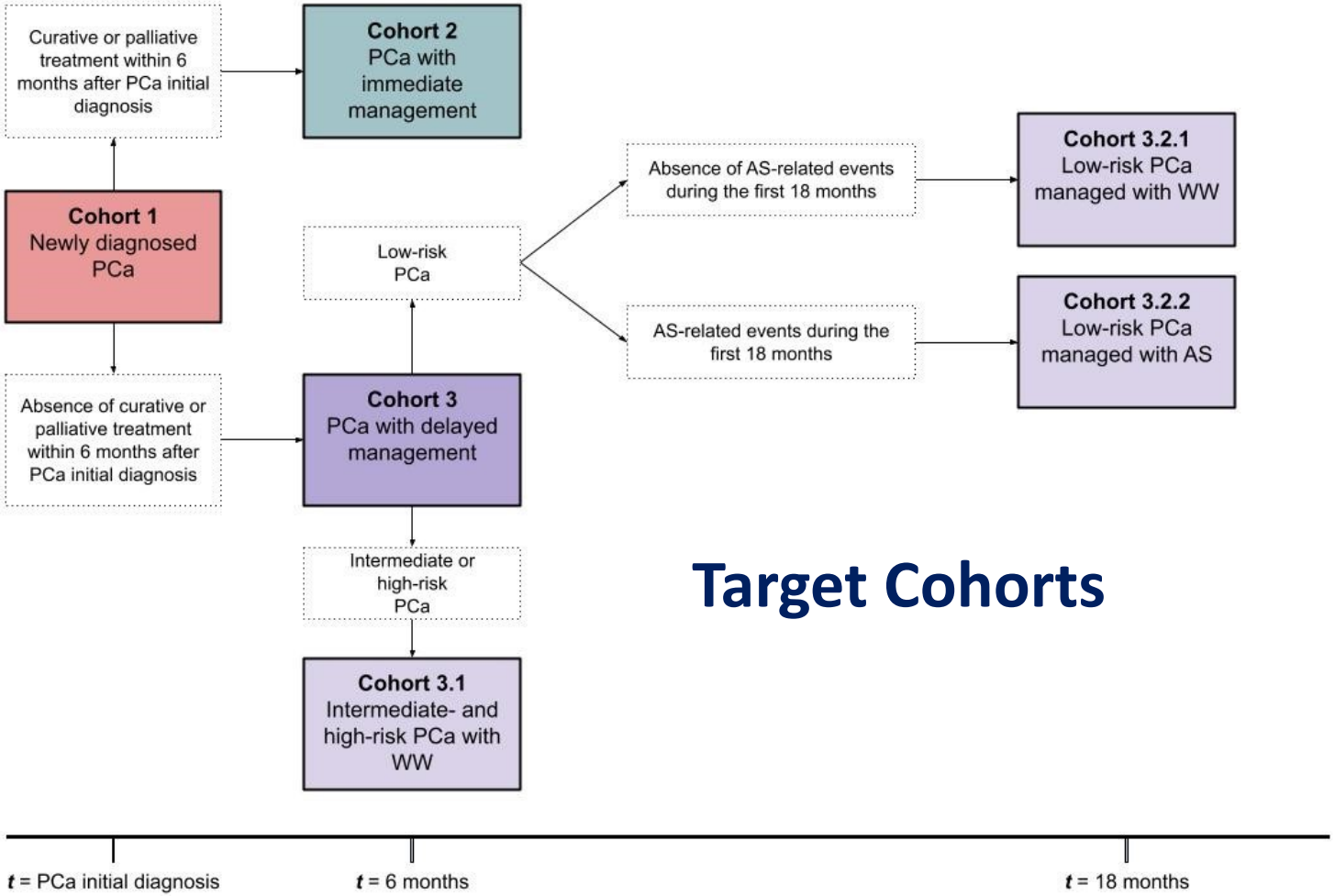


Challenges in Defining WW



Study-a-thon Approach

- arbitrary 6-months time-frame post PCa diagnosis to distinguish active treatment from conservative management
- No attempt to distinguish AS from WW at the time of diagnosis



Target Cohorts

PCa = prostate cancer; WW = watchful waiting; AS = active surveillance; t = index date

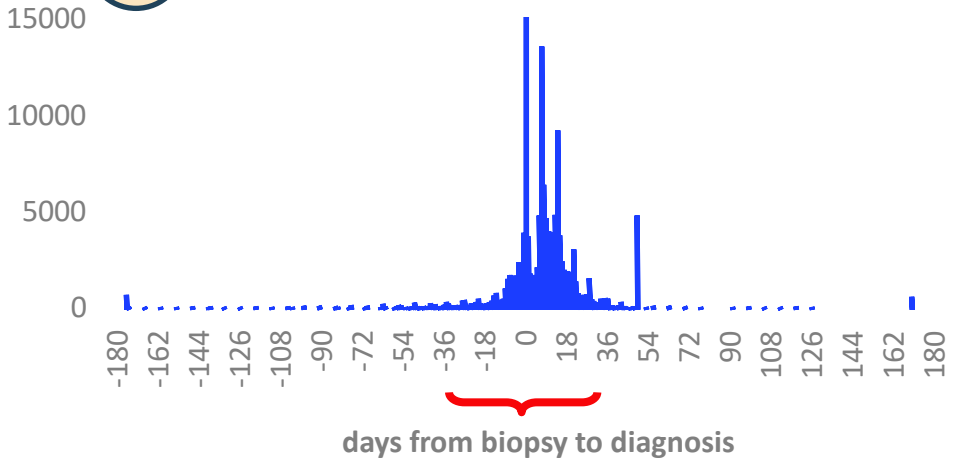
Defining Newly Diagnosed PCa

1 Literature

Validated definition in the literature:

- Combination of PCa diagnosis **AND** prostate biopsy
- Different date of biopsy
- Sensitivity and PPV >90%

2 Data



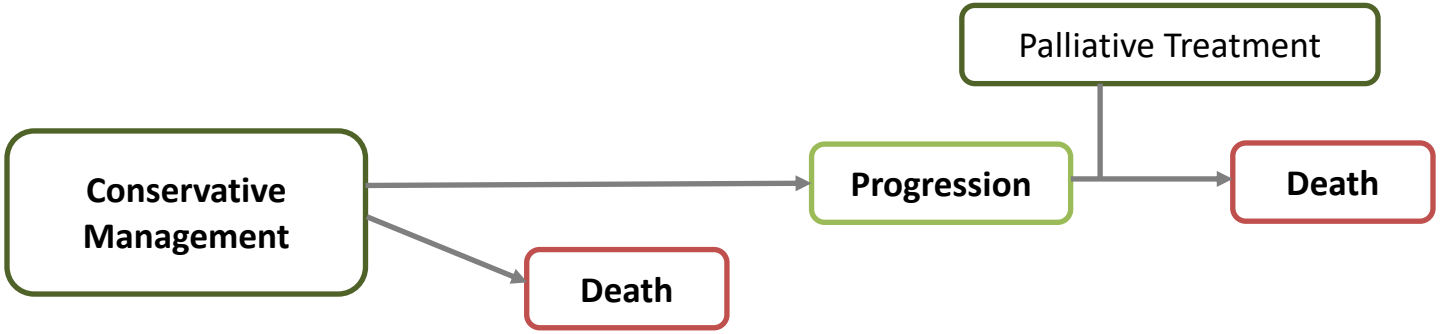
3 PIONEER Definition for newly diagnosed Pca

PCa diagnosis + Prostate biopsy +/- 30 days of Dx or PSA>50 ng/mL + No prior history of PCa or related conditions in the year before + No prior treatment with ADT or other hormonal therapies

Outcomes Cohorts

Patients on conservative management experience two main outcomes:

1. Death
2. Progression after which they receive treatment



Characterization Study: describe patients' characteristics and their outcomes

Prediction Study: identify patients likely to experience death before progression

Challenges in Defining Outcomes

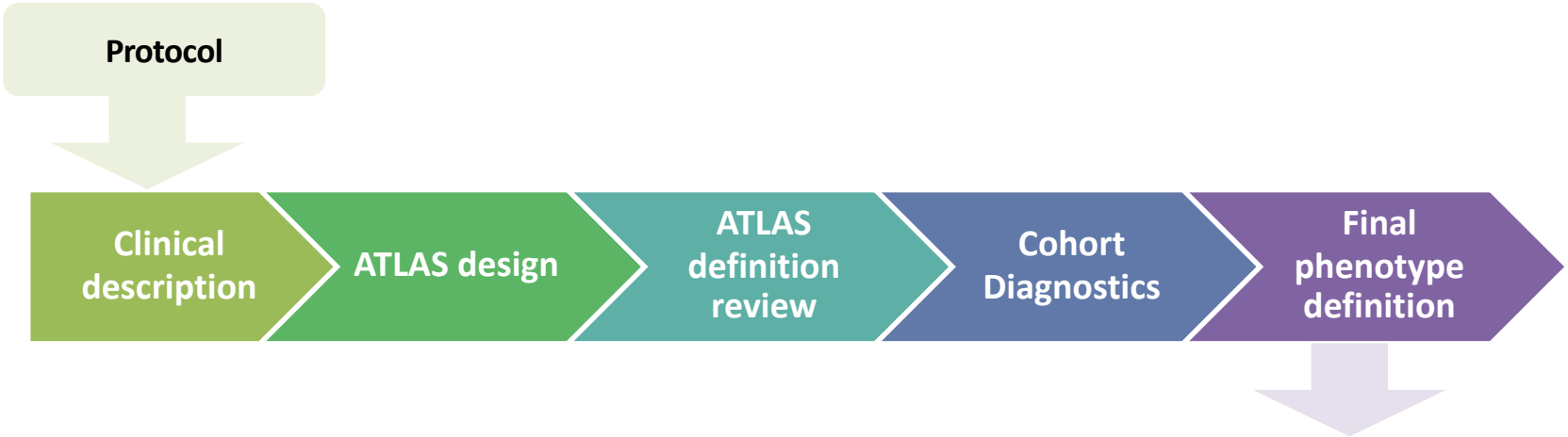
Limitations

- Data Capture
  - Missing information on tumor attributes (i.e., progression)
  - Unavailability/under-reporting of death
  - No/unreliable cause of death
- Competing risk: Not all patients experience both outcomes
- Non comparability of two probability prediction and probabilities over time

Solutions

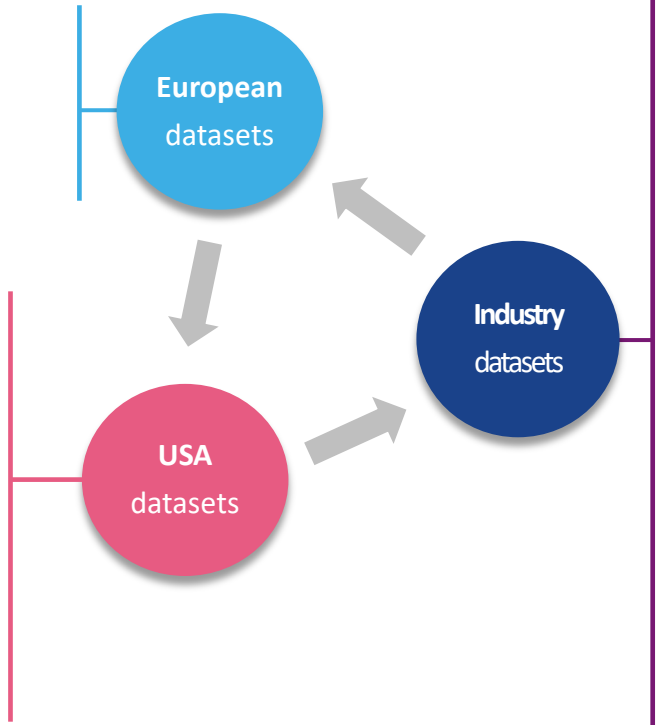
- Use symptoms to define symptomatic progression
- Treatment initiation as a surrogate for progression
- Death before progression/treatment initiation: death due to other cause

Phenotype-Cohort Development Process



- 62 new cohorts were defined: 20 target, 7 outcomes and 35 strata
  - All cohorts are loaded in PIONEER ATLAS (https://pioneer-atlas.thehyve.net/#/home
- **Cohort Diagnostics** results from ten different databases were reviewed

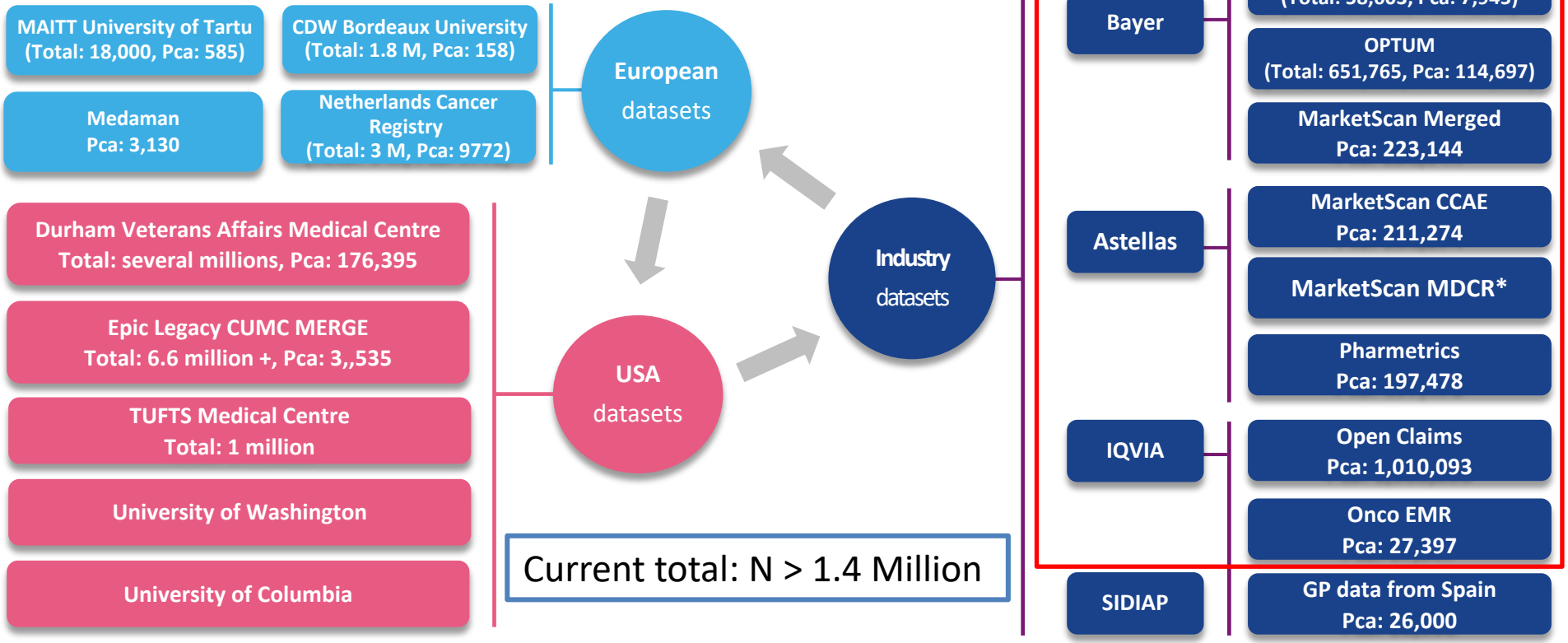
Studyathon Data Overview



- **Bayer**
  - CPRD (Total: 58,603, Pca: 7,945)
  - OPTUM (Total: 651,765, Pca: 114,697)
  - MarketScan Merged Pca: 223,144
- **Astellas**
  - MarketScan CCAE Pca: 211,274
  - MarketScan MDCR*
  - Pharmetrics Pca: 197,478
- **IQVIA**
  - Open Claims Pca: 1,010,093
  - Onco EMR Pca: 27,397

Studyathon Data Overview

PIONEER



PIONEER

Studyathon Goals & Achievements

Cohort diagnostics:
1.4M patients
Shiny app: https://bit.ly/3v6Tnz6

Debugged and functional R package for federated data analytics: bit.ly/3aa1liy

Patient voice included

Prediction models for
- **time to death**
- **symptomatic progression**
- **initiation of palliative treatment**

Communication channels built with EHDEN & OHDSI

Risk scores for risk of death, progression or treatment

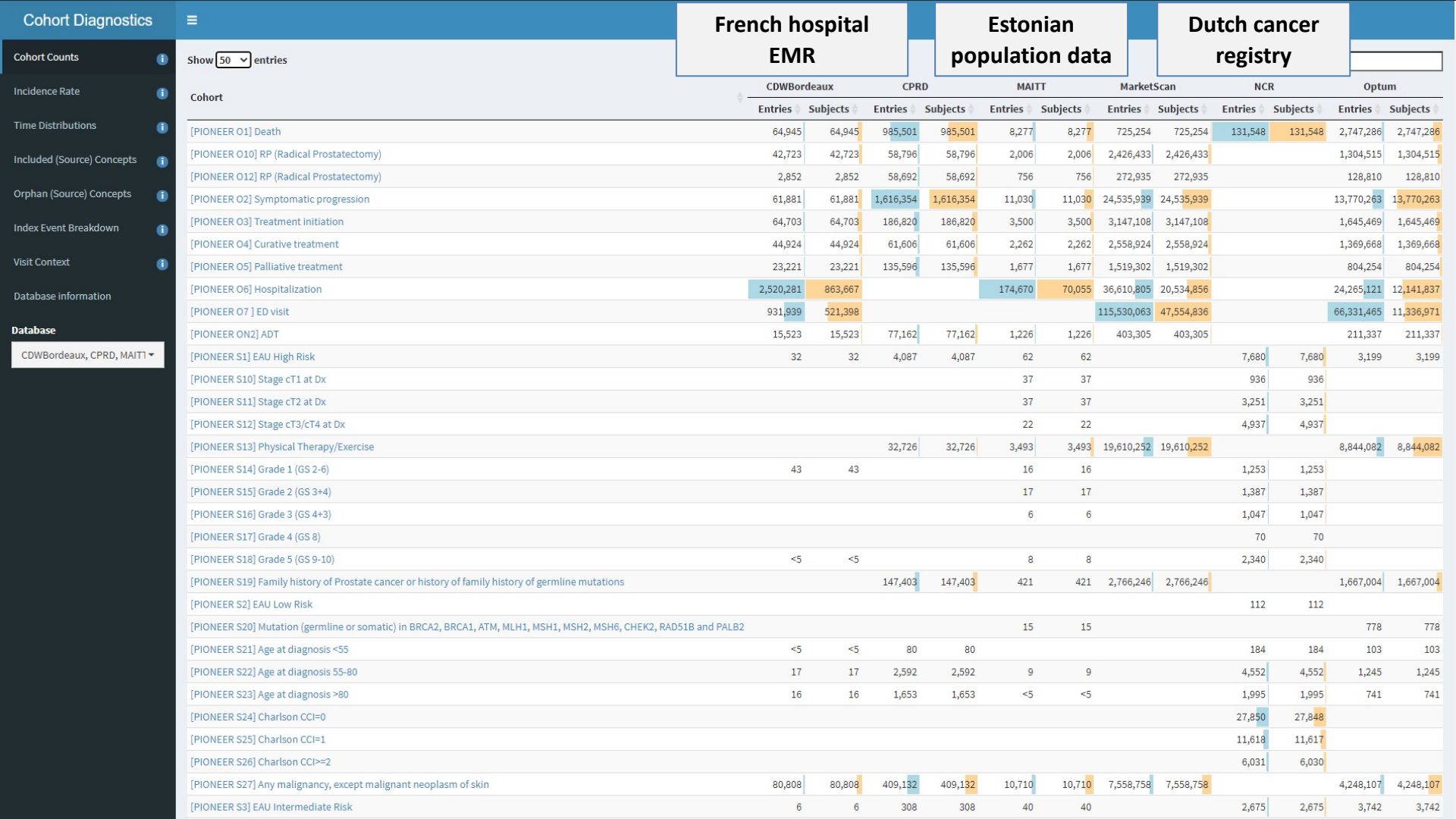
Study protocol available on: bit.ly/3vJI7ZK

Characterisation results now on Shiny App: bit.ly/3dTT8QK

PIONEER

Study Execution

- Simple successive tasks for data holders: (tasks #1 - #4); task #5 is planned for the PLP package
- New functionalities: time-to-event analysis and Kaplan-Meier plots (currently in debugging)
- We reused the OHDSI R packages BUT important to advertise this resource more!
- Feedback loop with phenotypes & clinicians



Cohort	CDWBordeaux		CPRD		MAITT		MarketScan		NCR		Optum	
	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects
[PIONEER O1] Death	64,945	64,945	985,501	985,501	8,277	8,277	725,254	725,254	131,548	131,548	2,747,286	2,747,286
[PIONEER O10] RP (Radical Prostatectomy)	42,723	42,723	58,796	58,796	2,006	2,006	2,426,433	2,426,433			1,304,515	1,304,515
[PIONEER O12] RP (Radical Prostatectomy)	2,852	2,852	58,692	58,692	756	756	272,935	272,935			128,810	128,810
[PIONEER O2] Symptomatic progression	61,881	61,881	1,616,354	1,616,354	11,030	11,030	24,535,939	24,535,939			13,770,263	13,770,263
[PIONEER O3] Treatment initiation	64,703	64,703	186,820	186,820	3,500	3,500	3,147,108	3,147,108			1,645,469	1,645,469
[PIONEER O4] Curative treatment	44,924	44,924	61,606	61,606	2,262	2,262	2,558,924	2,558,924			1,369,668	1,369,668
[PIONEER O5] Palliative treatment	23,221	23,221	135,596	135,596	1,677	1,677	1,519,302	1,519,302			804,254	804,254
[PIONEER O6] Hospitalization	2,520,281	863,667			174,670	70,055	36,610,805	20,534,856			24,265,121	12,141,837
[PIONEER O7] ED visit	931,939	521,398					115,530,063	47,554,836			66,331,465	11,336,971
[PIONEER ON2] ADT	15,523	15,523	77,162	77,162	1,226	1,226	403,305	403,305			211,337	211,337
[PIONEER S1] EAU High Risk	32	32	4,087	4,087	62	62			7,680	7,680	3,199	3,199
[PIONEER S10] Stage cT1 at Dx					37	37			936	936		
[PIONEER S11] Stage cT2 at Dx					37	37			3,251	3,251		
[PIONEER S12] Stage cT3/cT4 at Dx					22	22			4,937	4,937		
[PIONEER S13] Physical Therapy/Exercise			32,726	32,726	3,493	3,493	19,610,252	19,610,252			8,844,082	8,844,082
[PIONEER S14] Grade 1 (GS 2-6)	43	43			16	16			1,253	1,253		
[PIONEER S15] Grade 2 (GS 3+4)					17	17			1,387	1,387		
[PIONEER S16] Grade 3 (GS 4+3)					6	6			1,047	1,047		
[PIONEER S17] Grade 4 (GS 8)									70	70		
[PIONEER S18] Grade 5 (GS 9-10)	<5	<5			8	8			2,340	2,340		
[PIONEER S19] Family history of Prostate cancer or history of family history of germline mutations			147,403	147,403	421	421	2,766,246	2,766,246			1,667,004	1,667,004
[PIONEER S2] EAU Low Risk									112	112		
[PIONEER S20] Mutation (germline or somatic) in BRCA2, BRCA1, ATM, MLH1, MSH1, MSH2, MSH6, CHEK2, RAD51B and PALB2					15	15					778	778
[PIONEER S21] Age at diagnosis <55	<5	<5	80	80					184	184	103	103
[PIONEER S22] Age at diagnosis 55-80	17	17	2,592	2,592	9	9			4,552	4,552	1,245	1,245
[PIONEER S23] Age at diagnosis >80	16	16	1,653	1,653	<5	<5			1,995	1,995	741	741
[PIONEER S24] Charlson CCI=0									27,850	27,848		
[PIONEER S25] Charlson CCI=1									11,618	11,617		
[PIONEER S26] Charlson CCI>=2									6,031	6,030		
[PIONEER S27] Any malignancy, except malignant neoplasm of skin	80,808	80,808	409,132	409,132	10,710	10,710	7,558,758	7,558,758			4,248,107	4,248,107
[PIONEER S3] EAU Intermediate Risk	6	6	308	308	40	40			2,675	2,675	3,742	3,742

About
Cohorts
Cohort Counts
Cohort Characterization
Time To Event
Compare Cohort Char.
Database information
Change Log

Database
OpenClaims

Cohort (Target)
[T1a] Newly diagnosed Pca

Strata (Target)
All

Cohort (Comparator)
[T3a] PCa under conservativ

Strata (Comparator)
All

Domain
All

Time Window
-365d to -1d, -30d to -1d, inc

Target: [T1a] Newly diagnosed Pca (biopsy or PSA GT 50) (n= 1010093)

Comparator: [T3a] PCa under conservative management (biopsy or PSA GT 50) (n= 403150)

Download Data

(•) Table ( ) Plot

Show 25 entries

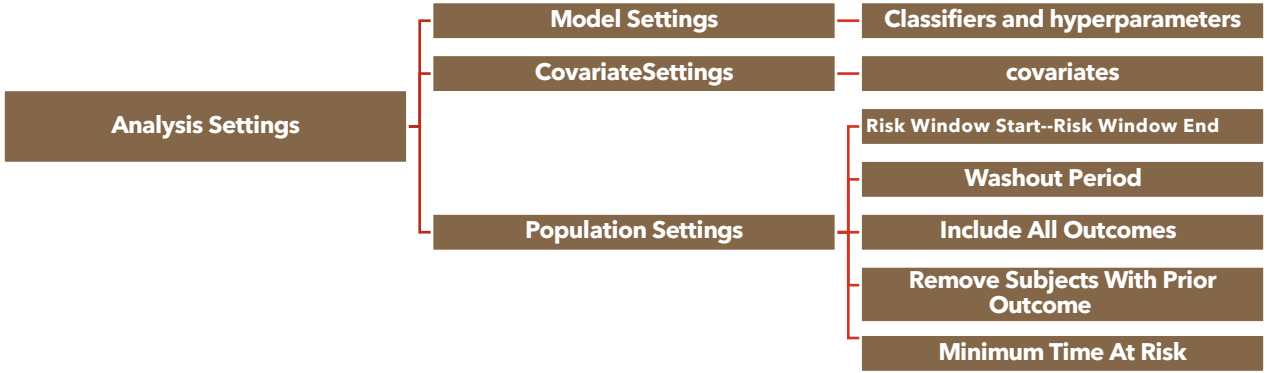
Search:

Covariate name	Mean Target	SD Target	Mean Comparator	SD Comparator	StdDiff
drug_era group during day 0 through 30 days relative to index: endocrine therapy	13.2%	0.34	0.2%	0.04	-0.38
drug_era group during day 0 through 30 days relative to index: antineoplastic and immunomodulating agents	14.9%	0.36	2.1%	0.14	-0.33
drug_era group during day 0 through 30 days relative to index: hormones and related agents	9.4%	0.29	0.1%	0.02	-0.32
drug_era group during day 0 through 30 days relative to index: hormone antagonists and related agents	5.7%	0.23	0.1%	0.03	-0.24
drug_era group during day 0 through 0 days relative to index: endocrine therapy	2.6%	0.16	0.1%	0.04	-0.15
drug_era group during day 0 through 0 days relative to index: hormones and related agents	1.7%	0.13	0.0%	0.02	-0.13
cohort during day 0 through 30 days start the index: 210	2.2%	0.15	0.2%	0.05	-0.13
drug_era group during day 0 through 0 days relative to index: antineoplastic and immunomodulating agents	4.0%	0.20	1.6%	0.13	-0.10
condition_era group during day -30 through -1 days relative to index: prostate specific antigen abnormal	74.0%	0.44	68.2%	0.47	-0.09
condition_era group during day -30 through -1 days relative to index: raised prostate specific antigen	74.0%	0.44	68.2%	0.47	-0.09
condition_era group during day -30 through -1 days relative to index: measurement finding outside reference range	76.2%	0.43	70.6%	0.46	-0.09
condition_era group during day -30 through -1 days relative to index: measurement finding above reference range	75.8%	0.43	70.2%	0.46	-0.09
drug_era group during day 0 through 0 days relative to index: hormone antagonists and related agents	1.0%	0.10	0.1%	0.03	-0.09
condition_era group during day 0 through 30 days relative to index: secondary malignant neoplastic disease	1.9%	0.14	0.6%	0.08	-0.08
condition_era group during day 0 through 30 days relative to index: secondary malignant neoplasm of bone	1.5%	0.12	0.4%	0.06	-0.08
cohort during day 0 through 0 days start the index: 304	1.7%	0.13	0.5%	0.07	-0.08
cohort during day 0 through 30 days start the index: 304	1.7%	0.13	0.5%	0.07	-0.08
condition_era group during day 0 through 30 days relative to index: malignant neoplasm of bone	1.5%	0.12	0.4%	0.06	-0.08
condition_era group during day 0 through 30 days relative to index: malignant neoplasm of skeletal system	1.5%	0.12	0.4%	0.06	-0.08
condition_era group during day 0 through 30 days relative to index: neoplasm of bone	1.5%	0.12	0.4%	0.06	-0.08
condition_era group during day 0 through 30 days relative to index: neoplasm of skeletal system	1.5%	0.12	0.4%	0.07	-0.08
condition_era group during day -365 through -1 days relative to index: prostate specific antigen abnormal	86.5%	0.34	82.7%	0.38	-0.07
condition_era group during day -365 through -1 days relative to index: raised prostate specific antigen	86.5%	0.34	82.7%	0.38	-0.07
condition_era group during day 0 through 30 days relative to index: mass of musculoskeletal structure	1.8%	0.13	0.7%	0.08	-0.07
condition_era group during day 0 through 30 days relative to index: neoplasm of musculoskeletal system	1.6%	0.13	0.6%	0.07	-0.07

Showing 1 to 25 of 60,200 entries

Previous 1 2 3 4 5 … 2408 Next

PLP: analysis settings

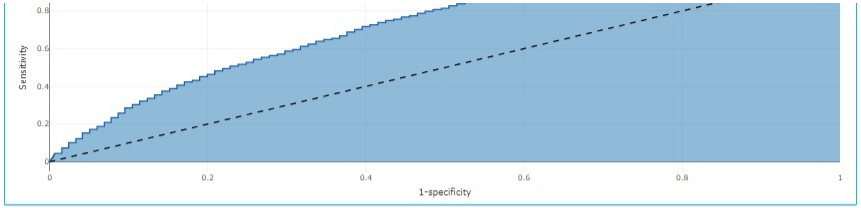


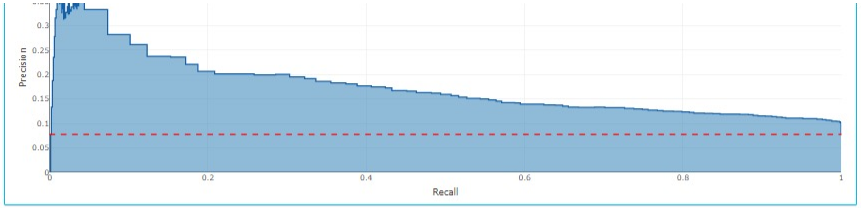
PREDICTION MODEL	Model hyper-parameters	Model setting
Random Forest	{"mtries":[squere root of number of features],"ntrees":[500],"maxDepth":[4,10,17],"varImp":[true]}	Seed
Naïve Bayes	None	None
Multi-layer preceptrons (MLP)	{"size":[4],"alpha":[0.00001]}	Seed
Lasso Logistic Regression	{"variance":0.01}	Seed
KNN (K-Nearest neighbour)	{"k":1000}	Seed
Gradient Boosting Machine	{"ntrees":[10,100],"nthread":20,"maxDepth":[4,6,17],"minRows":[20],"learnRate":[0.01,0.1]}	Seed
Decision Tree	{"maxDepth":[10],"minSamplesSplit":[2],"minSamplesLeaf":[10],"minImpurityDecrease":[1e-7],"classWeight":["None"],"plot":false}	Seed
AdaBoost	{"nEstimators":[50],"learningRate":[1]}	Seed

PLP: prediction of symptoms onset in OPTUM

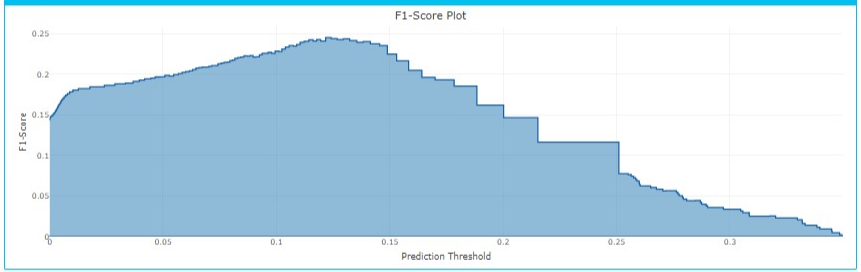
PIONEER

Log
Data Info
Help
Threshold: 0.074216
Result:
Dev: Optum - Val: Optum -T:
Development Optum
Validation Optum
Model Lasso Logistic Regression
T [PIONEER T3] new PCa cons...
O [PIONEER O1] Symptoms...

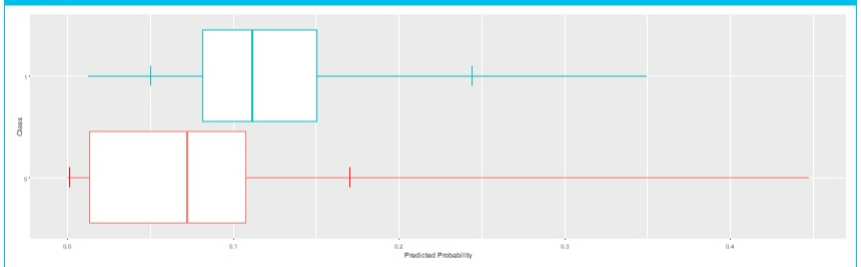




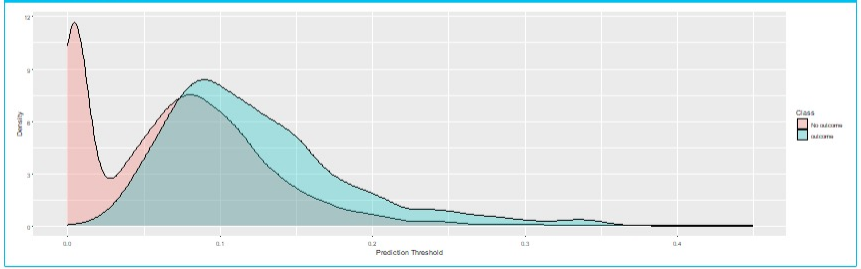
F1 Score Plot



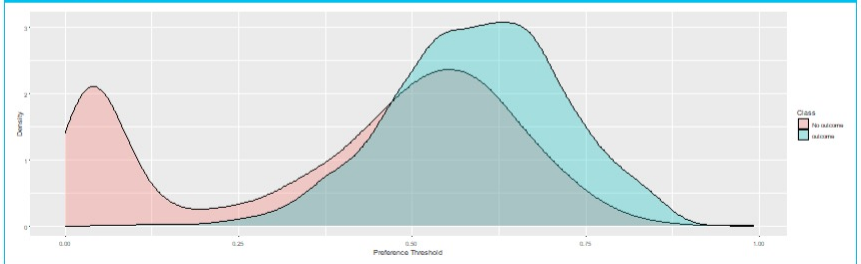
Box Plot



Prediction Score Distribution



Preference Score Distribution



PLP: status & next steps

So far:

- PLP package works on claims data
- Protocol is finalised & soon published
- Clinicians' input reached consensus

To do:

- Include specific covariates in the PLP package
- Send package to data partners for model development and validation (task #5)

Patient representatives

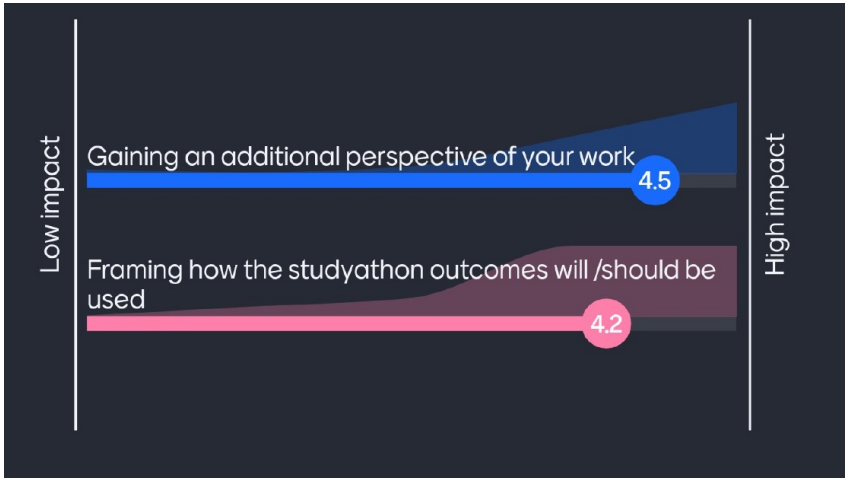
Gary Hooker

Ken Mastris

Robert Greene

Erik Briers

How would you rate the impact of patient participation on the study with regards to:

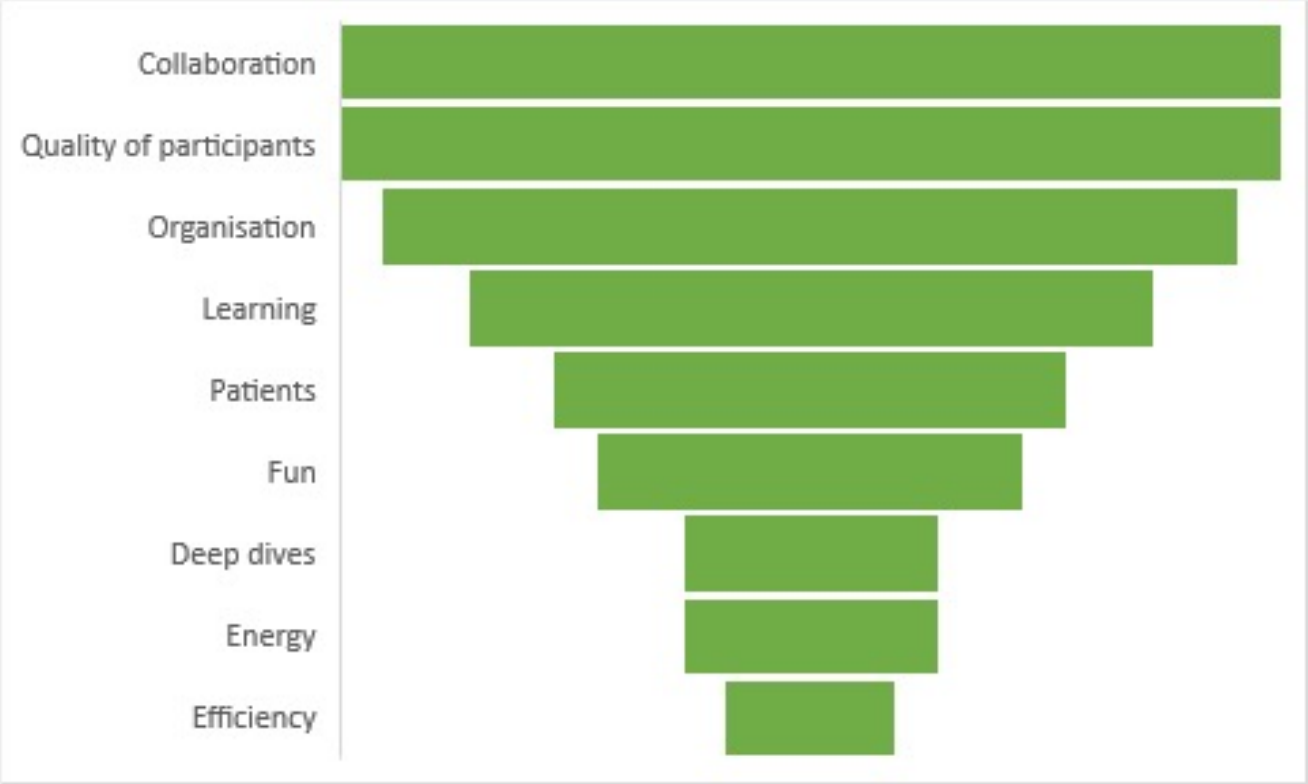


Shared their disease stories

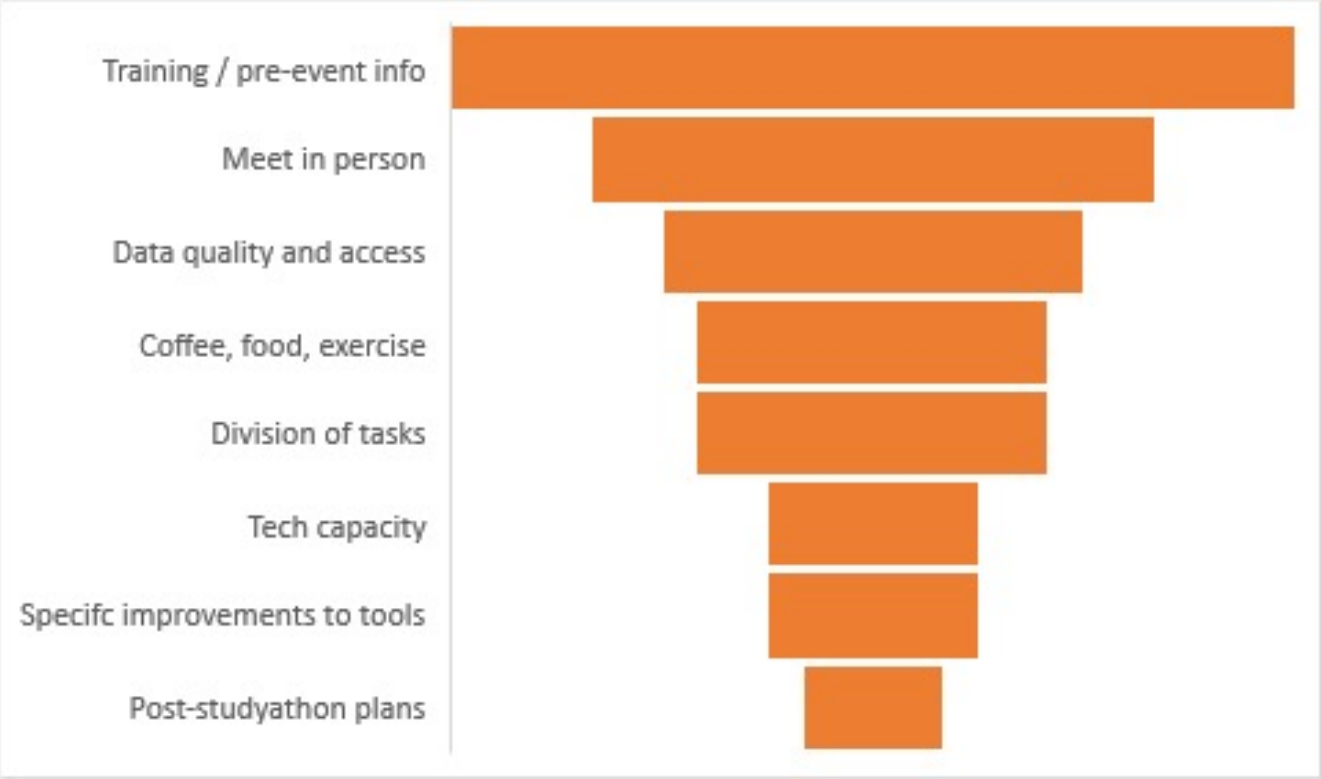
Q & A session

Joined studyathon teams

What participants valued

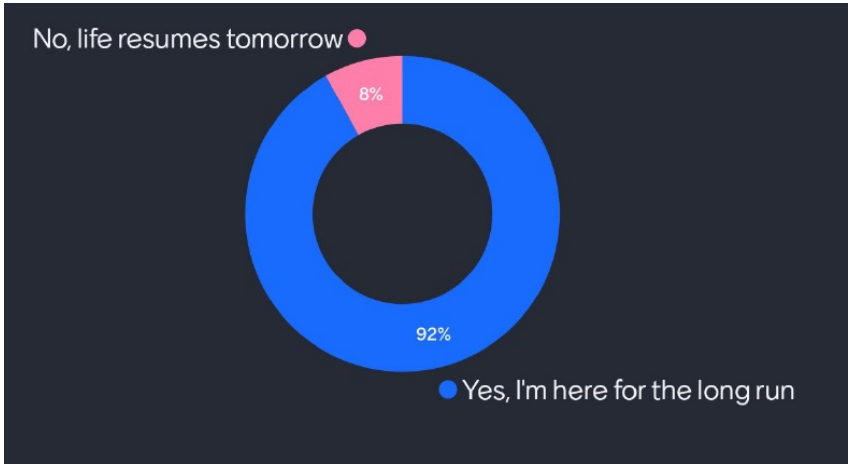


What could be improved



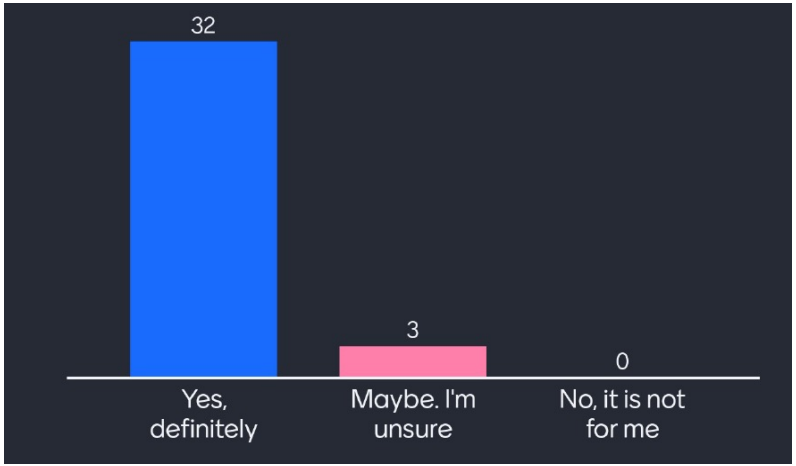
Willingness to engage afterwards

Will you be an active participant in the tasks post-studyathon?



n=37

Would you attend another studyathon?



n=35

Lessons learned

- Cancer data are particularly complex
- Importance of datathon well before studyathon
- Inferring critical information from shadows in the databases is challenging
- Studyathons can be led from outside of OHDSI, but benefit from
    - Guidance
    - Training for new participants
    - Library of re-useable resources
    - Knowledge of existing analytics tools
    - Post-event integration into OHDSI community

Future Outlook

Study-a-thon is still ongoing!!

- Publications
- Estimations
- Prediction

Data owners joined the group and keep on joining

Importance of patient representatives - key for future studyathons

Next steps

- Future studyathons
  - Different questions & data
- Explore new formats & approaches

Collaborative spirit bringing people & skills together – common goal

Thank you!

Together we can ensure each individual patient receives the right treatment for them at the right time.

PIONEER is funded through the IMI2 Joint Undertaking and is listed under grant agreement No. 777492 and is part of the Big Data for Better Outcomes Programme (BD4BO). IMI2 receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA)

Contact us for queries

Team Lead Contacts	
Clinical characterisation	Giorgio Gandaglia (giorgio.gandaglia@gmail.com)
Phenotyping	Asieh Golozar (golozar@ohdsi.org) Shilpa Ratwani (shilpa@ohdsi.org)
Prediction	Ronald Herrera (ronald.herrera@bayer.com)
Data sources/study execution	Susan Evans Axelsson (susan.evans_axelsson@med.lu.se)
General Contacts	
PIONEER	Carl Steinbeisser (carl@collaborate.eu) Emma Jane Smith (e.smith@uroweb.org)
EHDEN	Nigel Hughes (nhughes@its.jnj.com)
OHDSI	Peter Rijnbeek (rijnbeek@ohdsi.org)