Dataset Creation Plan

Project Initiation This Section must be Completed Prior to Project Dataset(s) Creation		
Project Title:	The Dr. Mom Cohort: A Population-Based Study of the Pregnancy Outcomes and Parental Leave Practices of Ontario's Women Physicians	
Project TRIM number:	2019 0906 355 000	
Research Program:	Cancer	
Site:	ICES Central	
Project Objectives:	*Insert Project Objectives as listed in the approved ICES Project PIA*	
	1. To determine if women physicians have different reproductive characteristics (e.g. age at first childbirth, number of childbirths, number of recognized pregnancies, utilization of assisted reproductive technologies) compared to their non-physician counterparts 2. To determine if time to first childbirth is different for women physicians compared to their non-physician counterparts	
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	
Principal Investigator (PI):	Dr. Andrea Simpson	
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*	
	Dr. Nancy Baxter	
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*	
	Mr. Eric McArthur	2018-Jul-25
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*	
	Dr. Maria Cusimano	2018-Feb-28
	Dr. Rinku Sutradhar	2018-Jul-25
	Dr. Amit Garg	2018-Jul-25
	Dr. Joel Ray	2018-Sept-06
	Dr. Simone Vigod	2018-Sept-06
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.*	
	Principal Investigator ☒	2019-04-04
	Responsible ICES Scientist or Second ICES Scientist/Lead ☐	yyyy-mon-dd
	ICES Research and Analysis Staff Creating the DCP ☐	yyyy-mon-dd

Project Initiation This Section must be Completed Prior to Project Dataset(s) Creation		
	ICES Analytic Staff ☒ yyyy-mon-dd	
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*	
DCP Creation Date and Author:	*Date DCP was finalized prior to Project Dataset(s) creation*	*Name of person who created the DCP*
	Date	***Name***
	2019-04-07	Maria Cusimano, Andrea Simpson

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)***
CIHI DAD	April 1, 1988 – Most recent
CIHI SDS	April 1, 1991 – Most recent
NACRS	July 1, 2000 – Most recent
OMHRS	October 1, 2005 – Most recent
CAPE	March 1, 1999 – Most recent
OLIS	January 1, 2007 – Most recent
General Use Datasets – Care Providers	
IPDB	January 1, 1992 – Most recent
General Use Datasets – Population	
RPDB	April 1, 1991 – Most recent
CENSUS	Most recent
General Use Datasets – Coding/Geography/Facilities	
LHIN	Most recent
INST	April 1, 1988 – Most recent
General Use Datasets – Other	
MOMBABY	April 1, 1988 – Most recent
ONMARG	Most recent
HYPER	April 1, 1991 – Most recent
ODD	April 1, 1991 – Most recent
Controlled Use Datasets	
ORGD	January 1, 2007 – Most recent
CIC	January 1985 - April 21, 2016
Other Datasets	
BORN	April 1, 2006 – Most recent
CPSO (imported to ICES)	January 1, 1990 – April 21, 2016

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***
2019-07-22	Andrea Simpson	Add CIC (permission obtained)

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***
yyyy-mon-dd		

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
yyyy-mon-dd

Project Cohort

Study Design

☒ Cohort study ☒ Matched cohort study ☐ Case-control study

☐ Cross-sectional study ☐ Other (specify):

Index Event/Inclusion Criteria

Population: All Ontario women of reproductive age (15-50 years) between January 1, 1995 and November 26, 2018 with valid IKNs

Inclusion Criteria:

1. Female
2. Age 15-50 years at index date
3. Index date between January 1, 1995 to November 26, 2018

Index Date:
Frequency matched analysis: Age 15 for both physicians and non-physicians
Individual matched analysis: Date of first registration with CPSO for physicians, or corresponding simulated date in non-physicians

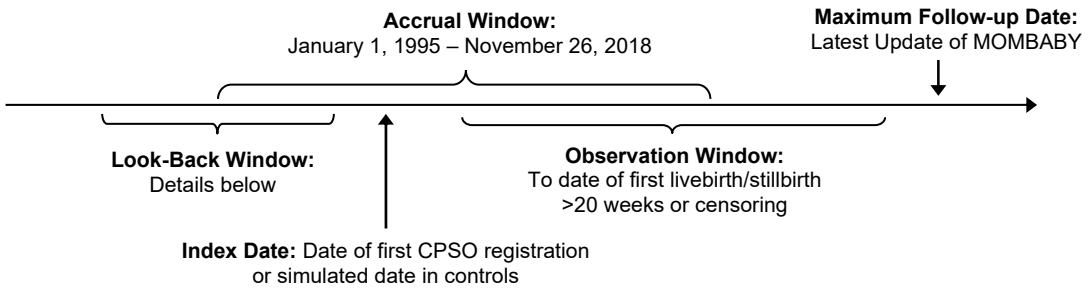
Estimated Size of Cohort

N/A

Exclusions (in order)

Step	Description
1	Non-Ontario resident (first two digits of 'PRCDDABLK' not 35 in RPDB)
2	Non-continuous OHIP eligibility from age 15 years

Project Time Frame Definitions



Accrual Start/End Dates	January 1, 1995 to April 21, 2016
Max Follow-up Date	Latest update of MOMBABY (March 31, 2019)
When does observation window terminate?	Latest update of MOMBABY
Lookback Window(s)	To maximum 2 years: - Covariates: Comorbidities (ADGs) Earliest available: - Covariates: Immigration status, previous livebirths

Variable Definitions (add additional rows as needed)

Main Exposure or Risk Factor	**Physician occupation**
	• Two-level covariate:
	  ○ Exposure: Registration as a physician (i.e. record in CPSO dataset)
	  ○ Comparator: Non-physician (i.e. no record in CPSO dataset)
	• Three-level time-varying covariate
	  ○ Non-physician (i.e. no record in CPSO dataset)
	  ○ Postgraduate trainee (i.e. record in CPSO dataset but not in OHIP)
	  ○ Independent physician (i.e. records in CPSO dataset and OHIP for billings)

Primary Outcome Definition	**First/next childbirth** (time-to-event) = Defined as the first livebirth/stillbirth ≥20 weeks GA experienced after the index date (presence of a record in MOMBABY)
Secondary Outcome Definition(s)	**Age at first/next childbirth** (continuous) = Defined as maternal age at first childbirth (derived from RPDB, MOMBABY)

Baseline Characteristics	**1) Age at index date** (from RPDB):
	• Report as continuous variable (with mean/SD and median/IQR)
	
	2) Year at study entry (from RPDB):
	• Defined as calendar year of index date
	• Also report as categorical variable (1995-2006, 2007-2018)
	
	3) Residential income quintile (from RPDB):
	• Report combined 6-level rural/income variable, with category for missing
	  ○ Rural
	  ○ Urban + Income Quintile 1

Variable Definitions (add additional rows as needed)

	○ Urban + Income Quintile 2 ○ Urban + Income Quintile 3 ○ Urban + Income Quintile 4 ○ Urban + Income Quintile 5 ○ Missing **4) Immigration status** (from CIC): • Report as categorical variable: Canadian born (no landing date or landing date prior to 1985); immigrant (landing date present) **5) Comorbidities** (from NACRS, DAD, SDS, OHIP): • Use ACG System software to collapse patient diagnoses from the preceding 2 years of health service utilization into appropriate Johns Hopkins Aggregated Diagnosis Groups (ADGs) as a measure of comorbidity • Report as categorical variable: no comorbidity (0), low comorbidity (1-5 ADGs), moderate comorbidity (6-9 ADGs), and high comorbidity (≥10 ADGs) **6) Childbirths before index date** (from MOMBABY) • Defined as the total number of livebirth ≥20 weeks GA before the index date (number of records in MOMBABY; full details in Appendix 1) • Report as categorical variable: 0, 1, 2+ • Report as count variable also
Work-Related Characteristics (relevant to Objective 3 on physicans only)	**1) Specialty at index date** (from IPDB, CPSO dataset): • Available for physicians with independent license • Report as binary variable: Family physician vs. other specialists • Algorithm to identify specialty: ○ Use 'Current Specialty' variable in CPSO dataset ○ If 'Current Specialty' variable is missing in CPSO dataset, attempt linkage of physician to either IPDB or OHIP ○ If specialty available in IPDB, then assign appropriately ○ If physnum/spec available in OHIP, then assign appropriately ○ If not linkable to IPDB or OHIP and graduation year ≥2013, then classify as "specialty not yet determined" due to ongoing training

Analysis Plan (expand/modify as needed)

DESCRIPTIVE TABLES

Table 1. Baseline characteristics of frequency-matched & individual-matched cohorts

Table 2. Reproductive outcomes of frequency-matched & individual-matched cohorts

FREQUENCY-MATCHED COHORT

1) Randomly select a subset of non-physicians from RPDB
2) Frequency sampling so ratio of non-physicians to physicians is 5:1 based on the distribution of their birth years

3) Randomly assign a dummy date in non-physicians corresponding to the distribution of CPSO registration dates among physicians; if the non-physician is not eligible on that date, then they should be replaced with a different non-physician
4) Individuals will be followed from their 15th birthday (time 0)
5) Compare baseline characteristics between physicians and non-physicians (standardized differences)
6) Run Cox proportional hazards model for time-to-childbirth & generate cumulative probability curves for childbirth (maternal age as the time scale) for physicians vs. non-physicians
    a. Overall and stratified by specialty
    b. Reproductive lifespan, and <37 years and ≥37 years
    c. With and without adjustment for area-level income quintile

INDIVIDUAL-MATCHED COHORT

1) Hard match each physician with 5 non-physicians (i.e., 1:5 ratio) on:
    - Exact age (years)
    - Year of licensing (actual or simulated date of CPSO registration)
    - Number of livebirths ≥20 weeks' GA before index date (0, 1, >2)
2) Individuals will be followed from the actual or simulated date of CPSO registration (time 0)
3) Compare baseline characteristics between physicians and non-physicians (standardized differences)
4) Run Cox proportional hazards model for time-to-childbirth & generate cumulative probability curves (years of follow-up as the time scale) for physicians vs. non-physicians
    a. Overall and stratified by specialty
    b. Binary variable and three-level time-varying covariate
    c. With and without adjustment for area-level income quintile

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	
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:		