Dataset Creation Plan

Project Initiation This Section must be Completed Prior to Project Dataset(s) Creation		
Project Title:	Assessment of the incidence, prevalence and economic impact of prostate cancer using administrative data in Ontario	
Project TRIM number:	2021 0970 231 000	
Research Program:	DAS	
Site:	ICES Central	
Project Objectives:	*Insert Project Objectives as listed in the approved ICES Project PIA*	
	The study aim is to estimate the incidence and prevalence of patients with prostate cancer who fall into on of the four disease states:Localized Non-metastatic, castration resistant (nmCRPC), Metastatic, castration sensitive (mCSPC), Metastatic, castration resistant (mCRPC). Upon classification of the disease states, health resource utilization and economic impact will be determined to demonstrate the economic burden of illness.	
ICES Project PIA Initial Approval Date:	*The ICES Employee or agent who is responsible for creating the Project Dataset(s) is responsible for ensuring there is an approved ICES Project PIA and verifying the date of approval prior to creating the Project Dataset(s)*	
	yyyy-mon-dd *(Private Sector Project: PIA NA)*	
Principal Investigator (PI):	Neerav Monga	
Check the applicable box if the PI is an ICES Student/Trainee	☐ ICES Student ☐ ICES Fellow ☐ ICES Post-Doctoral Trainee ☐ Visiting Scholar	
Responsible ICES Scientist:	*Name the Responsible ICES Scientist if the PI is not a Full Status ICES Scientist*	
	Refik Saskin	
Project Team Member(s) Responsible for Project Dataset Creation and/or Statistical Analysis and date joined (list all):	*All person(s) (ICES Analyst, Appointed Analyst, Analytic Epidemiologist, PI, and/or Student) responsible for creating the Project Dataset(s) and/or statistical analysis on the Research Analytics Environment (RAE) and the date they joined the project must be recorded*	
	Christopher Wallis	2020-Mar-09
Project Team Member(s) who will request RAE folder access (list all):	*List the project team member responsible for dataset creation who will request access for all members requiring RAE project folder access (e.g. analyst, methodologist, student).*	
	Bo Zhang	2020-Oct-08
Other ICES Project Team Members and date joined (list all):	*All other Research Project Team Members (e.g., Research Administrative Assistants, Research Assistants, Project Managers, Epidemiologists) and the date they joined the project must be recorded*	
	Jacob Etches, Jenna Novess, Lisa Ishiguro	2020-Nov-16
Confirmation that DCP is consistent with Project Objectives:	*The following individuals must confirm that the ICES Data provided for in this DCP is relevant (e.g., with respect to cohort, timeframe, and variables) and required to achieve the Project Objectives stated in the ICES Project PIA prior to initial Project Dataset creation: 1) PI; 2) Responsible ICES Scientist if the PI is not a Full Status ICES Scientist, or a second ICES Scientist or the Scientific Program Lead if the PI is creating both the DCP and the Project Dataset[s]; 3) ICES Research and Analysis Staff creating the DCP; and 4) ICES Analytic Staff (ICES Employee or agent responsible for creating the Project Dataset[s]). This may be delegated either verbally or via e-mail.*	

Project Initiation This Section must be Completed Prior to Project Dataset(s) Creation			
	Principal Investigator	☒	2021-Dec-20
	Responsible ICES Scientist or Second ICES Scientist/Lead	☒	2021-Dec-20
	ICES Research and Analysis Staff Creating the DCP	☒	2021-Dec-20
	ICES Analytic Staff	☒	2021-Dec-20
Designated ICES Research and Analysis Staff accountable for Project Documentation:	*The person named (ICES staff) is accountable for ensuring that the approved ICES Project PIA, ICES Project PIA Amendments, and DCP are saved on the T Drive, ensuring ICES Project PIA Amendments are submitted as required, ensuring DCP Amendments are documented, and sharing the final DCP with the PI/Responsible ICES Scientist at project completion*		
	Bo Zhang		
DCP Creation Date and Author:	*Date DCP was finalized prior to Project Dataset(s) creation*	*Name of person who created the DCP*	
	Date	***Name***	
	2020-Sep-29	Christopher Wallis	
	2020-Oct-08	Bo Zhang	

ICES Data This Section must be Completed Prior to Project Dataset(s) Creation	
The ICES Employee or agent who is responsible for creating the Project Dataset(s) must ensure that this list includes only data listed in the ICES Project PIA *Changes to this list after initial ICES Project PIA approval require an ICES Project PIA Amendment*	*Mandatory for all datasets that are available by individual year*
General Use Datasets – Health Services	***Years (where applicable)***
OHIP	2010 – most recent
CIHI DAD	2010 – most recent
CIHI SDS	2010 – most recent
ODB	2010 – most recent
NACRS	2010 – most recent
OLIS	2010 – most recent
General Use Datasets – Care Providers	
See list	
See list	
General Use Datasets – Population	
RPDB	2010 – most recent
See list	

ICES Data This Section must be Completed Prior to Project Dataset(s) Creation	
General Use Datasets – Coding/Geography	
LHIN	2010 – most recent
See list	
General Use Datasets - Facilities	
See list	
General Use Datasets - Other	
See list	
See list	
Controlled Use Datasets	
ALR	Earliest available – most recent
OCR	Earliest available – most recent
NDFP	Earliest available – most recent
Other Datasets (including PSD and PDC data)	

Project Amendments and Reconciliation			
ICES Project PIA Amendment History (add additional rows as needed):	*Privacy approval date*	*Person who submitted amendment*	*Note that any changes to the list of ICES Data or Project Objectives require an ICES Project PIA Amendment*
	Date	***Name***	***Amendment***
	yyyy-mon-dd		
DCP Amendment History (add additional rows as needed):	*Date DCP amended*	*Person who made the DCP amendment*	*Note that any DCP amendments involving changes to the list of ICES Data or Project Objectives require an ICES Project PIA Amendment*
	Date	***Name***	***Amendment***
	2020-Dec-23	Bo Zhang	1. Added a new category based on best_stage_grp='Unk' for patient groups, so that there will be 3 groups: a). Prostate cancer – non-metastatic at diagnosis [best_stage_grp in (I,II,III)] b). Prostate cancer – metastatic at diagnosis (best_stage_grp=IV) c). Prostate cancer – unknown stage (best_stage_grp=Unk) 2. Remove '1d' category from the patient group 'Prostate cancer – non-metastatic at diagnosis'
	2021-Jan-26	Bo Zhang	Extend cohort period back to the inceptip of OCR (i.e. 1964) and estimate PC prevalence from the inception of OCR.

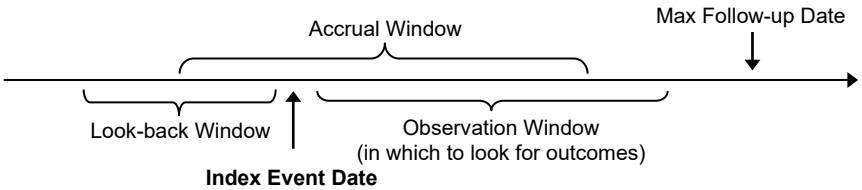
Project Amendments and Reconciliation			
	2021-Mar-25	Bo Zhang	Added risk categories among patients with non-metastatic disease (based on PSA, stage, and grade) and estimate its incidence over time
Date Programs/DCP reconciled	*The person(s) creating the dataset and/or analyzing the data are responsible for ensuring that the final DCP reflects the final program(s) when the project is completed*		
	2021-Dec-20		

Project Cohort			
Study Design	☐ Cohort study	☐ Matched cohort study	☐ Case-control study
	☒ Cross-sectional study	☐ Other (specify):	
Index Event / Inclusion Criteria *(please ensure index event / inclusion criteria are specified with data sources, variables, study period and values or codes)*	Diagnosis of prostate cancer (OCR; curr_topog_cd = ICD-10 O-3 C61.9)		
Estimated Size of Cohort	>10,000		

Exclusions *(in order)* *(common exclusions are listed in grey italics for consideration)*

Step	Description
1	Invalid IKN
2	Invalid birth date(e.g. missing or after index date), Invalid death date (e.g. before index date), Invalid gender (from RPDB) etc.
3	Non-ontario residents at index (substr(prcddablk,1,2) ne '35')
4	Female sex
5	
6	
7	
8	

Project Time Frame Definitions



Accrual Start/End Dates	January 1, 2010 – most recent available (31Dec2019) for incident cases; January 1, 1964 – most recent available (31Dec2019) for prevelant cases
Max Follow-up Date	Most recent available (currently 31Jan2021)

Project Time Frame Definitions	
When does observation window terminate?	End of study period, death, or emigration (dolc from RPDB)
Lookback Window(s) *(please ensure lookback windows are defined with start and end dates and in relation to the index event date)*	2 years prior to index for baseline covariate capture

Variable Definitions (add additional rows as needed)

A few key points to keep in mind:

1. *Please ensure codes, data sources, diagnosis types and lookback periods (if applicable) are provided for all definitions listed below and that codes are provided in Excel format. If borrowing codes from another project, please list all the codes here*
2. *There are maximum number of digits that can be specified using ICES data (ICD 9 CA codes are up to maximum of 4 digits, ICD 10 CA codes are 6 digits, OHIP diagnosis codes are 3 digits)*

Main Exposure or Risk Factor	Time (year) Plan for CROSS-SECTIONAL methodology with incidence and prevalence assessed repeated at choronically defined time intervals (quarterly and annually)
Primary Outcome Definition	Prostate cancer disease states, defined as follows: 1. Prostate cancer – non-metastatic at diagnosis (OCR; best_ stage_grp = 1/2/3) a. No treatment b. Local therapy (RP, brachytherapy, EBRT, or EBRT+brachytherapy; see table) without recurrence c. Local therapy with biochemical recurrence BCR Definition: i. if prior RP, Total PSA (LOINC 19197-3, 2857-1, 35741-8) >0.1 OR ii. if prior RT, Total PSA (LOINC 19197-3, 2857-1, 35741-8) > nadir (=lowest PSA after RT) + 2 ng/mL d. Non-metastatic castrate resistant prostate cancer i. CRPC definition: 1). Total PSA (LOINC 19197-3, 2857-1, 35741-8) level of at least 2 ng/mL above nadir (lowest that must occur after PC diagnosis and immediately precedes meeting the CRPC definition) PSA and also at least 25% higher than nadir level (after nadir), AND 2). A second total PSA test (any LOINC code from 'a') at least 3 weeks following PSA test meeting "a" criteria, AND 3). Total Testosterone level (LOINC 14913-8) < 1.7 nmol/L (drawn within 6 months of total PSA level in criterion "a" or "b"), AND ii. To qualify for nmCRPC, CPRC definition as above plus prior local therapy (RP or RT) and Total PSA <20 ng/mL e. Metastatic prostate cancer, castrate resistant

Variable Definitions (add additional rows as needed)

A few key points to keep in mind:

1. *Please ensure codes, data sources, diagnosis types and lookback periods (if applicable) are provided for all definitions listed below and that codes are provided in Excel format. If borrowing codes from another project, please list all the codes here*
2. *There are maximum number of digits that can be specified using ICES data (ICD 9 CA codes are up to maximum of 4 digits, ICD 10 CA codes are 6 digits, OHIP diagnosis codes are 3 digits)*

	i. Defined as 2+ codes for metastasis chronologically following the diagnosis of prostate cancer (OHIP dxcode in 196, 197 and 198, NACRS; DAD; ICD-10 C77, C78, C79) AND ii. CRPC definition: a). Total PSA (LOINC 19197-3, 2857-1, 35741-8) level of at least 2 ng/mL above nadir (lowest) PSA and also at least 25% higher than nadir level, AND b). A second total PSA test (any LOINC code from 'a') at least 3 weeks following PSA test meeting "a" criteria, AND c). Total Testosterone level (LOINC 14913-8) <1.7 nmol/mL (drawn within 6 months of total PSA level in criterion "a" or "b") 2. Prostate cancer – metastatic at diagnosis (OCR; best_ stage_grp = 4) a. Castrate sensitive: patients initially enter this category and remain here until meeting castration resistant definition (below) b. Castrate-resistant: i. CRPC definition: a). Total PSA (LOINC 19197-3, 2857-1, 35741-8) level of at least 2 ng/mL above nadir (lowest) PSA and also at least 25% higher than nadir level, AND b). A second total PSA test (any LOINC code from 'a') at least 3 weeks following PSA test meeting "a" criteria, AND c). Total Testosterone level (LOINC 14913-8) <1.7 nmol/mL (drawn within 6 months of total PSA level in criterion "a" or "b") 3. Prostate cancer – castration resistant, receiving or received systemic therapy for mCRPC (treatment after mCRPC) a. Receipt of docetaxel, abiraterone, enzalutamide, radium-223, or cabazitaxel b. Denote receipt of 1, 2, 3, or 4 of the above agents 4. Death 5. Prostate cancer – Unknown stage at diagnosis (best_stage_grp='Unk')
If meSecondary Outcome Definition(s)	
Baseline Characteristics	
Other Variables	

QUALITY CONTROL ANALYSIS:
Among patients who meet the castration resistance definition (2 PSA tests and testosterone <1.7 nmol/L), can you assess:
-proportion who meet nmCRPC definition?
-proportion who meet mCRPC definition?
-proportion who meet both definitions?
-proportion who meet neither definition?

NOTE: **Sub-group analysis** among men age >65 years at the time of endpoint assessment:

-for "Metastatic prostate cancer, castrate sensitive", ascertain whether ADT was initiated prior to meeting definition (***expect to be very low***)

-for "Metastatic prostate cancer, castrate resistant", ascertain whether patient is on ADT or after to meeting definition (***expect to be very high***)

Analysis Plan and Dummy Tables (expand/modify as needed) *(please ensure the analysis plan is outlined with dummy tables (can be a separate document) and clear specification of exposures / outcomes / covariates for each model)*	
Descriptive Tables (insert or append dummy tables), e.g.:	
Table 1. Baseline characteristics of study cohort	
Table 2. Baseline chacteristics over time (2 year increments) with test for trend	
Table 3. Incidence and prevalence of prostate cancer and disease states over time	
Table 4. Risk categories among patients with non-metastatic disease (based on PSA, stage, and grade)	
Statistical Model(s)	
Type of model	ARIMA (can discuss other time-series analysis approaches as well)
Primary independent variable	Time (measured quarterly)
Dependent variable	Disease state (overall pstate cancer incidence (potentially remove 2019 data from model and build future projections on data from 2014 onwards; for de novo metastatic disease (stage 4 at diagnosis) and for non-metastatic disease (stage 1-3 at diagnosis; mayconsider modeling the risk categories for risk categories among patients with non-metastatic disease (based on PSA, stage, and grade)
Covariates	
Sensitivity Analyses	
Type of model	
Primary independent variable	
Dependent variable	
Covariates	

Quality Assurance Activities	
RAE Directory of SAS Programs	
RAE Directory of Final Dataset(s)	*The final analytic dataset for each cohort includes all the data required to create the baseline tables and run all the models. It should include all covariates for all models such as patient risk factors, hospital characteristics, physician characteristics, exposure measures (continuous, categorical) and outcomes. It should include covariates that were considered but didn't make the final cut. This would permit an analyst to easily re-run the models in the future.*
RAE README file available:	☐Yes ☐No

Quality Assurance Activities		
Date results of quality assurance tools for final dataset shared with project team (where applicable):		
	%assign	yyyy-mon-dd
	%evolution	yyyy-mon-dd
	%dinexplore	yyyy-mon-dd
	%track / %exclude	yyyy-mon-dd
	%codebook	yyyy-mon-dd
Additional comments:		

Definitions of local treatment modality

	OHIP billing codes	CCI codes
Radical prostatectomy	S651	1.QT.91 (CCP: 724)
External beam radiotherapy	X310, X311, X312, X313, A343, A340, A341, K013	1.SQ.27.JA, 1.QT.27.JA, 1.QT.27.JA-DA (CCP: 0631), 1.QT.27.JA-DB (CCP: 0632), 1.QT.27.JA-DC(CCP: 0621), 1.QT.27.JA-DE (CCP: 0622), 1.QT.27.JA-DG (CCP: 0624), 1.QT.27.JX
Brachytherapy	S640, X323, X324, X325, (X313 and J138 same day)	1.QT.26.BA-EB (CCP: 0634), 1.QT.26.BA-EC (CCP: 0634), 1.QT.26.HA (CCP: 0634), 1.QT.26.HA-EB (CCP: 0634), 1.QT.26.HA-EC (CCP: 0634), 1.QT.26.LA (CCP: 0634), 1.QT.26.LA-EB (CCP: 0634), 1.QT.26.LA-EC (CCP: 0634), 1.QT.53.HA-EM, 1.QT.53.LA-EM

Definition of androgen deprivation therapy

Generic medication name	Drug Identification Number (Trade name and dose)
LHRH agonist/antagonist	
Buserelin acetate	01989677 (Suprefact 1mg/mL), 02225166 (Suprefact 1mg), 02225158 (Suprafact 1mg), 02228955 (Suprefact depot 2mo 6.3mg), 02240749 (Suprefact depot 3mo 9.45mg)
Leuprolide acetate	00727695(Lupron 5mg), 00884502 (Lupron depot 3.75mg), 00836273 (Lupron depot 7.5mg), 02239834 (Lupron depot 11.25mg), 02230248 (Lupron depot 22.5mg), 02239833 (Lupron depot 30mg), 02248239 (Eligard 7.5mg), 02248240 (Eligard 22.5mg), 02248999 (Eligard 30mg), 02268892 (Eligard 45mg)
Goserelin acetate	00857599 (Zoladex 3.6mg), 02049325 (Zoladex 3.6mg), 02225905 (Zoladex 10.8mg)
Triptorelin pamoate	09857199 (Trelstar 3.75mg/mL), 02240000 (Trelstar 3.75mg), 02243856 (Trelstar 11.25mg), 09857200 (Trelstar LA 11.25mg/mL), 02412322 (Trelstar 22.5mg)
Degarelix acetate	02337029 (Firmagon 80mg), 02337037 (Firmagon 120mg)
Anti-androgen	
Bicalutamide	02400731, 02325985, 02274337, 02374412, 02296063, 02360993, 02325233, 02382423, 02360098, 02184478, 02281139, 02357216, 02407531, 02302403, 02358557, 02336545, 02281163, 02275589, 02311038, 02371324, 02277700, 02276089, 02270226, 02428709
Flutamide	02230524, 00637726, 02059673, 02230875, 02238560, 02239045, 02239388, 02239082, 02233019, 02230104, 02230089
Nilutamide	02221861, 02221888, 00863904, 01989650, 00863890, 01989642
Surgical castration	
Bilateral orchiectomy	OHIP – S589 CCI – 1.QM.89.^^

2 Codes for Metastatic prostate cancer, castrate resistant

Data from OHIP

DXCODE (length of 3)	Category Name	Sub-category	Description
196	NEOPLASMS	Malignant Neoplasms	Secondary neoplasm of lymph nodes
197	NEOPLASMS	Malignant Neoplasms	Secondary neoplasm of respiratory and digestive systems
198	NEOPLASMS	Malignant Neoplasms	Metastatic or secondary malignant neoplasm, carcinomatosis

Data from DADSDS or NACRS

ICD10ca	ICD9
C770	1960
C771	1961
C772	1962
C773	1963
C774	1965
C775	1966
C778	1968
C779	1969
C780	1970
C781	1971
C782	1972
C783	1973
C784	1974
C785	1975
C786	1976
C787	1977
C788	1978
C790	1980
C791	1981
C792	1982
C793	1983
C794	1984
C795	1985
C796	1986
C797	1987
C7980	19881
C7988	19889
C799	19889

Mock Table 3. Incidence and prevalence of prostate cancer and disease states over time

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
PREVALENCE										
Overall prostate cancer (n)										
Overall prostate cancer (per capita)										
Non-metastatic PCa (n)										
-High risk localized or locally advanced (stage 3 or Gleason>=7 or PSA>=20 ng/mL)										
-no Rx										
-local Rx without recurrence										
-local Rx with BCR										
-nmCRPC										
Metastatic PCa (n)										
mCRPC on Rx (n)										
INCIDENCE										
Overall prostate cancer (n)										
Overall prostate cancer (per capita)										
Non-metastatic PCa (n)										
-no Rx										
-local Rx without recurrence										
-local Rx with BCR										
-nmCRPC										
Metastatic PCa (n)										
mCRPC on Rx (n)										
Death (n)										

Note: high risk group in nmPC patients, using stage and Gleason scores at PC diagnosis and PSA values in 3 months prior to and 2 months following PC diagnosis.

NCCN Guidelines

TABLE. Risk Stratification and Staging Workup of Prostate Cancer[1]

Risk Group	Clinical/Pathologic Features	Imaging	Molecular Testing of Tumor	Genetic Testing of Tumor
Very low	All of the following: • T1c • Gleason score ≤6/grade group 1 • PSA <10ng/mL • <3 prostate biopsy fragments/ cores positive, ≤50% cancer in each fragment/core • PSA density <0.15 ng/mL/g	Not indicated	Not indicated	Consider if there's a strong family history
Low	All of the following: • T1-T2a • Gleason score ≤6/grade group 1 • PSA <10ng/mL	Not indicated	Consider if life expectancy is ≥10 years	Consider if there's a strong family history
Intermediate-favorable	Any of the following: • T2b-T2c • Gleason score 3+4=7/grade group 2 • PSA 10-20 ng/mL PLUS percentage of positive biopsy cores <50%	• Bone imaging: not recommended for staging • Pelvic ± abdominal imaging: recommended if nomogram predicts >10% probability of pelvic lymph node involvement	Consider if life expectancy is ≥10 years	Consider if there's a strong family history
Intermediate-unfavorable	Any of the following: • T2b-T2c • Gleason score 3+4=7/grade group 2 or Gleason score 4+3=7/grade group 3 • PSA 10-20 ng/mL	• Bone imaging: recommended if T2 and PSA >10 ng/mL • Pelvic ± abdominal imaging: recommended if nomogram predicts >10% probability of pelvic lymph node involvement	Not routinely recommended	Consider if there's a strong family history
High	Any of the following: • T3a • Gleason score 8/grade group 4 or Gleason score 4+5=9/grade group 5 • PSA >20 ng/mL	• Bone imaging: recommended • Pelvic ± abdominal imaging: recommended if nomogram predicts >10% probability of pelvic lymph node involvement	Not routinely recommended	Consider
Very high	Any of the following: • T3b-T4 • Primary Gleason pattern 5 • >4 cores with Gleason core 8-10/ grade group 4 or 5	• Bone imaging: recommended • Pelvic ± abdominal imaging: recommended if nomogram predicts >10% probability of pelvic lymph node involvement	Not routinely recommended	Consider
Regional	Any T, N1, M0	Already performed	Consider tumor testing for: • homologous recombination gene mutations • MSI/dMMR	Consider
Metastatic	Any T, any N, M1	Already performed	Consider tumor testing for: • homologous recombination gene mutations • MSI/dMMR	Consider

dMMR indicates mismatch repair deficiency; MSI, microsatellite instability; PSA, prostate-specific antigen.