McMaster University

Family Medicine

@PresentersTwitterHandles

@McMasterFamMed

Using Electronic Medical Record Data for Primary Care Research

Jennifer Lawson, MLIS, MSc

Shuaib Hafid, MPH

Research Knowledge and Skills Builder

Watch a recording of the session here

January 16, 2024

Michael G. DeGroote SCHOOL OF MEDICINE

Learning Objectives

Attendees will develop an understanding of:

1. Primary care Electronic Medical Record (EMR) data in Research
   - Strengths, limitations, opportunities
2. Local, provincial and national Primary Care EMR data repositories
   - Broad function of Practice-based learning and research networks, data structure, data governance
3. Working with EMR data
   - Data request, skills, support, building capacity

What is electronic medical record data?

- Patient-level electronic medical documentation captured at clinics to track patients' health status and care patterns.
- Data can include structured and unstructured data:
  - Structured data: Demographic (i.e., age, sex) and clinical characteristics (i.e., diagnoses), billing fee codes, and laboratory test and examination results.
  - Unstructured data: Open-text chart notes for patients.

Working with EMR Data

Strengths

- Captures patient-level longitudinal data.
- Contains granular-level information beyond fee codes billed by physicians.
- Can capture care provided by non-physician health care professionals.
- Cost-effective by leveraging data that is already collected for care documentation (i.e., COUMT).
- Limits information biases such as recall bias, observation bias, and reporting bias.
- Potential to link to other datasets.

Limitations

- Data robustness/comprehensiveness relies on documentation tools/processes used in clinics during care provision.
- Information is siloed within a single care setting (i.e., primary care EMR data will only capture care provided in a primary care clinic, etc.).
- Risk of residual confounding which prevents causal inferences.
- Requires systems and personnel for data extraction, cleaning, and reformatting to make the data research ready.

EMR Data Accessibility

- EMR systems can be directly queried by trained personnel (e.g., OSCAR Demographic Reporting Tool that runs “canned” queries).
  - Works best on data fields known to be complete and clean e.g. Patient DOB, Patient Name and contact information (mostly), patient status (not always up to date)
- EMR data can be curated into data repositories that have governance, security and management to be used as research tools

EMR Research Types

- Epidemiological/Observational Designs
  - Cross-sectional
  - Retrospective
  - Quasi-experimental (pre-post design, interrupted time series)
- Quality Improvement → Learning Health Systems

Practice-based Research Networks

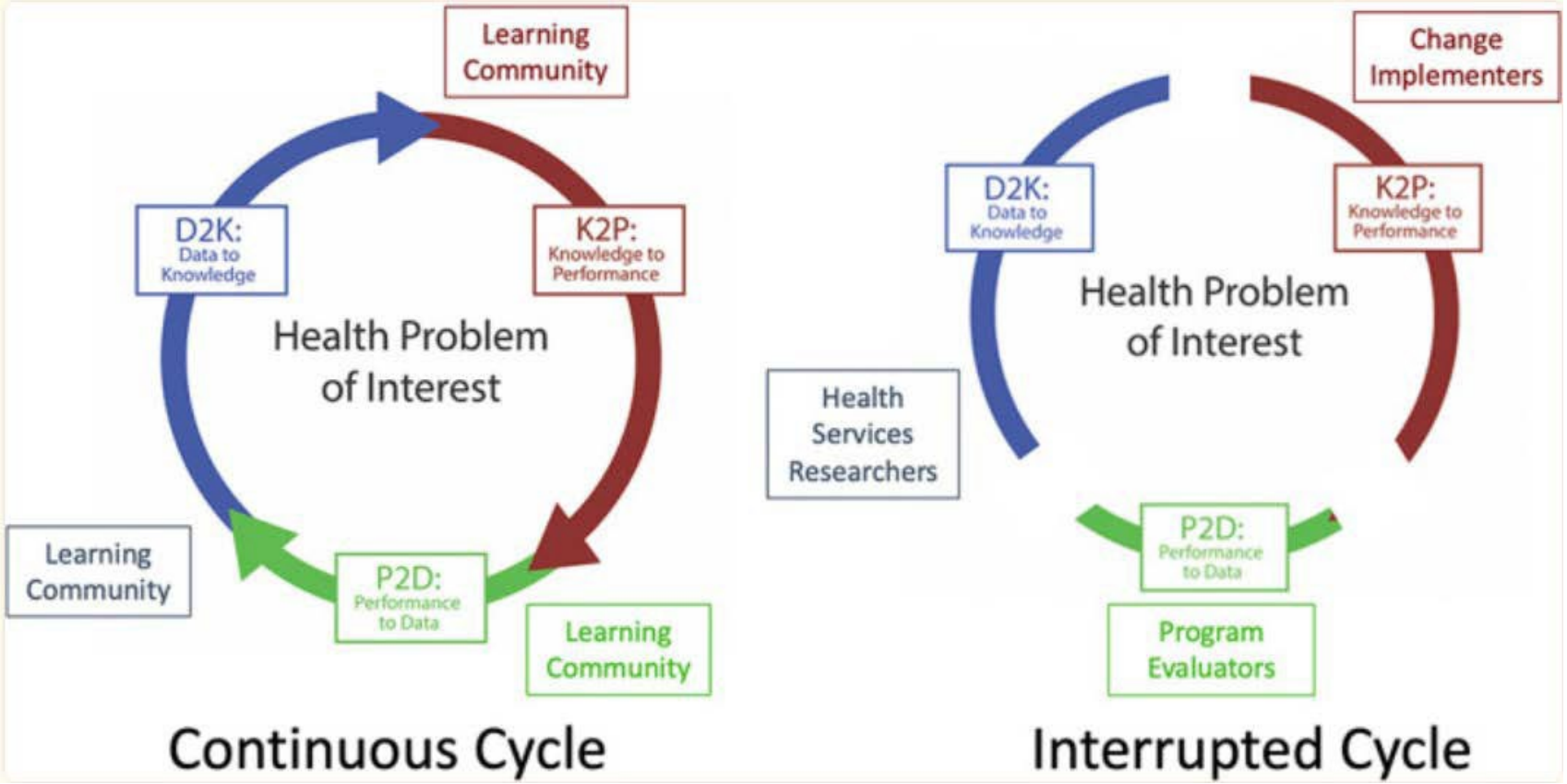
- PBRNs bring together academic researchers and primary care clinicians to conduct research that improves quality of care and health outcomes for patients and populations.
- Sustained collaborations between primary care professionals, researchers, members of the community and policymakers and funders.
- EMR data are a critical backbone to PBLRNs;
- EMR data can be transformed and shared to form data repositories

PBRN Operations

- Requires sustained funding, a governance mechanism and administrative and research support
- Network Director: to set and lead mission
- Network coordinator: recruitment and operational approvals
- Data managers: transform EMR data for research readiness
- Data analysts: methods/techniques for insights from the data
- Researchers to write and disseminate findings.
- 25 years across North America, UK, Europe, Australia, New Zealand
- **PBLRNs** to align with the momentum of Learning Health Systems

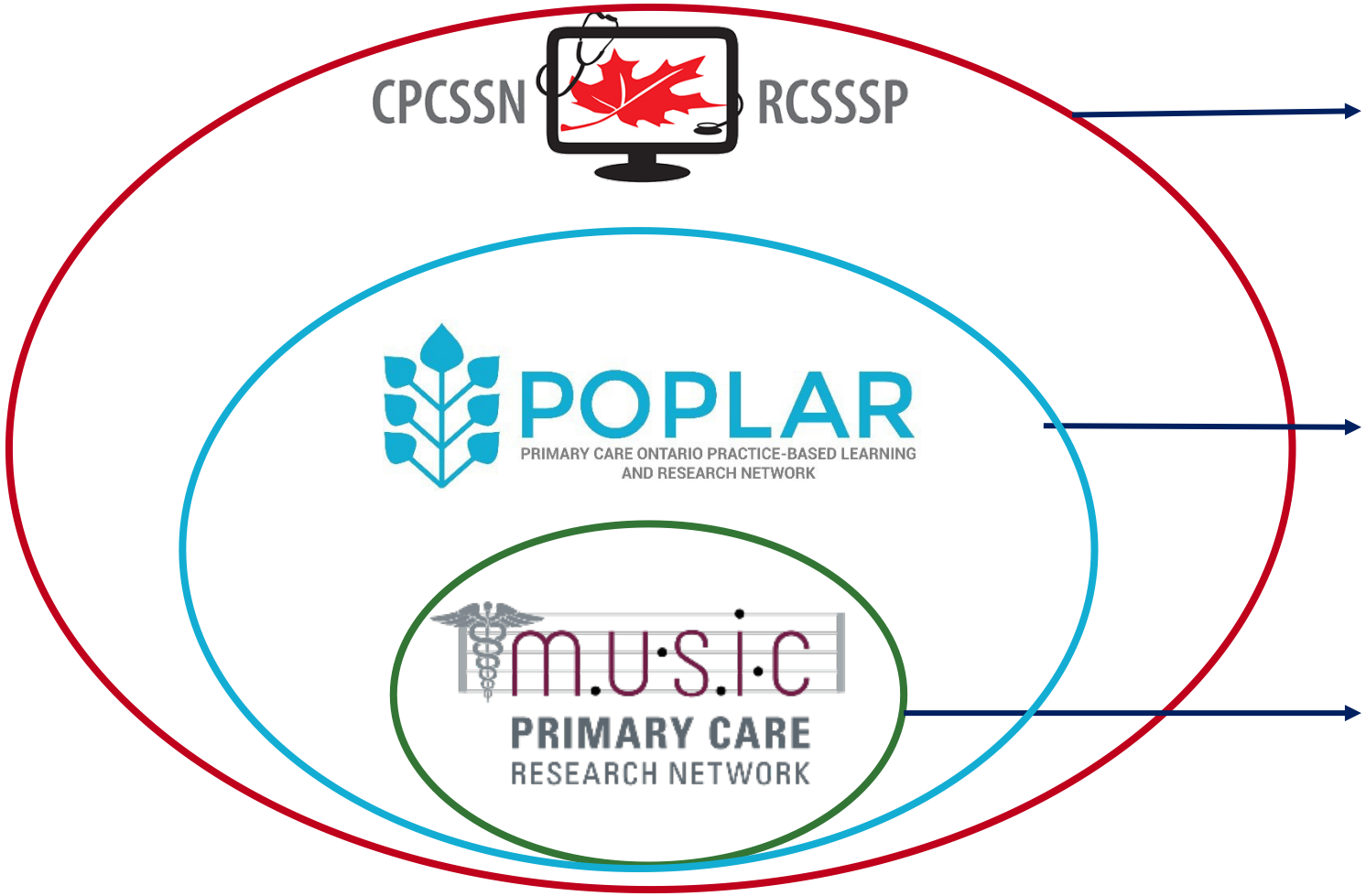
Learning Health System

- Learning health systems (LHSs) are organizations where “science, informatics, incentives, and culture are aligned for continuous improvement and innovation”



Friedman CP. What is unique about learning health systems? Learn Health Syst. 2022 Jul 15;6(3):e10328.

Primary care EMR Data for Research



- National: Canada's first multi-disease electronic medical record surveillance system
- Queen's HSREB is Board of Record

- Provincial: Ontario-wide network
- Funded by Ministry of Health
- Program of INSPIRE-PHC at Queen's
- Currently in development

- Regional: Practice-based Research Network
- HiREB approval
- Upcomng CTO REB Application

All data is housed at Queen's Centre for Advanced Computing – a high security data facility

MUSIC Network: Local/Regional

- **M**cMaster **U**niversity **S**entinel and **I**nformation **C**ollaboration
- Est. in 2014 led by Dee Mangin; research ready database by 2016
- Composed of 4 clinical sites: MFP, SFHC, Burlington, St. Catharines
- Michelle Howard: Lead, Jen Lawson: Co-Ordinator & Analyst; Kris Adamczyk: Data Manager, Rebecca Clark: Data Analyst; Shuaib; Data Analyst
- PC Providers agree to contribute their patients' data as Network Collaborators
- Deidentified data spans 2010 - present; 55 physicians; 89,803 patients (all status); 39,817 patient (1 yr Encounter, 12 yrs old)
- Data schema follows that of CPCSSN and may soon mirror POPLAR's

POPLAR Network: Provincial

- **Primary Care Ontario Practice-based Learning and Research Network**
- Initiated in 2021; funded by the Ministry through INSPIRE-PHC
- A provincial collaboration:
  - 7 Ontario Practice-based Networks; 1000 family physicians that agree to contribute data; 1.5 million patients.
- POPLAR aims to securely collect and de-identify electronic medical record (EMR) data for Research and Quality improvement, aligning the Quintuple Aim.
- Data are housed at the Centre for Advanced Computing at Queens

POPLAR's Aims

Data

- Primary care EMR data
- Data to be collected following Research Ethics Board approval; consistent with privacy legislation
- Data System located in secure, restricted environment
- Data Governance maintained by PBLRNs

Research & QI

- Projects in the public benefit
  - Data projects
  - Support for clinical research
- Linking providers to innovative studies
- Feedback to clinics
  - Support for CareCanvas QI Dashboard

Partnerships

- Network of Networks
- Collaboration: Learning Health System
- Building regional and provincial partnerships
- Contributing data to ICES, CPCSSN

CPCSSN: National

Canadian **P**rimary **C**are **S**entinel **S**urveillance **N**etwork **(CPCSSN):**

- Established in 2008; supported by College of Family Physicians of Canada and the Public Health Agency of Canada**.**
- Co-directorship, core paid staff, steering committee (local Network leads) and several working groups
- Today 11 practice-based research networks across 6 provinces.
- A centralized deidentified patient-level database at the CAC-Queen U
- ~2 million patients across 1,500 primary care practices in Canada.
- Funding has wavered through its 15-year operational span; now CPCRN and PHAC support

CPCSSN PBLRN Composition

1. British Columbia Canadian Primary Care Sentinel Surveillance Network (BC-CPCSSN)
2. Southern Alberta Primary Care Research Network (SAPCReN)
3. Northern Alberta Primary Care Research Network (NAPCReN)
4. Manitoba Primary Care Research Network (MaPCReN)
5. Réseau de recherche en soins primaires de l'Université de Montréal (RRSPUM)
6. Atlantic Practice Based Research Network (APBRN)
7. Maritime Family Practice Research Network (MaRNet)

Ontario CPSSN = POPLAR Networks

8. Deliver Primary Healthcare Information (DELPHI)
9. **McMaster University Sentinel and Information Collaboration (MUSIC)**
10. University of Toronto Practice-Based Research Network (UPLEARN)
11. Eastern Ontario Network (EON)
12. Ottawa Practice Enhancement Network (OPEN)
13. Northern Ontario School of Medicine Research Toward Health Hub (NORTHH)
14. EPIC Practice-based Learning Network (EPIC)

CPCSSN Data Tables

Table Name	Description
Patient	Patient demographic information for all patients linked to a participating provider
Billing	All billing codes submitted to the provinces for each patient.
Encounter	All unique patient encounters.
Encounter Diagnosis	All diagnoses resulting from an encounter with a patient.
Exam	Results of physical exams performed on patients at encounters.
Lab	Results of all requested and completed laboratory tests.
Medication	All medications (type, dosage,) for each patient.
Referral	Referrals made by the provider/practice for the patient.
Disease Case	Patients and disorder details among validated CPCSSN disease case definitions.
Vaccine	All vaccinations administered to the patient.

CPCSSN Data Dictionary

CPCSSN Disease Case Definitions

Adult Asthma	Diabetic Neuropathy
CKD	Diabetic Retinopathy
COPD	Dyslipidemia
Cardiovascular Disease	Epilepsy
Cerebrovascular Disease	Herpes Zoster
Chronic Heart Failure	Hypertension
Cirrhosis	Multiple Sclerosis
Coronary Artery Disease	Non-Valvular Atrial Fibrillation
Dementia	Osteoarthritis
Depression	Parkinson's Disease
Diabetes Mellitus	Pediatric Asthma
	Pediatric Diabetes Mellitus

CPCSSN Supports & Tools

- Data Analyst Working Group – Led by Rachel Morkem

CPCSSN Analyst Guide – Table by Table FAQ

The data dictionary (hyperlink) provides general information on what can kind of data is found within each table, as well as a detailed look at all the variables within each table.

Below you will find answers to common FAQ about each table:

Q. *How complete are the variables?*

Q. *Which variables can you forget about using, and which ones (calc vs orig) are essential to make the table useful?*

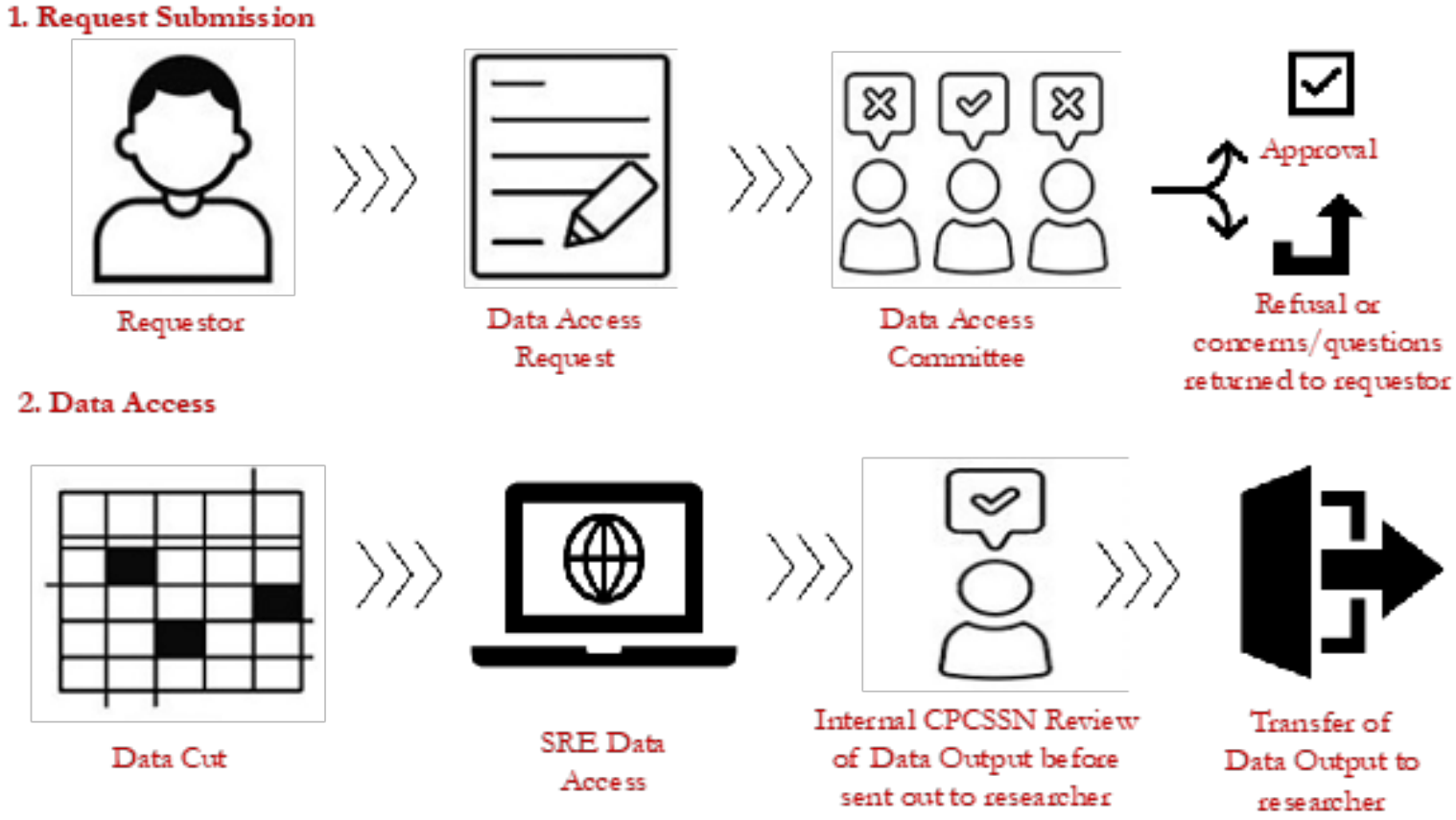
Q. *How are some of the calculated fields derived?*

Q. *How can I use this table effectively?*

Q. *What date variable should I use?*

Selecting EMR Data for a Research Project

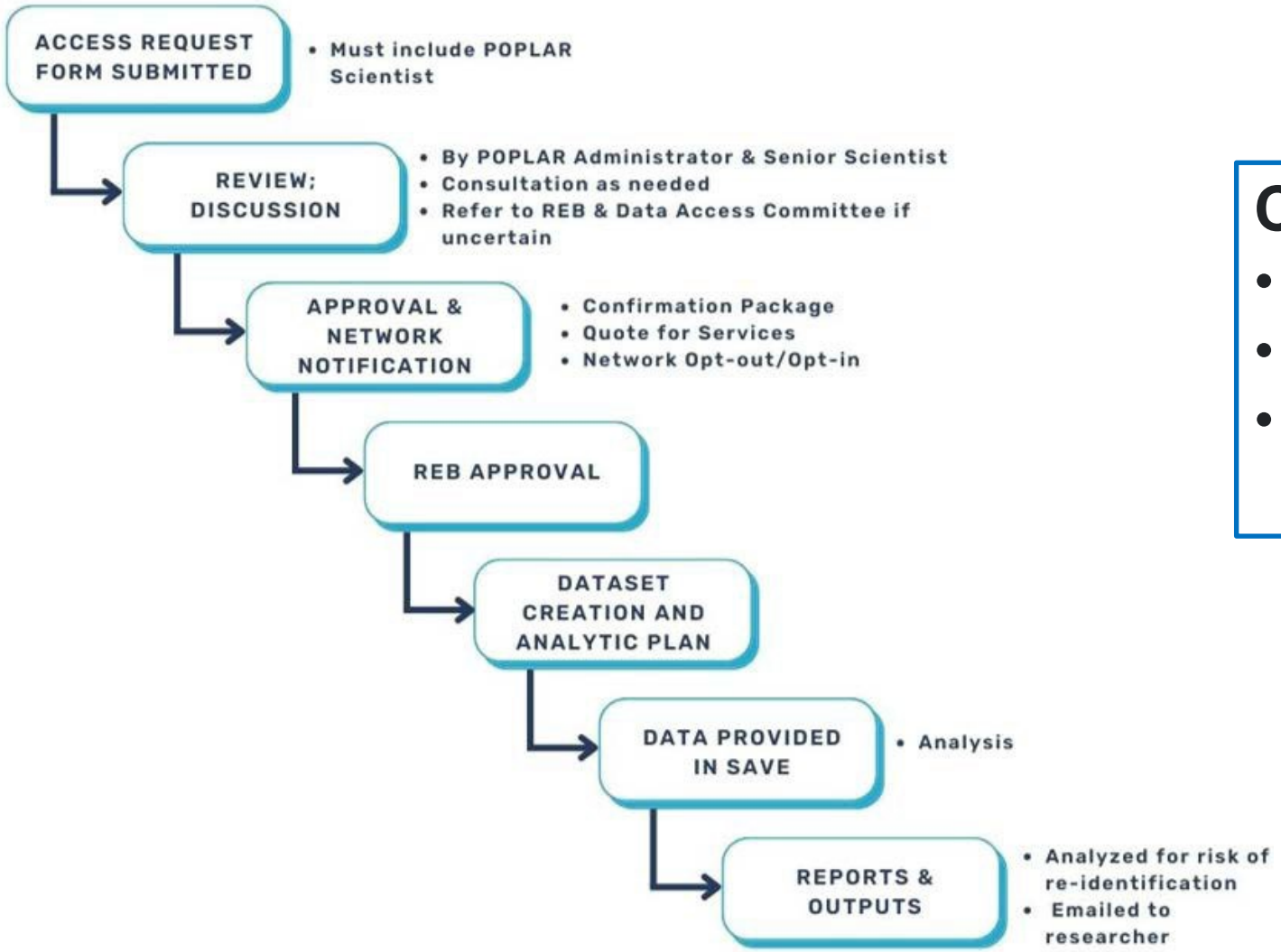
- Data set considerations: local, provincial and national datasets
  - Geographic scope
  - Funding
  - Data availability
  - Timing
  - Data quality



Activating a CPCSSN Project

1. Draft research project protocol that support your research question.
2. Review the CPCSSN Data Dictionary to identify which variables you would like to access and whether your research project is feasible.
3. Submit and obtain research ethics approval for research project protocol.
4. Review the Data Access Process Document and Data Access Fee chart.
5. Submit the following documents to CPCSSN:
    1. Data Access Request form
    2. Secure Research Environment User Access form and Agreement Document
    3. Research project protocol and project ethics approval

Activating a POPLAR Project



Criteria for Approval:

- POPLAR Investigator
- Project Relevance
- Methodologically sound and feasible

Activating a MUSIC project

- MUSIC Application Form

  → vetted MUSIC Governance Committee

- Research Application Clinic Level

  → vetted Health Service Operations

- Define data needs and feasibility of retrieval - consultation
  - Structured, deidentified data follows CPCSSN schema currently
  - Ancillary data by way of querying EMR and linking via patient key

Data Preparation

- Confidence in wrangling with large data.
- Different approaches towards data preparation depending on your EMR source:
  - RIPPLE-C Project (CPCSSN data)
    - Relying on validated definitions that standardize EMR data from across Canada but can be a limitation by not capturing the most granular-level information.
  - Frailty & COVID-19 Project (MUSIC Data)
    - Ability to capture more granular patterns through provincial billing codes but requires competency in working with billing data (i.e., knowledge of billing code structures).

Legacy Drug Prescribing Patterns Study

- Filter applied to MedicationName field to remove non-pharmaceutical data such as, prescribed massage, diabetic needles and messages between the doctors and pharmacist.
- Medication name to ATC code translation checks.
- Drug prescribing durations could not rely on start and stop dates of each prescription alone due to data anomalies and so start date of first prescription and end date of last prescription were also taken into account .

Duration Analysis and Legacy Criteria

Patient_ID	StartDate	StopDate	Duration (d)	Drug Name
853146	**3/31/16**	6/29/16	90	CYMBALTA 60MG
853146	7/11/16	10/9/16	90	CYMBALTA 60MG
853146	11/3/16	2/1/17	90	CYMBALTA 60MG
853146	12/21/16	**3/16/18**	450	CYMBALTA 60MG
Sum Duration			**720**	
Start-Stop Duration			**715**	

Data Analysis & Visualization Skills

- Confident in ‘data gymnastics’
- Experience with database programming languages (i.e., SQL).
- Experience with statistical programming languages (i.e., SAS, R, Python [pandas library]) – do not use SPSS.
- Using programming languages to produce customizable exhibits:
  - R: GGPLOT2, forestplot
  - Python: matplotlib, seaborn

**Hyperlinks lead to educational resources.*

MUSIC Ex. The association between patients' frailty status, multimorbidity, and demographic characteristics and changes in primary care for chronic conditions during the COVID-19 pandemic: a pre-post study

Shireen Fikree, Shuaib Hafid, Jennifer Lawson, Gina Agarwal, Lauren E Griffith, Liisa Jaakkimainen, Derelie Mangin, Michelle Howard
Family Practice 2023; 40(4); 523-530.

Objectives:

- To assess the impact of the COVID-19 pandemic on primary care management of older adults with common chronic conditions (diabetes, hypertension, and chronic kidney disease).
- To examine whether any changes were associated with age, sex, neighbourhood income, multimorbidity, and frailty.

Study Design:

- Retrospective pre-post design cohort study (n = 2,043) using MUSIC data who had a valid Clinical Frailty Scale assessment completed.

MUSIC Ex. The association between patients' frailty status, multimorbidity, and demographic characteristics and changes in primary care for chronic conditions during the COVID-19 pandemic: a pre-post study

Shireen Fikree, Shuaib Hafid, Jennifer Lawson, Gina Agarwal, Lauren E Griffith, Liisa Jaakkimainen, Derelie Mangin, Michelle Howard
Family Practice 2023; 40(4); 523-530.

Data Tables Accessed:

- Independently collected frailty scores were linked to MUSIC-CPCSSN data
  - **Disease Registry:** To identify baseline comorbidity status.
  - **Billing:** To capture encounter and encounter types.
  - **Patient:** To describe patient demographic characteristics.
  - **Medication:** To capture medication prescribing patterns.
  - **Risk Factors:** To capture patients' smoking history.
  - **Labs:** To capture results from completed laboratory tests.
  - **Measure/Exam:** To capture measurements from exams conducted at encounters (i.e., BP, weight, height, BMI).

MUSIC Ex. The association between patients' frailty status, multimorbidity, and demographic characteristics and changes in primary care for chronic conditions during the COVID-19 pandemic: a pre-post study

Shireen Fikree, Shuaib Hafid, Jennifer Lawson, Gina Agarwal, Lauren E Griffith, Liisa Jaakkimainen, Derelie Mangin, Michelle Howard
Family Practice 2023; 40(4); 523-530.

Subgroup	No. Blood Pressure Measures	Mean Change (95% CI)	No. Prescriptions for HTN Medications (HTN patients only)	Mean Change (95% CI)	No. Prescriptions for HTN Medications (DM patients only)	Mean Change (95% CI)	No. Prescriptions for Metformin	Mean Change (95% CI)
Total Cohort		−1.11 (−1.28 to −1.00)		−0.48 (−0.75 to −0.20)		−0.41 (−0.76 to −0.06)		−0.13 (−0.30 to −0.04)
Age group (years):								
65–69		−1.05 (−1.30 to −0.80)		−0.51 (−1.00 to −0.03)		−0.04 (−0.55 to 0.48)		−0.54 (−0.79 to −0.28)
70–74		−1.21 (−1.47 to −0.94)		−0.70 (−1.27 to −0.14)		−0.58 (−1.30 to 0.14)		0.20 (−0.17 to 0.56)
75–79		−1.23 (−1.55 to −0.91)		−0.20 (−0.89 to 0.50)		−0.68 (−1.68 to 0.28)		−0.02 (−0.41 to 0.37)
80–84		−1.09 (−1.56 to −0.62)		−0.48 (−1.26 to 0.30)		−0.45 (−1.64 to 0.73)		0.00 (−0.54 to 0.54)
>85		−1.12 (−1.53 to −0.71)		−0.30 (−0.98 to 0.38)		−0.67 (−1.83 to 0.49)		−0.05 (−0.47 to 0.37)
Sex:								
Male		−1.17 (−1.40 to −0.97)		−0.58 (−1.01 to −0.16)		−0.36 (−0.83 to 0.11)		−0.18 (−0.45 to 0.09)
Female		−1.12 (−1.30 to −0.93)		−0.42 (−0.78 to −0.06)		−0.45 (−0.97 to 0.07)		−0.08 (−0.30 to 0.13)
Frailty score category:								
Low (0 – 3)		−1.20 (−1.39 to −1.00)		−0.65 (−1.02 to −0.29)		−0.27 (−0.72 to 0.19)		−0.15 (−0.39 to 0.10)
Moderate (4 – 6)		−1.11 (−1.32 to −0.90)		−0.14 (−0.58 to 0.31)		−0.56 (−1.11 to 0.00)		−0.12 (−0.35 to 0.12)
High (7 – 9)		−0.60 (−1.21 to 0.01)		−1.73 (−2.89 to −0.56)		−0.50 (−5.09 to 4.09)		0.00 (−1.79 to 1.79)
Median neighbourhood income quintile:								
1 (lowest)		−0.72 (−1.15 to −0.29)		−0.13 (−1.08 to 0.82)		−0.70 (−3.13 to 1.73)		−0.47 (−1.08 to 0.14)
2		−1.26 (−1.53 to −0.98)		−0.93 (−1.51 to −0.36)		−0.05 (−0.40 to 0.50)		−0.15 (−0.45 to 0.15)
3		−1.28 (−1.53 to −1.03)		−0.29 (−0.74 to 0.16)		−0.44 (−0.97 to 0.10)		−0.01 (−0.27 to 0.25)
4		−0.94 (−1.20 to −0.69)		−0.37 (−0.90 to 0.15)		−0.35 (−1.02 to 0.32)		−0.09 (−0.43 to 0.25)
5 (highest)		−0.89 (−1.71 to −0.08)		−0.27 (−1.65 to 1.11)		−3.60 (−9.85 to 2.65)		−1.14 (−3.94 to 1.65)
Number of prevalent chronic conditions:								
1		−1.07 (−1.63 to −0.51)		0.00 (−1.28 to 1.28)		−0.55 (−1.85 to 0.85)		−0.55 (−1.37 to 0.28)
2		−1.43 (−1.77 to −1.08)		−0.62 (−1.25 to 0.02)		−0.36 (0.88 to 0.16)		−0.04 (−0.40 to 0.31)
3		−1.16 (−1.52 to −0.80)		−0.92 (−1.52 to −0.31)		−0.23 (−0.88 to 0.41)		−0.08 (−0.25 to 0.41)
>4		−1.02 (−1.19 to −0.85)		−0.32 (−0.69 to 0.05)		−0.51 (−1.20 to 0.18)		−0.18 (−0.42 to 0.05)

HTN = Hypertension, DM= Diabetes mellitus

MUSIC Ex. The association between patients' frailty status, multimorbidity, and demographic characteristics and changes in primary care for chronic conditions during the COVID-19 pandemic: a pre-post study

Shireen Fikree, Shuaib Hafid, Jennifer Lawson, Gina Agarwal, Lauren E Griffith, Liisa Jaakkimainen, Derelie Mangin, Michelle Howard
Family Practice 2023; 40(4); 523-530.

Conclusions:

- During the COVID-19 pandemic, chronic condition care declined slightly.
- Measures such as blood pressure and HbA1c (Hemoglobin A1c) did not change.
- Changes in care were not associated with patient health and demographic factors.
- Virtual primary care likely contributed to maintaining chronic condition care.

CPCSSN Ex. Describing primary care patterns before and during the COVID-19 pandemic across Canada: a quasi-experimental pre-post design cohort study using national practice-based research network data.

Shuaib Hafid, Karla Freeman, Kris Aubrey-Bassler, John A. Queenan, Jennifer Lawson, Meredith Vanstone, Kathryn Nicholson, Marie-Thérèse Lussier, Derelie Mangin, Michelle Howard [Submitted for publication to British Journal of General Practice]

Objective:

- To analyze how the pandemic affected primary care access and comprehensiveness in chronic disease management by comparing primary care patterns before and during the early COVID-19 pandemic.

Study Design:

- Quasi-experimental pre-post design cohort study (n = 919,928) using CPCSSN data from January 1, 2018 – December 31, 2021.

CPCSSN Ex. Describing primary care patterns before and during the COVID-19 pandemic across Canada: a quasi-experimental pre-post design cohort study using national practice-based research network data.

Shuaib Hafid, Karla Freeman, Kris Aubrey-Bassler, John A. Queenan, Jennifer Lawson, Meredith Vanstone, Kathryn Nicholson, Marie-Thérèse Lussier, Derelie Mangin, Michelle Howard [Submitted for publication to British Journal of General Practice]

Data Tables Accessed:

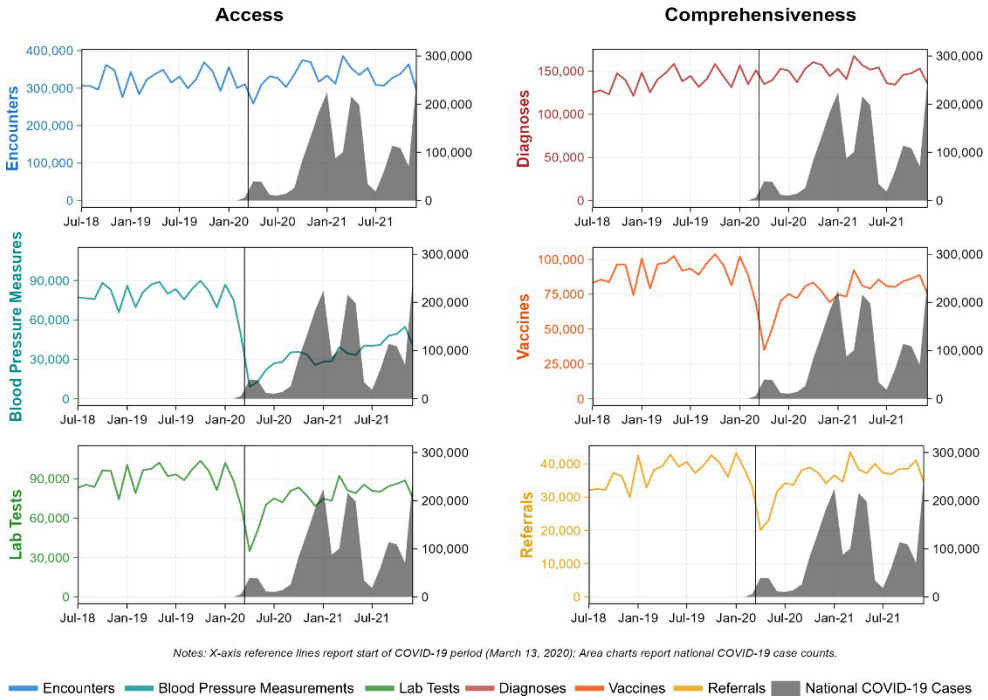
- **Patient**: To capture demographic information about patients and to apply age exclusion criteria.
- **Encounter**: To capture the number of unique encounters that occurred.
- **Encounter Diagnosis**: To capture the number of unique diagnoses addressed.
- **Exam**: To capture the number of unique exams completed.
- **Lab**: To capture the number of unique laboratory tests completed.
- **Referral**: To capture the number of unique referrals sent to specialists.
- **Vaccine**: To capture the number of non-COVID-19 vaccinations administered.
- **Disease Case**: To describe the patients baseline comorbidity status using validated CPCSSN case definitions.

CPCSSN Ex. Describing primary care patterns before and during the COVID-19 pandemic across Canada: a quasi-experimental pre-post design cohort study using national practice-based research network data.

Shuaib Hafid, Karla Freeman, Kris Aubrey-Bassler, John A. Queenan, Jennifer Lawson, Meredith Vanstone, Kathryn Nicholson, Marie-Thérèse Lussier, Derelie Mangin, Michelle Howard [Submitted for publication to British Journal of General Practice]

Conclusions:

- Primary care encounters remained consistent during the pandemic, but in-person care, such as lab tests and blood pressure measurements, decreased.
- In-person care indicators followed reacted temporally to national COVID-19 case counts during the pandemic.



Family Medicine | McMaster University

Building EMR Research Capacity

- Composing novel, relevant research questions aligned with EMR data
- Local pilot projects and scaled provincial or national projects
- Call for funding and personnel to build standardized data infrastructure, test, improve, maintain data repositories
- Advocate for key external partners: EMR vendors, communities and OHTs, formation and good functioning of regulatory and auditing authorities
- PC community buy-in and good governance structures that ensure that data is robust, meaningful and used for the greater good

Questions?

If you have any additional questions, please reach out to us directly.

Thank you!

McMaster University

Family Medicine

Department of Family Medicine
Michael G. DeGroote School of Medicine
Faculty of Health Sciences

fammedmcmaster.ca
@McMasterFamMed

McMaster University

Family Medicine

Supplementary Slides