Identifying children with medical complexity in administrative datasets in a Canadian context: study protocol

Holly McCulloch 1,2, Sydney Breneol 3, Samuel A Stewart,4 Sandra Magalhaes,5 Mari Somerville 1,2, Jordan Sheriko,6 Shauna Best,1 Stacy Burgess,1 Elizabeth Jeffers,6 Mary-Ann Standing,7 Sarah King,1 Julie Clegg,1 Janet A Curran 1,2,5

ABSTRACT

INTRODUCTION

Children with medical complexity and their families are an important population of interest within the Canadian healthcare system. Despite representing less than 1% of the paediatric population, children with medical complexity require extensive care and account for one third of paediatric healthcare expenditures. Opportunities to conduct research to assess disparities in care and appropriate allocation of health resources relies on the ability to accurately identify this heterogeneous group of children. This study aims to better understand the population of children with medical complexity in the Canadian Maritimes, including Nova Scotia (NS), New Brunswick (NB) and Prince Edward Island (PEI). This will be achieved through three objectives: (1) Evaluate the performance of three algorithms to identify children with medical complexity in the Canadian Maritimes in administrative data; then using the ‘best fit’ algorithm (2) Estimate the prevalence of children with medical complexity in the Canadian Maritimes from 2003 to 2017 and (3) Describe patterns of healthcare utilisation for this cohort of children across the Canadian Maritimes.

Methods and analysis

The research will be conducted in three phases. In Phase 1, an expert panel will codevelop a gold-standard definition of paediatric medical complexity relevant to the Canadian Maritime population. A two-gate validation process will then be conducted using NS data and the gold-standard definition to determine the ‘best fit’ algorithm. During phase 2 the ‘best fit’ algorithm will be applied to estimate the prevalence of children with medical complexity in NS, NB and PEI. Finally, in phase 3 will describe patterns of healthcare utilisation across the Canadian Maritimes.

Ethics and dissemination

Ethics approval for this protocol was granted by the institutional research ethics board at the IWK Health Centre (REB # 1026245). A waiver of consent was approved. This study will use administrative data; then using the ‘best fit’ algorithm. During phase 2 the ‘best fit’ algorithm will be applied to estimate the prevalence of children with medical complexity in NS, NB and PEI. Finally, in phase 3 will describe patterns of healthcare utilisation across the Canadian Maritimes.

Strengths and limitations of this study

- Clinical experts and families with lived experience will develop and operationalise a gold-standard definition for children with medical complexity in the Maritimes.
- Multiple methods will be employed to validate the ‘best fit’ algorithm.
- Certain clinical variables relevant to describing children with medical complexity may not be available within health administrative data.
- Health administrative data is limited by type of provider/service and reporting and extraction practices.

INTRODUCTION

Since 2010, children with medical complexity have gained increasing attention as an important population in critical need of practice and policy reform within the Canadian healthcare system.1,2 Medical complexity is generally characterised as having one or more complex chronic condition(s) associated with significant functional limitations, high health resource use and family-identified needs.1 3 One seminal Canadian study conducted nearly 10 years ago, suggested that despite representing less than 1% of the paediatric population in Ontario, children with medical complexity account for one third of paediatric healthcare expenditures.1 Adding to the findings stemming from this study, a recent publication from The Canadian Institute for Health Information (CIHI) reported that in 2015–2016 there were 948 per 100 000 children and youth with medical complexity.4 Recent findings from the United States also suggest that the prevalence of children living with medical complexity is increasing,5 6 likely due to the increased survival rates of a variety of life-limiting and life-threatening conditions.7 Information regarding children with medical complexity has been primarily
derived using reports from the United States with only a few seminal reports stemming from Canadian-based data. However, due to important population differences, it is critical to understand this vulnerable population in the Canadian context to assist in mapping health outcomes and healthcare utilisation for this vulnerable population. Doing so may ensure that more relevant, needs based and sustainable models of service delivery are developed to optimise quality health experiences and outcomes for children and their families.

Caregivers of children with medical complexity carry a tremendous amount of responsibility, stress and financial burden to attend to their intensive care needs. Opportunities to assess disparities in care, and appropriately allocate health resources to better serve children with medical complexity, are dependent on being able to accurately identify them. A standard definition of children with medical complexity does not currently exist in Canada and determining a classification system for these children is a necessary first step to optimise health service delivery for this vulnerable group of children. Given the heterogeneity of this group, with variation in the severity and combination of comorbid chronic conditions, there are unique challenges when attempting to identify a cohort at a population level. Identification begins with selecting a method to classify and characterise children with medical complexity.

The Canadian Maritimes is a unique paediatric care setting composed of three provinces: Nova Scotia (NS), New Brunswick (NB), and Prince-Edward-Island (PEI). The only paediatric tertiary care facility in the Maritimes is in NS, which results in families crossing provincial jurisdictions for specialty care. This adds a layer of contextual difference that may intersect with and/or contribute to medical complexity and healthcare utilisation in the Maritimes.

Health administrative data algorithms

Three methods commonly used with health administrative data in the study of children with medical complexity include: (1) Cohen et al’s list of complex chronic conditions, technological assistance and neurological impairment; (2) Simon et al’s Paediatric Medical Complexity Algorithm (PMCA) and (3) Feudtner et al’s complex chronic conditions classification system. Cohen et al’s algorithm was developed in Ontario using hospital discharge data but has not yet been validated. Algorithms developed by Simon et al and Feudtner et al have not undergone validation in Canadian health data.

Algorithm 1

Cohen et al operationalised a definitional framework in an Ontario administrative dataset by using sets of ICD-10-CA codes relevant to complex chronic conditions, neurological impairment and technological assistance. This definitional framework aligns with the work of other experts conducting research on this population outside of Canada. However, the sensitivity and specificity of the lists of Canadian Classification of Procedures and Interventions for identifying children with medical complexity in Canada has not yet been reported.

Algorithm 2

Simon et al used the Chronic Disability Payment System as a guide to develop the PMCA at the Seattle Children’s Hospital in Washington State. The authors employed a systematic process beginning with the development of consensus definitions for three levels of medical complexity (complex chronic, noncomplex chronic and no chronic conditions). They classified children with medical complexity as those who fit their definition for having a complex chronic condition. They proceeded by selecting and modifying an existing algorithm to conform to the consensus definitions, and selected a gold-standard paediatric population through medical chart review to evaluate the sensitivity and specificity of the algorithm. The PMCA had high sensitivity and specificity (complex chronic: 86% sensitivity, 86% specificity; non-complex chronic: 65% sensitivity, 84% specificity; children without complex chronic: 77% sensitivity and 93% specificity) for identifying children with medical complexity and was subsequently updated and validated for International Classification of Diseases, Tenth Revision, Clinical Modifications (ICD-10-CM) codes in 2018. This algorithm has a least and more conservative version depending on the type of data available to researchers. The least conservative version was shown to perform better in hospital discharge data and the most conservative version was shown to perform better in claims and billing data.

Algorithm 3

Feudtner et al followed a similar approach as Simon et al in Washington State, creating a working definition of complex chronic conditions and subsequently operationalised the definition using clinical knowledge and existing literature. This resulted in a list of conditions and their corresponding ICD-9-CM codes that are highly sensitive (87%) for identifying children with complex chronic conditions. The algorithm was updated in 2014 and the ICD codes were translated into the 10th edition (ICD-10-CM), including both diagnostic and procedural codes indicative of technology dependence or organ transplantation. Lindley et al have evaluated the predictive ability of the original and modified versions and determined that the modified version, which will be used in this study, better operationalises medical complexity.

Aims and objectives

The aim of this study is to better understand the Canadian Maritime’s population of children with medical complexity. The ability to accurately identify children with medical complexity is necessary to conduct research that may inform the design and implementation of successful population-level policies and interventions. This study will be conducted in the Canadian Maritime provinces: NS, NB, and PEI. This study will: (1) evaluate
the performance of three administrative data algorithms to identify children with medical complexity; then use the ‘best fit’ algorithm to (2) estimate the prevalence; and (3) describe patterns of healthcare utilisation for this cohort of children from 2003 to 2017.

METHODS AND ANALYSIS

Data sources

The study cohorts will be constructed using the CIHI Discharge Abstract Databases (CIHI DAD) from NS and the prevalence and healthcare utilisation estimates will use the CIHI DAD, Vital Statistics (VITAL), Insured Patient Registry (MASTER), and Physician Billing Databases (MED) from each province. The administrative datasets for NS, NB and PEI are housed by their respective data repository organisations: Health Data Nova Scotia (HDNS), NB Institute for Research Data and Training (NB-IRDT), and Secure Island Data Repository (SIDR).

Study design

This study will be conducted in three phases to identify children with medical complexity in the Canadian Maritimes and their healthcare utilisation over a 15-year period (figure 1).

In phase 1, the ‘best fit’ algorithm for identifying children with medical complexity in the Maritimes will be selected using sensitivity and specificity. Using NS administrative data, a cohort will be established for each of the three algorithms. A gold-standard definition will be developed for children with medical complexity in NS and will be edited through email. The definition will be circulated to the expert stakeholder group after the meeting for their review and comments. A coding manual will be developed to accompany the medical chart extraction tool.

In phase 2, the prevalence of children with medical complexity in the Maritime provinces will be estimated. This will be done in all three Maritime provinces using NS, NB and PEI administrative data cohorts identified by the previously selected ‘best fit’ algorithm.

In phase 3, patterns of healthcare utilisation of children with medical complexity across the Maritime provinces will be described.

Phase 1: identify the ‘best fit’ algorithm for identifying children with medical complexity in the Maritimes

Developing a gold-standard definition

A multidisciplinary expert consensus meeting will be convened to co-develop a gold-standard definition of children with medical complexity relevant to the Canadian Maritimes. The consensus meeting will involve relevant stakeholders, including a minimum of two clinicians, two researchers and a parent from each of the three Maritime provinces. Parent participants will be invited to attend a pre-meeting virtual session to introduce them to the consensus meeting agenda and provide them with an opportunity to ask questions related to the project. A parent partner and research coordinator will facilitate the pre-meeting session. An infographic will be developed and distributed prior to the consensus meeting to introduce participants to existing frameworks and literature that will help guide the conceptualisation of medical complexity, functional limitations, and other related concepts. Outputs from relevant exploratory work conducted by members of the research team will be presented during the first session of the consensus meeting. Parent and clinician experts will participate in structured breakout sessions to examine all meeting materials in the context of their clinical or lived experience. Sessions will be audio-recorded, and a note-taker will track participation. After the breakout sessions are complete, a consensus based decision-making process will be used where participants work together through discussion to reach agreement on the elements of a clinically meaningful gold-standard definition for this paediatric population in the Maritimes. The definition will be circulated to the expert stakeholder group after the meeting for their review and comments and will be edited through email.

To operationalise the definition, a subgroup of clinicians with expertise working with this patient population will use the gold-standard definition to develop a list of clinical variables that will form the basis of a medical chart extraction tool. The expert group will complete this task over a series of meetings, and the final tool will be reviewed by clinicians from both NB and PEI. A coding manual will be developed to accompany the medical chart extraction tool.
Establishing an NS administrative data cohort

The algorithm validation process will be undertaken using administrative data from the NS provincial data repository organisation, HDNS. Children living in NS with medical complexity will be identified and classified with specific ICD-10-CA codes defined within each of the three respective algorithms, resulting in a distinct cohort being derived from each algorithm. Due to the uncertainty of how the Simon et al algorithm will perform using Canadian health administrative data, both the least and more conservative versions will be used. All children aged 0–18 years of age who have a discharge record in NS CIHI DAD from 2003 to 2017 will be eligible for inclusion.

Examining the ‘best fit’

Given the prevalence of children with medical complexity is rare, a two-gate case-control study design will be used to validate the algorithms (figure 2).18 This will be done in two steps. First, a stratified random sample of 100 children will be selected from each of the cohorts derived by the algorithms to participate in a chart audit.19 The chart audit will be conducted by nurses trained in abstracting records of children with medical complexity using the medical chart extraction tool developed by the expert panel. The audit will begin with all abstractors independently extracting the same five charts to calibrate coding. Results will be compared, and the inter-rater reliability will be estimated using a kappa score. If the kappa score is low and there are many discrepancies, the chart audit tool will be refined, and further training provided to the abstractors. An additional five charts will then be extracted and compared. This process will be repeated until the chart extraction is determined to be reliable. A total of 30% of the records in each cohort will be independently abstracted by a second reviewer.

In the second step, two clinically derived groups of children (complex and not complex) will be established by healthcare providers from across NS to contribute to the validation procedure. Healthcare providers who contribute to this process will be from varying specialties and settings (ie, primary and tertiary care) and provide direct patient care or care coordination activities to children with medical complexity and their families. Sixty children with medical complexity will be identified from relevant clinical areas that fit within the study’s gold-standard definition. Additionally, a cohort of 400 controls who are children without medical complexity, according to the study’s gold standard definition will be identified. All clinical experts participating in this step will undergo training prior to identifying patients for inclusion to ensure they understand the gold-standard definition and the clinical variables that informed the development of the medical chart audit tool.

Sample size and data analysis

There are approximately 3200 Nova Scotian children discharged each year from the paediatric tertiary care facility in the Maritimes. Cohen et al has estimated a 0.67% prevalence of children with medical complexity, therefore, we estimated approximately 21 Nova Scotian children with medical complexity are discharged yearly.3 However, local expert clinicians estimated that there are closer to 50 Nova Scotian children discharged each year with medical complexity. To ensure a large enough cohort to power a regression analysis, prevalence will be examined over a 15-year period.

The validation of the algorithms will occur in three parts (figure 3). The algorithm-identified cohorts from step 1 will be evaluated using positive predictive values, to determine what percentage of the algorithm-identified patients truly qualified as medically complex. The gold-standard sample from step 2 will be evaluated using sensitivity and specificity, to evaluate how well the algorithms can identify both complex and not complex patients. Finally, the two samples will be combined into a single analysis. All statistics will be estimated with 95% CIs, and F-statistics will be used to investigate the trade-offs between sensitivity and specificity.20 As medical complexity is by definition a rare occurrence, particular attention will be paid to the specificity, in order to ensure that the algorithm is not overwhelming the final cohort with false positives.

Phase 2: prevalence of children with medical complexity in all three Maritime provinces

The algorithm identified as the ‘best fit’ in phase 1, based on NS data, will be applied to NB’s and PEI’s CIHI DAD through NB-IRDT and SIDR respectively, to identify the cohort of children with medical complexity using the same...
eligibility parameters as NS. Statistical code to generate the cohorts developed in NS during the validation step, will be used, with minor modification as needed.

Prevalence of medical complexity will be calculated using DAD during the years 2003–2017 in each Maritime province. Individual level data will use encrypted health card numbers and birthdate in month/year as unique identifiers to link each child’s data across datasets and over time.

**Sample size and data analysis**

To estimate the annual prevalence, and corresponding 95% CIs, of children with medical complexity per 100,000 paediatric population in the Maritime provinces during the study period, the number of children identified by the selected algorithm will be divided by the total number of children (≤18 years) in the 2016 paediatric population for each province. Prevalence estimates will be stratified based on relevant variables (eg, age, sex).

**Phase 3: patterns of healthcare utilisation**

Patterns of healthcare utilisation including hospital admissions, outpatient visits and same-day surgeries during the years 2003–2017, or up to 18 years of age will be described in each Maritime province. Individual level data will use encrypted health card numbers and birthdate in month/year as unique identifiers to link each child’s data across datasets and over time.

**Sample size and data analysis**

Annual rates, and their corresponding 95% CIs, will be estimated for several types of healthcare utilisation. Specifically, data availability allows describing rates of acute hospitalisations, unplanned readmissions, day surgeries, hospital-based outpatient visits, community-based primary care visits and specialty care visits. Healthcare utilisation follow-up will begin at the index date. The index date will be defined as the first discharge date with a complex chronic condition, by the ‘best fit’ algorithm. Children will be censored at the end of the study period or sooner if they die or turn 18 before 2017. The numerator will be the total number of annual records for each child in the cohort. Each child will contribute 1 year of person-time to the denominator for every year they are alive and living in the province. Results from each province will be described separately by summarising findings and comparing healthcare utilisation across provinces.

This project is currently in phase one in the process of identifying the ‘best fit’ algorithm using the two-gate case–control study. The project is anticipated to be completed by December 2022.

**Patient and Public Involvement**

This study will use an integrated knowledge translation (iKT) approach, where end users are involved in each stage of the project.21 The team of researchers, clinicians, administrators and patient partners will work together throughout the research process to develop outputs that are relevant to knowledge user needs, including development of the research objectives and methods. Opportunities will be created to engage additional parents, clinicians, and administrators to inform different stages of the work as the project unfolds. The proposed iKT method was chosen to increase uptake of the research into policy and practice.22

---

**Figure 3** Analysis of gold-standard sample of cases with medical complexity (complex) and controls (not complex). PPV, positive predictive values.
ETHICS AND DISSEMINATION
Ethical approval was received from the primary research institution in NS (IWK Health Centre; REB#: 1026245). A waiver of consent has been granted by the research ethics board, and therefore, informed participant consent is not required for this study. Ethical approval will be obtained at the appropriate institutions in NB and PEI before data abstraction commences in these provinces.

This study will use an iKT approach, where end users are involved in each stage of the project, which could increase uptake of the research into policy and practice. The findings of this research study will be submitted for publication and dissemination through conference presentations and with our end users.

DISCUSSION
To our knowledge, this will be the first time the performance of these three algorithms will be evaluated using Canadian health administrative data. While the recent CIHI report identified children with medical complexity across Canada to provide population-level prevalence rates and an overview of their healthcare utilisation, this study will add to the existing knowledge by including a longer to identify any potential trends.

Findings from this work may contribute to a priority task, identified by Children’s Healthcare Canada, for a national effort to support system change for children with medical complexity. The opportunity to comprehensively identify a ‘best fit’ algorithm, that is able to describe children with medical complexity, addresses a significant gap in health research and is essential for better management and improvement of health outcomes. The methods developed (eg, two-gate validation procedure, chart audit tool) may be useful to other jurisdictions across Canada wishing to validate algorithms specifically for their region. Further, identifying a ‘best fit’ algorithm to describe children with medical complexity may allow us to characterise population-level prevalence and identify patterns in health service utilisation, as well as identify gaps in service, for this vulnerable population. The formation of this cohort may act as a starting point to conduct future population-based research that can contribute evidence to support health system planning and policy-making and help improve health outcomes. Key stakeholders are being involved throughout this project to help ensure research findings related to service and resource gaps are translated to health system and policy change.

Limitations and mitigating strategies
The team has reflected on potential challenges in the conduct of the study and have planned strategies accordingly to mitigate these difficulties should they arise. A challenge lies in the lack of consensus about how to operationalise a gold-standard definition of children with medical complexity. The plan to convene an expert panel and employ multiple methods of algorithm validation was chosen to address this limitation. Simon et al’s and Feudtner et al’s administrative data algorithms were developed and validated using ICD-CM codes, while Canada uses the ICD-CA iteration of the global ICD codes. The differences between these iterations are expected to be negligible as they relate to administrative needs, not the medical diagnoses. Additional limitations concerning the use of health data are the availability and accuracy of important variables of interest. For example, the order that diagnoses are coded (eg, primary diagnosis, secondary diagnoses) is not always consistent. To help mitigate this, all diagnoses from a patient encounter coded in the CIHI DAD will be included. Finally, estimating healthcare utilisation using health administrative data has limitations as not all services are reflected in existing databases. The team of patient partners, researchers, clinicians and administrators recognise the importance of capturing alternative or relevant non-health service utilisation (eg, massage therapy, informal respite care) and indirect costs for families (eg, travel, lost time at work). The use of additional administrative data from other sources is being explored as well as qualitative interviews with patients, families and providers in a future study.
REFERENCES