AWARD NUMBER: W81XWH-19-1-0177

TITLE: Using Administrative Health Data to Identify Patients with NF1 in Ontario, Canada, and to Assess Prevalence, Mortality, and Health Care Utilization Patterns

PRINCIPAL INVESTIGATOR: Carolina Barnett-Tapia

CONTRACTING ORGANIZATION: University Health Network

REPORT DATE: July 2022

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2022	Annual	15JUN2021 - 14JUN2022

4. TITLE AND SUBTITLE

Using Administrative Health Data to Identify Patients with NF1 in Ontario, Canada, and to Assess Prevalence, Mortality, and Health Care Utilization Patterns

5a. CONTRACT NUMBER
W81XWH-19-1-0177

5b. GRANT NUMBER
NF180027

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Barnett-Tapia, Carolina

E-Mail: c.barnetttapia@utoronto.ca

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University Health Network
200 Elizabeth St, Toronto, ON
M5G 2C4

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We are using electronic medical records and administrative database to study health care use of people living with NF1 in Ontario, Canada. Using electronic medical record database (EMRPC, previously called EMRALD), we estimated a minimum prevalence of NF1 between 1:2532 to 1:3851.

A previously developed billing algorithm has poor performance and won't allow proper assessment of health utilization patterns. Therefore, we have moved forward with a mitigation strategy, whereby we have created a registry of people with confirmed NF1 followed at tertiary centers This registry has been linked to administrative database, and matched to healthy controls to compare mortally and use of health care.

We are also studying mortality and healthcare use in patients identified through electronic medical records, using EMRPC and UTOPIAN databases

15. SUBJECT TERMS
NF1, prevalence, billing codes, EMR

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	47	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

INTRODUCTION:

The objective of this study is to develop an algorithm to identify people with NF1 living in the province of Ontario, Canada (population ~ 14 million), through administrative health databases. This algorithm will be used to study incidence, prevalence, NF1-related mortality and health care utilization patterns of patients with NF1. The specific aims for this project are outlined below.

Specific Aim 1: To develop and validate an algorithm to identify patients with NF1 in the province of Ontario.

Specific Aim 2: To estimate the incidence, prevalence and mortality of patients with NF1 in Ontario, Canada. We hypothesize that we can obtain reliable estimates of incidence and prevalence of NF1 in the province which we expect to be within previously published ranges. We also plan to calculate mortality ratios stratified by age and hypothesize that individuals with NF1 have higher mortality ratios compared to the general population.

Specific Aim 3: To study the health-utilization patterns of NF1 patients which includes the number of primary care, specialist and emergency visits, outpatient surgeries, hospital admissions, mental health care and pain treatments Compared to a matched cohort of healthy controls, we hypothesize that patients with NF1 will have, on average, significantly more visits at all levels of health care.

1. KEYWORDS:

NF1, neurofibromatosis 1, electronic medical record, EMR, administrative health databases, algorithm, prevalence, health utilization.

2. ACCOMPLISHMENTS:

Mayor goals/tasks for reporting period (as stated in SOW)

1. **Specific Aim 1, Major task 4:** 100% completion Manuscript submitted, revisions requested and re-submitted
2. **Specific Aim 2, Major task 1 :** Create dataset and obtain data: 100% completion. Datasets cleaned up in late 2021 and transferred to ICES. Data validated and duplicated removed, final cohort meeting inclusion/exclusion criteria has 1,213 individuals with NF1.
3. **Specific Aim 2, Major task 2:** Analyze data for incidence, prevalence, mortality: 90% completed. Prevalence estimated from EMRALD data (manuscript submitted), re-assessed in UTOPIAN data (in preparation). From registry data, 15 (1.2%) NF1 individuals died during study window. Will compare to matched controls.
4. **Specific aim 3, Major task 1:** 95% complete. NF1 registry complete and linked to ICES. Now finishing matched control cohort (have matched ~50% of cases, with 1:4 ratio)
5. **Specific Aim 3, Major task 2: 15% complete.** Common diagnostic and assessment billing codes in primary care practices assessed though UTOPIAN database. ER visits, admissions, Mental health-related use and pain management use to be assessed this coming year.

As stated in our annual report for the first year, the billing algorithm had poor performance, therefore we are using our mitigation strategies to accomplish Aims 2 and 3. In our proposal we had 2 mitigation strategies and we are using both: using electronic medical records data to assess health utilization, and also creating a registry of individuals with NF1 from tertiary care and linking it to the administrative databases at ICES.

We have used a different EMR database (UTOPIAN), which is also linked to ICES. We demonstrated that our original EMR search strategy was feasible in UTOPIAN, although required minimal validation of identified cases by a clinician. Overall, we found that using NF1-secific terms in the cumulative patient profile to identify cases of NF1, works reasonably well in the UTOPIAN database even though the specific neurofibromatosis terms are different. Within UTOPIAN eligible individuals (n=**421,971)** we identified 127 individuals with possible/probable NF1 (estimated prevalence 1 in 3,322). The 127 NF1 cases were matched to 635 controls. We found a higher number of primary care encounters with a diagnosis of anxiety (86.8/100 NF1 patients vs. 54.0/100 controls) essential hypertension, HIV, skin conditions (see data attachment, manuscript in preparation). The controls had more encounters primarily for contraception management, asthma, disorders of menstruation)

We developed a registry of individuals with NF1 seen until December 2020, at the main academic centres with specialised NF1 care in Ontario: the University Health Network and the Hospital for Sick Children. The cohort has 1,213 individuals with NF1 (after removing duplicated and those with exclusion criteria). We have linked this registry to ICES databases, and are creating the matched control cohort. In the final year of the grant we will compare health care utilization including: ER visits, specialty visits, admissions to hospital, admission for mental health care, skin-related procedures covered by the provincial health care plan, pain medications, including opioids, nerve blocks, etc.

What opportunities for training and professional development has the project provided?

Nothing to report

How were the results disseminated to communities of interest?

Presented internally during Canadian neurofibromatosis rounds.

What do you plan to do during the next reporting period to accomplish the goals?

i. We will complete manuscript of EMR search validation using UTOPIAN data and billing patterns in primary care.

ii. We will compare mortality and health utilization patterns between NF1 individuals to matched controls. This will include NF1 individuals from the registry and those identified form EMR databases.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

i. We have created a large, fully linked registry of individuals with confirmed NF1, and will match to non-NF1 controls. This will allow the comparison of health utilization parameters between groups, including for pain procedures, mediation, metal health resources, and death.

ii. We have assessed the feasibility of using our previously developed EMR algorithm in a new database. We have preliminary assessment of commonly billed codes by primary care providers looking after people with NF1.

What was the impact on other disciplines?

Nothing to Report

What was the impact on technology transfer?

We have already validated a simple EMR search that may be used in any healthcare setting that uses electronic medical records, to identify individuals with NF1. We have also created a large registry of NF1 individuals, fully linked to administrative health care databases. This will allow to complete our Aim3, but will also allow large-scale research in NF1.

What was the impact on society beyond science and technology?

Nothing to Report

5. CHANGES/PROBLEMS:

No new issues in the past year.

Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

Actual or anticipated problems or delays and actions or plans to resolve them

As expected, we had many delays over the past year, as ICES has prioritized COVID-related projects. However, we completed all administrative approvals and were able to successfully transfer and link registry data. We have the final year to complete analyses between NF1 individuals and controls.

Changes that had a significant impact on expenditures

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report

Significant changes in use or care of human subjects

We have obtained approvals to link registry data for tertiary care to ICES data. All the analyses are the same, and all data are deidentified.

Significant changes in use or care of vertebrate animals

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Barnet C, Candido E, Chen B, Pequeno P, Parkin P, Tu K. Development of algorithms to identify individuals with Neurofibromatosis type 1 within administrative data and electronic medical records in Ontario, Canada. Under review (revisions requested, re-submitted, Orphanet)

Books or other non-periodical, one-time publications.

Nothing to Report

Other publications, conference papers and presentations.

Barnet C, Candido E, Chen B, Pequeno P, Parkin P, Tu K. Development of algorithms to identify individuals with Neurofibromatosis type 1 within administrative data and electronic medical records in Ontario, Canada. preprint https://doi.org/10.21203/rs.3.rs-1666441/v1

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

EMR search strategy to identify NF1 cases, in manuscript.

- **Inventions, patent applications, and/or licenses**

Nothing to Report

-

Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	*Carolina Barnett-Tapia*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	ORCID ID 0000-0001-5546-0221
Nearest person month worked:	3.6
Contribution to Project:	*Dr. Barnett-Tapia has coordinated this project. She performed final chart review to classify records as definitive or possible NF1. She reviewed all billing codes from cases to develop the algorithms*
Funding Support:	

Name:	*Elisa Candido*
Project Role:	Co-investigator
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1.2
Contribution to Project:	*Ms. Candido coordinated the ICES personnel, directed the algorithm development, and helped draft mitigation strategies for Aims 2 and 3*
Funding Support:	NA

Name:	Karen Tu
Project Role:	*Co-investigator*
Researcher Identifier (e.g. ORCID ID):	ORCID ID: 0000-0003-0883-4934
Nearest person month worked:	1.2
Contribution to Project:	*Dr. Tu provided expertise in developing search strategy within EMRALD and developing billing algorithms. She has provided access to UTOPIAN for external validation of EMR algorithm*
Funding Support:	

Name:	*Patricia Parkin*
Project Role:	*Co-investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1.0
Contribution to Project:	*Dr. Parkin has provided clinical expertise to determine billing codes to use in algorithms; she also has helped with registry of patients followed at the Hospital for Sick Children since the 1990s.*
Funding Support:	

Name:	*Branson Chen*
Project Role:	Health Information Analyst
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1
Contribution to Project:	*Mr. Chen conducted the search within EMRALD, analyzed abstractor reliability, and developed EMR algorithm*
Funding Support:	

Name:	*Priscila Pequeno*
Project Role:	Senior Research Analyst
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1
Contribution to Project:	*Ms. Pequeno developed and tested all the billing algorithms*
Funding Support:	

Name:	Meg Mendoza
Project Role:	Research Analyst
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	*Mr.* Mendoza prepared UHN REB application for linkage to ICES data and coordinated with HSC
Funding Support:	

Name:	Samantha Lee
Project Role:	ICES epidemiologist
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	MS. Lee has helped coordinate linkage project, and plan data collection for health care utilization
Funding Support:	

Name:	Alejandro Hernandez
Project Role:	ICES research analyst
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2
Contribution to Project:	Mr. Hernandez is working on the linked data, creating matched cohort and will work on health care utilization analysis
Funding Support:	

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report

What other organizations were involved as partners?

1. **Organization Name:** Institute for Clinical Evaluative Science (ICES)
2. **Location of Organization:** *Toronto, Canada*
3. **Partner's contribution to the projec**t: Collaboration

The ICES hosts all health administrative data for the province of Ontario, and we contracted their services for this study.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: *NA*

QUAD CHARTS: *NA*

9. APPENDICES:

- UTOPIAN data analysis
- Submitted manuscript (revised version)

Validation of an EMR algorithm to identify individuals with Neurofibromatosis type 1

Analysis Report

Version 4.0

February 16, 2022

Report authors:

Ellen Stephenson

Jemisha Apajee

Brief Summary

Project aim: The purpose of this project was to show that the EMR search strategy for identifying cases of NF1 in EMRPC was translatable to a different database, UTOPIAN Data Safe Haven. The second goal was to compare health utilization among NF1 patients identified in EMRPC and UTOPIAN databases.

Brief conclusion: The exact EMR search strategy validated in EMRPC did not translate well to UTOPIAN. The limited set of terms used to search the CPP for NF1 cases missed several definite or possible NF1 cases in UTOPIAN because of differences in how physicians document the diagnosis in free text. By expanding the search strategy to include more terms indicative of neurofibromatosis, we were able to find a comparable number of definite and possible cases of NF1, similar to the prevalence originally found in EMRPC. In EMRPC the search strategy was validated using chart abstraction; this was not done in UTOPIAN so the precise sensitivity and specificity of the original or modified search strategy is unknown. If the cases detected with the additional terms are all true cases of NF1, then the sensitivity of the precise EMRPC search strategy is very poor when applied to UTOPIAN EMR data (<50%). However, if the search strategy using in EMRPC is taken to mean searching for neurofibromatosis and related terms in the cumulative patient profile to identify cases of NF1, then this approach does appear to work reasonably well in the UTOPIAN database even though the specific neurofibromatosis terms are different.

The following section includes a comparison of billing patterns for NF1 patients identified in the UTOPIAN database against a sample of matched controls. Cases were matched based sex, birth year and EMR start year (5 control patients: 1 NF1 patient).

Case definition used to identify patients with NF1 in UTOPIAN

Cases with NF1 in the UTOPIAN database were identified by searching for certain key words in the CPP followed by a manual review by a clinician. The set of key words consisted of terms closely related to neurofibromatosis: “NF1”, “Neurofibromatosis 1”, “Neurofibromatosis type 1” and “Von Recklinghausen”. The search method involved using regular expressions to account for typos as well as differences in how physicians document the diagnosis in free text. Once the list of diagnoses with at least one occurrence of any of the key words was created, a clinician reviewed the text and labelled them as “Definitely NF1”, “NF2”, “Not NF1”, “Possibly NF1” or Possibly NF2”. In this report patients with NF1 are those who have at least one diagnosis in the CPP labelled as “Definitely NF1” or “Possibly NF1”.

Cohort Selection

Project Cohort			
Data Extract: (e.g., 2020Q2)	2020 Q4		
Study Design:	Cross-sectional		
Inclusion Criteria:	**NF1 cases** Patients in the UTOPIAN database with sufficient data quality who have been identified as having Neurofibromatosis **Matched cases** Patients in the UTOPIAN database with sufficient data quality who can be matched with NF1 patients on sex, birth year and EMR start date		
Cohort selection	*Step* Description	Number of physicians	Number of patients
	0 *2020 Q4 data extract*	**429**	
	1 *Sufficient physician-level data quality*	**392**	
	2 *Sufficient patient level data quality* • *valid birth date and sex recorded* • *Patients who are rostered at time of data extraction or active within the family practice (i.e., family practice visit in the past 3 years or other visit type within the past 3 years and a populated cumulative patient profile in family practice EMR)* • *Patients with at least 1 year since EMR start date*		**421971**
	3a Definite NF1 term in Health condition table of CPP		61
	3b Possible NF1 term in Health condition table of CPP		72
	3a or 3b Any neurofibromatosis keyword in Health Conditions Table of CPP		**127**

Table 1. Demographic characteristics of NF1 and control cohorts

	Cohort			
	Control patients		Definitely/Possibly NF1	
	N	%	N	%
Total	635		127	
Sex*				
Female	340	53.5%	68	53.5%
Male	295	46.5%	59	46.5%
Neighborhood income quintile				
1 (lowest)	137	21.6%	30	23.6%
2	114	18.0%	21	16.5%
3	107	16.9%	25	19.7%
4	102	16.1%	16	12.6%
5 (highest)	160	25.2%	34	26.8%
Missing	15	2.4%	1	0.8%
Rurality				
Rural	60	9.4%	11	8.7%
Urban	568	89.4%	115	90.6%
Missing	7	1.1%	1	0.8%
Age at December 2020*				
Mean age in years (SD)	36.39	22.61	36.39	22.61
Median age in years (IQR)	33	(19-55)	33	(19-55)
Age at last visit				
Mean age in years (SD)	34.93	23.03	35.17	22.90
Median age in years (IQR)	31	18-54	32	18-54
Length of time on the EMR*				
Mean number of years (SD)	8.27	4.68	8.10	4.50
Median number of years (IQR)	8	(5-10)	8	(5-10)

* These variables were used for case-control matching.

Most common outpatient diagnostic OHIP codes

Table 2a. Number of patients with at least one of the most common diagnostic codes from 2015-2019

Diagnosis code	Description of diagnosis code*	Matched controls		NF1 cases	
		N	% of cohort	N	% of cohort
460	Acute nasopharyngitis [common cold]	146	22.99	32	25.20
300	Anxiety, dissociative and somatoform disorders	114	17.95	28	22.05
787	Symptoms involving digestive system	107	16.85	27	21.26
799	Other ill-defined and unknown causes of morbidity and mortality	122	19.21	22	17.32
917	General medical examination	77	12.13	22	17.32
916	Other healthy infant or child receiving	75	11.81	16	12.60
691	Atopic dermatitis and related conditions	79	12.44	16	12.60
896	Need for prophylactic vaccination and inoculation against combinations of diseases	94	14.80	16	12.60
401	Essential hypertension	62	9.76	14	11.02
781	Symptoms involving nervous and musculoskeletal systems	93	14.65	13	10.24
724	Other and unspecified disorders of back	43	6.77	13	10.24
709	Other disorders of skin and subcutaneous tissue	34	5.35	13	10.24
780	General symptoms	56	8.82	11	8.66
372	Disorders of conjunctiva	32	5.04	11	8.66
250	Diabetes mellitus	41	6.46	10	7.87
009	Ill-defined intestinal infections	42	6.61	10	7.87
079	Viral and chlamydial infection in conditions classified elsewhere and of unspecified site	33	5.20	9	7.09
650	Normal delivery	26	4.09	8	6.30
599	Other disorders of urethra and urinary tract	45	7.09	8	6.30
466	Acute bronchitis and bronchiolitis	27	4.25	8	6.30
715	Osteoarthrosis and allied disorders	40	6.30	7	5.51
895	Encounter for contraceptive management	53	8.35	7	5.51
311	Depressive disorder, not elsewhere classified	27	4.25	6	4.72
627	Menopausal and postmenopausal disorders	20	3.15	6	4.72
682	Other cellulitis and abscess	21	3.31	6	4.72
786	Symptoms involving respiratory system and other chest symptoms	55	8.66	6	4.72
796	Other nonspecific abnormal findings	49	7.72	6	4.72
840	Sprains and strains of shoulder and upper arm	20	3.15	6	4.72

785	Symptoms involving cardiovascular system	47	7.40	5	3.94
388	Other disorders of ear	16	2.52	5	3.94
847	Sprains and strains of other and unspecified parts of back	28	4.41	5	3.94
461	Acute sinusitis	31	4.88	5	3.94
493	Asthma	43	6.77	5	3.94
626	Disorders of menstruation and other abnormal bleeding from female genital tract	40	6.30	5	3.94
379	Other disorders of eye	5	0.79	5	3.94
382	Suppurative and unspecified otitis media	17	2.68	5	3.94
381	Nonsuppurative otitis media and eustachian tube disorders	33	5.20	5	3.94
564	Functional digestive disorders, not elsewhere classified	21	3.31	5	3.94
706	Diseases of sebaceous glands	17	2.68	5	3.94

* The description of the diagnosis code is taken from the UTOPIAN database; it does not necessarily match the description in the OHIP manual.

Codes occurring for fewer than 5 patients with NF1 are not shown.

Table 2b. Number of records of the most common diagnostic codes from 2015-2019

Diagnosis code	Description of diagnosis code*	Matched controls		NF1 cases	
		N	Rate per 100 patients	N	Rate per 100 patients
916	Other healthy infant or child receiving	837	131.81	148	116.54
300	Anxiety, dissociative and somatoform disorders	343	54.02	110	86.61
250	Diabetes mellitus	399	62.83	99	77.95
460	Acute nasopharyngitis [common cold]	392	61.73	80	62.99
401	Essential hypertension	235	37.01	69	54.33
650	Normal delivery	227	35.75	58	45.67
044	Human immunodeficiency virus [HIV] disease	127	20.00	56	44.09
799	Other ill-defined and unknown causes of morbidity and mortality	297	46.77	54	42.52
917	General medical examination	201	31.65	49	38.58
787	Symptoms involving digestive system	229	36.06	45	35.43
781	Symptoms involving nervous and musculoskeletal systems	238	37.48	43	33.86
691	Atopic dermatitis and related conditions	149	23.46	39	30.71
896	Need for prophylactic vaccination and inoculation against combinations of diseases	277	43.62	34	26.77
724	Other and unspecified disorders of back	91	14.33	33	25.98
785	Symptoms involving cardiovascular system	80	12.60	29	22.83
311	Depressive disorder, not elsewhere classified	64	10.08	23	18.11
079	Viral and chlamydial infection in conditions classified elsewhere and of unspecified site	75	11.81	22	17.32
599	Other disorders of urethra and urinary tract	114	17.95	20	15.75
780	General symptoms	92	14.49	19	14.96
709	Other disorders of skin and subcutaneous tissue	50	7.87	18	14.17
627	Menopausal and postmenopausal disorders	67	10.55	17	13.39
372	Disorders of conjunctiva	53	8.35	16	12.60
009	Ill-defined intestinal infections	64	10.08	16	12.60

715	Osteoarthrosis and allied disorders	83	13.07	16	12.60
895	Encounter for contraceptive management	152	23.94	16	12.60
682	Other cellulitis and abscess	41	6.46	16	12.60
491	Chronic bronchitis	10	1.57	16	12.60
466	Acute bronchitis and bronchiolitis	37	5.83	15	11.81
216	Benign neoplasm of skin	23	3.62	14	11.02
808	Fracture of pelvis		0.00	14	11.02
786	Symptoms involving respiratory system and other chest symptoms	100	15.75	13	10.24
388	Other disorders of ear	56	8.82	12	9.45
847	Sprains and strains of other and unspecified parts of back	59	9.29	12	9.45
461	Acute sinusitis	46	7.24	11	8.66
493	Asthma	103	16.22	11	8.66
427	Cardiac dysrhythmias	47	7.40	11	8.66
626	Disorders of menstruation and other abnormal bleeding from female genital tract	99	15.59	10	7.87
727	Other disorders of synovium, tendon, and bursa	84	13.23	10	7.87
796	Other nonspecific abnormal findings	80	12.60	9	7.09
840	Sprains and strains of shoulder and upper arm	34	5.35	9	7.09
379	Other disorders of eye	7	1.10	9	7.09
616	Inflammatory disease of cervix, vagina, and vulva	53	8.35	9	7.09
382	Suppurative and unspecified otitis media	33	5.20	8	6.30
536	Disorders of function of stomach	18	2.83	8	6.30
381	Nonsuppurative otitis media and eustachian tube disorders	60	9.45	7	5.51
564	Functional digestive disorders, not elsewhere classified	50	7.87	7	5.51
706	Diseases of sebaceous glands	44	6.93	7	5.51
595	Cystitis	33	5.20	7	5.51
967	Need for prophylactic vaccination and inoculation against influenza	22	3.46	7	5.51
733	Other disorders of bone and cartilage	18	2.83	7	5.51
848	Other and ill-defined sprains and strains	29	4.57	7	5.51
884	Multiple and unspecified open wound of upper limb	8	1.26	7	5.51

579	Intestinal malabsorption		0.00	7	5.51
034	Streptococcal sore throat and scarlet fever	52	8.19	6	4.72
242	Thyrotoxicosis with or without goiter	6	0.94	6	4.72
244	Acquired hypothyroidism	54	8.50	5	3.94
692	Contact dermatitis and other eczema	45	7.09	5	3.94
078	Other diseases due to viruses and chlamydiae	72	11.34	5	3.94
278	Overweight, obesity and other hyperalimentation	23	3.62	5	3.94
477	Allergic rhinitis	59	9.29	5	3.94
625	Pain and other symptoms associated with female genital organs	15	2.36	5	3.94
909	Social maladjustment	4	0.63	5	3.94
680	Carbuncle and furuncle	4	0.63	5	3.94
831	Dislocation of shoulder	7	1.10	5	3.94
902	Educational circumstances	1	0.16	5	3.94
269	Other nutritional deficiencies	2	0.31	5	3.94
332	Parkinson's disease	9	1.42	5	3.94
807	Fracture of rib(s), sternum, larynx, and trachea	1	0.16	5	3.94
879	Open wound of other and unspecified sites, except limbs	8	1.26	5	3.94
788	Symptoms involving urinary system	19	2.99	4	3.15
486	Pneumonia, organism unspecified	39	6.14	4	3.15
739	Nonallopathic lesions, not elsewhere classified	34	5.35	4	3.15
127	Other intestinal helminthiases	1	0.16	4	3.15
303	Alcohol dependence syndrome	14	2.20	4	3.15
288	Diseases of white blood cells		0.00	4	3.15
290	Dementias	17	2.68	4	3.15
492	Emphysema	3	0.47	4	3.15
560	Intestinal obstruction without mention of hernia		0.00	4	3.15
829	Fracture of unspecified bones	4	0.63	4	3.15
313	Disturbance of emotions specific to childhood and adolescence	14	2.20	3	2.36
373	Inflammation of eyelids	21	3.31	3	2.36
280	Iron deficiency anemias	25	3.94	3	2.36
356	Hereditary and idiopathic peripheral neuropathy	4	0.63	3	2.36

429	Ill-defined descriptions and complications of heart disease	14	2.20	3	2.36
437	Other and ill-defined cerebrovascular disease		0.00	3	2.36
455	Hemorrhoids	8	1.26	3	2.36
464	Acute laryngitis and tracheitis	13	2.05	3	2.36
487	Influenza	41	6.46	3	2.36
845	Sprains and strains of ankle and foot	24	3.78	3	2.36
919	Superficial injury of other, multiple, and unspecified sites	43	6.77	3	2.36
010	Primary tuberculous infection	10	1.57	3	2.36
075	Infectious mononucleosis	1	0.16	3	2.36
153	Malignant neoplasm of colon	4	0.63	3	2.36
193	Malignant neoplasm of thyroid gland	9	1.42	3	2.36
451	Phlebitis and thrombophlebitis	4	0.63	3	2.36
214	Lipoma	4	0.63	2	1.57
346	Migraine	26	4.09	2	1.57
349	Other and unspecified disorders of the nervous system	4	0.63	2	1.57
463	Acute tonsillitis	44	6.93	2	1.57
550	Inguinal hernia	12	1.89	2	1.57
553	Other hernia of abdominal cavity without mention of obstruction or gangrene	2	0.31	2	1.57
611	Other disorders of breast	9	1.42	2	1.57
844	Sprains and strains of knee and leg	49	7.72	2	1.57
963	Need for prophylactic vaccination and inoculation against measles-mumps-rubella (MMR)	4	0.63	2	1.57
969	Unspecified combined vaccine	29	4.57	2	1.57
977	Unspecified systemic agent causing adverse effects in therapeutic	4	0.63	2	1.57
070	Viral hepatitis	5	0.79	2	1.57
154	Malignant neoplasm of rectum, rectosigmoid junction, and anus	2	0.31	2	1.57
232	Carcinoma in situ of skin	1	0.16	2	1.57
272	Disorders of lipoid metabolism	46	7.24	2	1.57
315	Specific delays in development	2	0.31	2	1.57
368	Visual disturbances	1	0.16	2	1.57
438	Late effects of cerebrovascular disease		0.00	2	1.57
521	Diseases of hard tissues of teeth	13	2.05	2	1.57

535	Gastritis and duodenitis	17	2.68	2	1.57
606	Infertility, male	1	0.16	2	1.57
714	Rheumatoid arthritis and other inflammatory polyarthropathies	19	2.99	2	1.57
729	Other disorders of soft tissues	6	0.94	2	1.57
901	Family disruption	9	1.42	2	1.57
930	Foreign body on external eye	3	0.47	2	1.57
V70.0	Routine general medical examination at a health care facility	13	2.05	2	1.57
005	Other food poisoning (bacterial)		0.00	1	0.79
057	Other viral exanthemata	5	0.79	1	0.79
099	Other venereal diseases	36	5.67	1	0.79
117	Other mycoses	11	1.73	1	0.79
173	Other malignant neoplasm of skin	9	1.42	1	0.79
174	Malignant neoplasm of female breast	1	0.16	1	0.79
184	Malignant neoplasm of other and unspecified female genital organs	1	0.16	1	0.79
217	Benign neoplasm of breast	12	1.89	1	0.79
225	Benign neoplasm of brain and other parts of nervous system	6	0.94	1	0.79
302	Sexual and gender identity disorders	2	0.31	1	0.79
305	Nondependent abuse of drugs	4	0.63	1	0.79
307	Special symptoms or syndromes, not elsewhere classified	63	9.92	1	0.79
309	Adjustment reaction	12	1.89	1	0.79
345	Epilepsy and recurrent seizures	20	3.15	1	0.79
350	Trigeminal nerve disorders		0.00	1	0.79
365	Glaucoma	4	0.63	1	0.79
374	Other disorders of eyelids	3	0.47	1	0.79
380	Disorders of external ear	12	1.89	1	0.79
386	Vertiginous syndromes and other disorders of vestibular system	2	0.31	1	0.79
426	Conduction disorders	1	0.16	1	0.79
443	Other peripheral vascular disease	5	0.79	1	0.79
454	Varicose veins of lower extremities	9	1.42	1	0.79
494	Bronchiectasis		0.00	1	0.79
518	Other diseases of lung	1	0.16	1	0.79
525	Other diseases and conditions of the teeth and supporting structures	3	0.47	1	0.79
530	Diseases of esophagus	19	2.99	1	0.79
532	Duodenal ulcer	9	1.42	1	0.79

548	Special screening for malignant neoplasms of colon	8	1.26	1	0.79
569	Other disorders of intestine	3	0.47	1	0.79
610	Benign mammary dysplasias	7	1.10	1	0.79
634	Spontaneous abortion	5	0.79	1	0.79
640	Hemorrhage in early pregnancy	3	0.47	1	0.79
685	Pilonidal cyst	6	0.94	1	0.79
686	Other local infections of skin and subcutaneous tissue	4	0.63	1	0.79
696	Psoriasis and similar disorders	6	0.94	1	0.79
698	Pruritus and related conditions	6	0.94	1	0.79
704	Diseases of hair and hair follicles	17	2.68	1	0.79
708	Urticaria	9	1.42	1	0.79
730	Osteomyelitis, periostitis, and other infections involving bone	6	0.94	1	0.79
737	Curvature of spine	5	0.79	1	0.79
744	Congenital anomalies of ear, face, and neck	2	0.31	1	0.79
756	Other congenital musculoskeletal anomalies		0.00	1	0.79
790	Nonspecific findings on examination of blood	17	2.68	1	0.79
850	Concussion	50	7.87	1	0.79
898	Counseling for marital and partner problems, unspecified	10	1.57	1	0.79
903	Illegitimacy or illegitimate pregnancy		0.00	1	0.79
904	Social maladjustment	14	2.20	1	0.79
961	Need for prophylactic vaccination and inoculation against diptheria-tetanus-pertussis with poliomyelitis [DTP + polio]	1	0.16	1	0.79
989	Toxic effect of other substances, chiefly nonmedicinal as to source	6	0.94	1	0.79

* The description of the diagnosis code is taken from the UTOPIAN database; it does not necessarily match the description in the OHIP manual.

Codes with a frequency of zero in the NF1 cohort are not shown.

Most common service OHIP billing codes

Table 3a. Number of patients with at least one of the most common service codes from 2015-2019

Service code	Service code description	Matched controls		NF1 cases	
		N	% of cohort	N	% of cohort
A007	intermediate assessment	550	86.61	106	83.46
Q012	After-Hours Premium	235	37.01	52	40.94
Q200	per Patient Rostering Fee Code	195	30.71	47	37.01
A001	minor assessment	241	37.95	43	33.86
K131	periodic health visit-adult age 18-64 years	117	18.43	36	28.35
A888	Partial assessment	112	17.64	25	19.69
G590	administering a flu / influenza shot	119	18.74	21	16.54
E430	When Papanicolaou smear is performed outside of hospital, to G365,add	97	15.28	21	16.54
A003	Major assessment	103	16.22	21	16.54
G847	Diphtheria, Tetanus, acellular Pertussis (Tdap) – adult	99	15.59	19	14.96
K005	primary mental health care	67	10.55	17	13.39
K013	counselling-one or more people-per 1/2hr	103	16.22	17	13.39
G365	Gynaecology-Papanicolaou Smear-Periodic	96	15.12	16	12.60
Q011	Tacking code pap smears	83	13.07	14	11.02
Q150	Colorectal Cancer Screening Fee	64	10.08	13	10.24
K017	periodic health visit-child age 2-15 years	73	11.50	12	9.45
Q133	Colorectal Cancer Screening tracking code	40	6.30	12	9.45
G841	Diphtheria, Tetanus, acellular Pertussis, Inactivated Polio Virus, Haemophilus influenza type b (DTaP-IPV-Hib) - paediatric	50	7.87	11	8.66
Q131	Mammography tracking code	58	9.13	11	8.66
G538	Other immunizing agents not listed above	79	12.44	11	8.66
G010	Laboratory medicine diagnostic & therapeutic procedures - One or more parts of above without microscopy	64	10.08	10	7.87
G848	Varicella (VAR)	43	6.77	10	7.87
Q015	Newborn care fee	39	6.14	9	7.09

G700	Basic fee-per-visit premium	48	7.56	9	7.09
G846	Pneumococcal Conjugate	54	8.50	9	7.09
G845	Measles, Mumps, Rubella (MMR)	56	8.82	9	7.09
A002	enhanced 18-month well baby visit	33	5.20	9	7.09
K030	diabetic management fee	31	4.88	8	6.30
Q040	Diabetes Management Incentive	23	3.62	8	6.30
Q590	basic flu shot fee-per-visit premium FHN/FHO	44	6.93	8	6.30
K132	periodic health visit - adult 65 years of age and older	36	5.67	7	5.51
G489	Venipuncture - adolescent or adult	24	3.78	6	4.72
Q130	influenza vaccine tracking code	30	4.72	5	3.94
P004	Minor prenatal assessment	15	2.36	5	3.94
RNPE		21	3.31	5	3.94
E079	Smoking cessation	24	3.78	5	3.94
E080	First visit after hospital discharge premium	25	3.94	5	3.94
G014	Rapid Strep	35	5.51	5	3.94
A903	Pre-dental/pre-operative general assessment	30	4.72	5	3.94
G844	Meningococcal C Conjugate (Men-C)	30	4.72	5	3.94

Codes occurring for fewer than 5 patients with NF1 are not shown.

Table 3b. Number of records of the most common service codes from 2015-2019

Service code	Service code description	Matched controls		NF1 cases	
		N	Rate per 100 patients	N	Rate per 100 patients
A007	intermediate assessment	3691	581.2598	733	577.17
Q012	After-Hours Premium	539	84.88189	122	96.06
A001	minor assessment	555	87.40157	121	95.28
K030	diabetic management fee	223	35.11811	62	48.82
G271		55	8.661417	60	47.24
K131	periodic health visit-adult age 18-64 years	177	27.87402	52	40.94
Q200	per Patient Rostering Fee Code	221	34.80315	51	40.16
K005	primary mental health care	141	22.20472	49	38.58
G590		221	34.80315	46	36.22
A888	Partial assessment	204	32.12598	37	29.13
G010	Laboratory medicine diagnostic & therapeutic procedures - One or more parts of above without microscopy	130	20.47244	31	24.41
E430	When Papanicolaou smear is performed outside of hospital, to G365,add	131	20.62992	30	23.62
K013	counselling-one or more people-per 1/2hr	151	23.77953	28	22.05
Q015	Newborn care fee	166	26.14173	26	20.47
A003	Major assessment	145	22.83465	25	19.69
Q150	Colorectal Cancer Screening Fee	93	14.64567	23	18.11
G841	Diphtheria, Tetanus, acellular Pertussis, Inactivated Polio Virus, Haemophilus influenza type b (DTaP-IPV-Hib) - paediatric	129	20.31496	23	18.11
Q040	Diabetes Management Incentive	65	10.23622	23	18.11
G847	Diphtheria, Tetanus, acellular Pertussis (Tdap) - adult	105	16.53543	21	16.54
G365	Gynaecology-Papanicolaou Smear-Periodic	127	20	21	16.54
K017	periodic health visit-child age 2-15 years	120	18.89764	21	16.54
Q011	Tacking code pap smears	117	18.4252	19	14.96
Q131	Mammography tracking code	116	18.26772	17	13.39
G700	Basic fee-per-visit premium	92	14.48819	17	13.39

Q130	influenza vaccine tracking code	82	12.91339	17	13.39
Q133	Colorectal Cancer Screening tracking code	54	8.503937	15	11.81
Q590	basic flu shot fee-per-visit premium FHN/FHO	82	12.91339	15	11.81
P004	Minor prenatal assessment	100	15.74803	15	11.81
G846	Pneumococcal Conjugate	111	17.48031	14	11.02
K132	periodic health visit - adult 65 years of age and older	60	9.448819	14	11.02
G489	Venipuncture - adolescent or adult	79	12.44094	14	11.02
G538	Other immunizing agents not listed above	123	19.37008	12	9.45
RNPE		69	10.86614	12	9.45
G848	Varicella (VAR)	54	8.503937	11	8.66
G845	Measles, Mumps, Rubella (MMR)	61	9.606299	11	8.66
K007	psychotherapy	2	0.314961	10	7.87
A002	enhanced 18-month well baby visit	34	5.354331	9	7.09
G373	Tuberculosis (TB) skin test – sole reason for visit	48	7.559055	8	6.3
G372	Tuberculosis (TB) skin test – with visit	59	9.291339	8	6.3
_IHPPR007A		15	2.362205	8	6.3
E079	Smoking cessation	37	5.826772	7	5.51
E080	First visit after hospital discharge premium	27	4.251969	7	5.51
_IHPAC001A		22	3.464567	7	5.51
G014	Rapid Strep	43	6.771654	6	4.72
_IHPOU021A		4	0.629921	6	4.72
A903	Pre-dental/pre-operative general assessment	37	5.826772	5	3.94
G844	Meningococcal C Conjugate (Men-C)	31	4.88189	5	3.94
Q132	Childhood Immunizations tracking code	32	5.03937	5	3.94
Q142	Colorectal Cancer Screening exclusion code	45	7.086614	5	3.94
RNIM		6	0.944882	5	3.94
APS		1	0.15748	5	3.94

Q140	Pap smear exclusion code	14	2.204724	5	3.94
RN61		45	7.086614	5	3.94
_IHPHR010A		8	1.259843	5	3.94
RNWB		3	0.472441	5	3.94
_IHPHR001A		25	3.937008	5	3.94

Codes with a frequency of <5 in the NF1 cohort are not shown.

Development of algorithms to identify individuals with Neurofibromatosis type 1 within administrative data and electronic medical records in Ontario, Canada.

Carolina Barnett, MD, PhD[1,2,3], Elisa Candido, MPH[4], Branson Chen, MSc[4] , Priscila Pequeno[4], Patricia C. Parkin, MD, FRCPC[2,5], Karen Tu, MD, MSc[2,6,7,8]

1. Elisabeth Raab Neurofibromatosis Clinic, Toronto General Hospital, University of Toronto

2. Institute of Health Policy, Management and Evaluation, University of Toronto

3. Division of Neurology, Department of Medicine, University of Toronto

4. ICES

5. Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Canada

6. North York General Hospital, Toronto, Canada

7. Department of Family and Community Medicine, University of Toronto

8. Toronto Western Hospital Family Health Team, University Health Network, Toronto, Canada

Correspondence to: Carolina Barnett, MD, PhD

200 Elizabeth St. 5EC Room 334. Toronto, ON. M5G

2C4

Word count

Abstract: 350

Manuscript: 3357

Number of tables: 4

Abstract

Background: There is limited population-based data on Neurofibromatosis Type 1 (NF1) in North America. We aimed to develop and validate algorithms using administrative health data and electronic medical records (EMRs) to identify individuals with NF1 in Ontario, Canada.

Methods: We conducted an electronic free-text search of 15 commonly-used terms related to NF1 in the Electronic Medical Records Primary Care Database. Records were reviewed by two trained abstractors who classified them as confirmed, possible, and not NF1. An investigator with clinical expertise performed final NF1 classification. Patients were classified as confirmed if there was a documented diagnosis, meeting NIH criteria. Patients were classified as possible if 1) NF1 was recorded in the cumulative patient profile, but no clinical information to support the diagnosis; 2) only one criterion for diagnosis (e.g. child of confirmed case) but no further data to confirm or rule out. We tested different combinations of outpatient and inpatient billing codes, and applied a free-text search algorithm to identify NF1 cases in administrative data and EMRs, respectively.

Results: Of 273,440 eligible patients, 2,058 had one or more NF1 terms in their medical records. The terms "NF", "café-au-lait", or "sheath tumour" were constrained to appear in combination with another NF1 term. This resulted in 837 patients: 37 with possible and 71 with confirmed NF1. The population prevalence ranged from 1 in 3851 (confirmed NF1) to 1 in 2532 (possible and confirmed NF1). Billing code algorithms had poor performance, with overall low PPV (highest being 71%). The accuracy of the free-text EMR algorithm in identifying patients with NF1 was: sensitivity 85% (95% CI:74-92%), specificity 100% (95% CI:100-100%), positive predictive value 80% (95% CI:69-88%), negative predictive value 100% (95% CI:100-100%), and false positive rate 20% (95% CI:11-33%). Of false positives, 53% were possible NF1.

Conclusions: A free-text search algorithm within the EMR had high sensitivity, specificity and predictive values. Algorithms using billing codes had poor performance, likely due to the lack of

NF-specific codes for outpatient visits. While NF1 ICD-9 and 10 codes are used for hospital admissions, only ~30% of confirmed NF1 cases had a hospitalization associated with an NF1 code.

Keywords:

NF1, EMR, administrative database, algorithm

Introduction:

Neurofibromatosis Type 1 (NF1) is one of the most common autosomal dominant disorders, with minimum birth incidence estimated at 1 in 2500 births, and a prevalence of approximately 1 in 3,000 to 4,000 people. [1] NF1 is a multi-systemic disorder, affecting the skin in virtually all individuals. The most common manifestations of NF1 are cutaneous, including cafe-au-lait macules, axillary and inguinal freckling, and neurofibromas. [2] Cutaneous neurofibromas are peripheral nerve sheath tumors, which despite being benign can cause pain, discomfort, disfiguration, and are a major cause of anxiety and reduced quality of life. [2, 3] Up to 50% of individuals with NF1 have plexiform neurofibromas;[4] these, beyond causing disfigurement and pain, are associated with potential for transformation to malignant peripheral nerve sheath tumors (MPNST), which have a high mortality rate. [5] In addition to the risk of MPNST, individuals with NF1 have a higher risk of other malignancies, including optic glioma and brain tumors, breast cancer, gastrointestinal stromal tumors, and pheochromocytoma, among others. [2]

Population-based studies provide an opportunity to study how individuals use the health care system, and can provide insights into outcomes and association with other diseases. This is especially useful in rare diseases, such as NF1, where some outcomes may not be observable in small cohort studies. Some population-based studies have relied on diagnostic codes associated with billings, typically ICD-9 or ICD-10, to identify NF1 individuals.[6, 7] However, most studies have not validated case ascertainment using medical records, so there is a high risk of misclassification. For example, there may be clerical errors in coding, or sometimes a code may be used when there is a suspicion of a disease that is later ruled out. Therefore, it is important to develop validated algorithms to correctly identify individuals with the disease. Another approach to conduct population-based research in NF1 is creation of a registry, such as the Finnish registry, that was created through an extensive search through medical records in the country, to identify individuals with NF1. This registry is linked with several healthcare databases, and has provided rich insights into population-based outcomes of individuals with NF1. [1, 8]

The Canadian single-payer system is managed at the provincial level, meaning that each province has its own set of coverage rules and a specific health-insurance provider. In Ontario, Canada's most populous province, the Ontario Health Insurance Plan (OHIP) covers ≥95% of the more than 14 million residents of the province. In this large population, approximately 3 times the population of Finland, it is not feasible to identify all individuals with NF1 through extensive medical record search. However, it might be feasible to create a population-based cohort using health administrative data.[9, 10]

We aimed to develop and validate algorithms to identify individuals with NF1 living in Ontario, Canada, using administrative data, as well as in primary care electronic medical records (EMRs). Once developed, these algorithms can help to study population-based outcomes and healthcare utilization in this population. As secondary outcome, we aimed to study prevalence of NF1 in the province of Ontario, Canada.

Methods:

Data Source and Population:

We used the Electronic Medical Record Primary Care (EMRPC) database (formerly known as EMRALD) to identify individuals with NF1 for algorithm development. EMRPC is a database that consists of *all* clinically relevant information from family physician electronic medical records (EMRs). EMRPC contains data from almost 400 family physicians and ~ 500,000 patients distributed throughout Ontario. Affiliated physicians participate on a voluntary basis, using TELUS Practice Solutions® EMR software. EMRPC allows free text search in all available data, including a cumulative patient profile (CPP) which has a problem list, past medical history, family history, allergies, immunizations, and risk factors. EMRPC also includes family physician progress notes, specialist consultation reports, diagnostic tests, discharge summaries, laboratory tests and prescriptions. In general, the population enrolled in EMRPC is considered representative of the

entire Ontario population.[11]

Case determination:

We considered EMRPC patients eligible if they were rostered (i.e. patients registered with a participating family physician or family practice to provide their primary care), with a valid health insurance number, had a clinical encounter within the past 2 years, and were on the EMR for ≥ 1 year. We then conducted an automated free-text search in all the EMR data available, including the CPP, but also progress notes and specialist letters, for all eligible patients, to identify potential NF1 individuals. We used a liberal search for the following terms: "NF", "NF1", "neurofibromatosis", "neurofibroma", "café-au-lait", "Lisch", "plexiform", "axillary freckling", "inguinal freckling", "peripheral nerve sheath tumor", "malignant peripheral nerve sheath tumor", "MPNST", and "optic glioma". We examined the full patient chart from first entry, to the date of last data extraction, no later than March 31, 2016.

Two trained chart abstractors reviewed flagged charts to classify them as: "Confirmed NF1"—if there was a note from a specialty clinic or genetic test result confirming the diagnosis—"possible NF1" if the abstractors were uncertain of the diagnosis, "Not NF1"— if NF1 was clearly ruled out by a specialist, and "blank" if the chart made absolutely no mention to NF1. We tested intra-rater reliability of the chart abstractors, whereby 5% of the reviewed charts were selected at random, and were presented again to the original abstractors, blinded to their initial classification. To test inter-rater reliability, another random sample of 5% of the charts that were reviewed by one abstractor was then presented to the other abstractor and vice versa.

To validate the classification, one investigator with expert clinical knowledge of NF1 (CB) reviewed all charts initially classified as "confirmed" and "possible" and reclassified as confirmed, possible and not NF1. Cases were classified as confirmed if there was a diagnosis of NF1 based on 1988 NIH criteria,[12] or if there was positive genetic testing with consistent clinical manifestations. Cases were classified as possible if there was a suspicion of NF1, but insufficient data on the charts

to rule it in or out. For example, young children with ≥6 cafe-au-lait spots who are too young to have other NF1 manifestations; or children of an individual with confirmed NF1, but without specific assessments in the EMR to assess clinical NF1 manifestations. Cases were also classified as possible if there was a diagnosis of NF1 in the CPP but without any clinical records describing NF1 specific manifestations or concerns, to confirm the diagnosis. All other records were classified as not NF1.

The patients classified as confirmed NF1 were considered the true positives for the algorithms; all other patients in the EMRPC database were considered non-NF1.

Algorithm development:

Physicians submit claims to OHIP for each clinical encounter either through fee-for-service or shadow billing; these are submitted with a corresponding diagnostic code. Each person insured by OHIP has a unique 10-digit health insurance number which is encrypted and used for linkage to different ICES databases. This includes a variety of administrative and other health-related data, including OHIP data for physician services (from 1991), and hospital services such as hospitalizations, emergency department visits and day surgery. These databases are linked together using unique encoded identifiers and analyzed at ICES.

We surveyed health providers who routinely treat patients with NF1, to develop a list of commonly used OHIP billing codes and procedures. These are the codes used for billing outpatient services. For hospital services we used the Canadian Institute for Health Information (CIHI) database that contains discharge diagnosis on the discharge abstract database (DAD); we also used diagnostic codes within the same day surgery (SDS) data from the National Ambulatory Care Reporting System (NACRS) database. These databases use ICD-9 codes for services before 2002 and ICD-10 thereafter. After identifying documented NF1 cases in EMRPC through chart abstraction, we assessed all outpatient and inpatient billing codes in the patients with confirmed NF1. For the

algorithm testing, we chose billing codes that had previously been identified by specialists treating NF1 individuals, and that also were present in the list of highly used billing codes for the confirmed NF1 cases.

We tested different combinations of the selected outpatient (OHIP) and hospital-related diagnostic codes (CIHI, NACRS and SDS databases), with different timeframes (e.g. ever billed or billed ≥1, 2 or 3 times in 1,2 and 3 years). We also included algorithms that specified the specialties associated with the outpatient codes, focusing on specialties that frequently care for individuals with NF1 in the province (e.g. neurology, neurosurgery, paediatrics, plastic surgery, dermatology).

We also tested the performance of a free-text search on the EMR to accurately identify individuals with NF1, using NF1-related terms in the problem list and past medical history sections of the CPP. For all algorithms we calculated sensitivity, specificity, positive and negative predictive values (PPV and NPV) with 95% confidence intervals (CI).

Sample size:

We used nomograms for sample size estimation for diagnostic accuracy. With a predicted prevalence of 1 in 4,000 (0.00025), for 90% sensitivity and specificity, and accuracy (half-width of the confidence interval) of 0.05, a minimum of 10,000 individuals (cases and non-cases) were needed. [13]

Ethics:

ICES is a prescribed entity under Ontario's Personal Health Information Protection Act (PHIPA). Section 45 of PHIPA authorizes ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management of, evaluation or monitoring of, the allocation of resources to or planning for all or part of the health system. Projects that use data collected by ICES under section 45 of PHIPA, and use no other data, are exempt from REB review. At ICES, all datasets are encoded in a process that removes direct

personal identifiers to ensure patient confidentiality. The use of the data in this project is authorized under section 45 and approved by ICES' Privacy and Legal Office.

Results

Chart abstraction:

At the time of the search (January 2019), EMRPC database had 273,440 patients that met eligibility criteria. The mean age was 43.9 ± 23 years and 55.8% were female. Mean time on the EMR was 6.45±3.5 years. On the initial search, 2,058 records had at least one search term. While pilot-testing the abstraction platform, we found that the terms: "NF", "Cafe-au-lait" and "sheath tumour" when found alone, were not related to NF1. For example, "NF" usually referred to the Canadian province of Newfoundland or the verification code of the health card. Therefore, we modified the search strategy so that these terms had to be found with at least another search term to flag the record. This modified search resulted in 837 patients whose charts were fully abstracted. The trained abstractors marked 42 patients as confirmed NF1, 228 as possible (confirmed and possible n=270) and 567 as "no NF1" or "blank" (no mention of NF1). Figure 1 depicts the chart abstraction and validation process. The inter and intra-rater agreement between chart abstractors was high, between 96% and 100%.

Validation of the 270 records flagged as confirmed and possible yielded 71 confirmed NF1 cases and 37 possible NF1 cases (confirmed and possible n=108); the remaining 162 were re-classified as non NF1 for a total of 729 non-NF1 individuals. To assess the validity of the modified search strategy, we also abstracted 300 random charts, 100 for each of the terms we excluded when present alone ('Cafe-au-lait", "NF", "sheath tumour"). After validation, none of these records corresponded to NF1 cases. The minimum prevalence of NF1 in this population was 1 in 3,851 (95% CI: 1 in 3,107 to 1 in 4,921), considering the confirmed cases only, and 1 in 2,532 (95% CI: 1 in 2,183 to 1 in 3,038) considering both confirmed and possible cases.

Diagnostic algorithms:

From the 71 confirmed NF1 cases, we identified the following OHIP billing codes (outpatient services) that were also in the billing list from specialists: 192—"malignant tumours of the cranial nerves or spinal cord", 216—"skin lesion: e.g. dermatofibroma", 225—" benign tumours of the brain, spinal cord or peripheral nerves", 709—"other disorders of the skin and subcutaneous tissue" and 758—"chromosomal abnormalities". ICD-9 codes were 237.7: "NF unspecified" and 237.71: "NF1". The ICD-10 code for NF is Q85.0. Only 22 of the 71 (31%) patients with confirmed NF1 had a hospital-related encounter (hospitalization, emergency department visit or same day surgery) with an NF1 diagnostic code (ICD-9 or ICD-10).

We tested several administrative algorithms in a reference cohort of 273,402 eligible individuals within EMRPC, of which 71 were confirmed NF1. The demographic characteristics of this cohort are in table 1. Overall, these algorithms had poor performance; the highest PPV was 71%, but sensitivity was only 21%, with 79% false negatives. Table 2 depicts the most relevant algorithms tested and their performance. We then assessed the performance of a free-text search strategy within the EMR to identify the true cases within the initial eligible sample of 273,440 records. Patients were considered positive if those fields had mention of "NF1", "Recklinghausen", or "neurofibromatosis" (and various misspellings), unless the disease term was in close proximity to an exclusion word/phrase (e.g. "no sign of", "screen for", "not", "unconfirmed", "NF2"). Specific search terms of this algorithm are summarized in Table 3. This algorithm resulted in a sensitivity 85% (95% CI:74-92%), specificity 100% (95% CI:100-100%), positive predictive value (PPV) 80% (95% CI:69-88%), negative predictive value 100% (95% CI:100-100%), and a false positive rate of 20% (95% CI:11-33%). Of the 15 false positives, 8 (53%) were possible NF1 cases.

Assessing billing patterns on EMR databases

From the confirmed NF1 records identified in EMRPC, we identified the most commonly used outpatient diagnostic codes, as seen in Table 4. Approximately 35% of outpatient encounters did not have a specific diagnostic code. The most common diagnostic code was code 304: drug dependence/drug addiction, which was billed 1216 times (4.3% of clinical encounters). When

combining all codes related to skin diagnoses, these were billed 822 times (2.9% of clinical encounters), The remaining diagnostic codes were each billed in 2% or less encounters.

Discussion:

In this study, we identified records of individuals with confirmed NF1 within a large, population-based database of EMRs. We found that a free-text search in the cumulative patient profile of primary care EMRs is a simple, yet effective way of identifying individuals with NF1. Because NF1 is a multisystemic disease where patients are often seen by different specialties, being able to identify people with NF1 from primary care records may provide more complete information than records from single specialties. Additionally, primary care records may be more reflective of the general population compared to a single specialty centre or clinic. This simple search algorithm can be used in the future to study other outcomes in other databases, so there is potential for generalizability; however, this algorithm should be validated in other EMRs before use, as the way physicians document the diagnosis of NF1 may differ in different healthcare settings. Because the uptake of EMRs is increasing, we believe that in the future this algorithm can be adapted to different settings to identify larger numbers of individuals living with NF1.

Our administrative data algorithms using physician billing codes performed poorly and were not able to accurately identify individuals with NF1. There are several reasons for this. First, there is no specific NF1 diagnostic code in OHIP for outpatient visits. Therefore, clinicians use codes that apply to their specialty practices, such as codes for skin lesions in the case of dermatology/plastic surgery, brain tumours in the case of neurosurgery, or codes for general genetic diseases for other specialties. We had hypothesized that a combination of these codes as well as the specific medical specialties could help identify individuals with NF1. For example, we thought that NF1 patients would have codes for specific skin lesions in combination with codes for genetic diseases, plus visits by certain specialties (e.g. neurology/neurosurgery). However, algorithms that followed this logic resulted in high numbers of false positives. The best performing algorithms were those related to hospital services that use ICD codes. However, only 22 out of the 71 confirmed cases had

a hospital-related billing (including same day surgery) associated with an NF1 code. It is not clear if this reflects few NF1-related hospital services in this cohort, or lack of awareness of NF1 by healthcare providers, whereby NF1 was not identified as a main health problem for a given admission.

Our estimate of minimum prevalence in this population is within published ranges; however, this is likely an underestimate of the true prevalence. For example, if the 37 probable cases were confirmed, the prevalence would increase from 1 in 3,851 to 1 in 2,532. Additionally, in our study we only identified individuals who had a diagnosis of NF1, but previous work has shown that with detailed screening programs on NF1 manifestations, the prevalence can be as high as 1 in 1,000, identifying many individuals without a previous diagnosis.[14, 15] This may be due to some individuals with very mild clinical features who may not receive a formal diagnosis of NF1.

A surprising finding from this study was that the diagnostic billing code for drug addiction/dependency was the most commonly used outpatient code in the 71 confirmed cases, accounting for 4.3% of clinical encounters. As a comparison, in previous studies from the general population of Ontario, drug addiction was not in the top 10 diagnostic codes.[16] Further work is needed to assess the prevalence of substance dependence in larger population-based studies of individuals with NF1, to determine if this finding is replicated in other settings. Because cutaneous manifestations of NF1 are highly prevalent in NF1 (>90% of individuals), and given the absence of specific NF1 diagnostic codes, we combined the 3 most commonly used skin-related outpatient diagnostic codes used in confirmed NF1 cases. With this approach, these became the second most common outpatient diagnostic code accounting for 2.9% of all billings. Further work is needed to understand the healthcare utilization patterns of individuals living with NF1.

This study has some limitations. Our findings are biased towards individuals with NF1 who actively use the healthcare system, so we will have missed those with mild disease who do not seek care, as well as very young children without clinical manifestations, and also those who have limited access

to healthcare due to social or geographical reasons. Our chart abstraction strategy was comprehensive, but it is possible that some individuals with clinically evident NF1 who do use the healthcare system, do not have their NF1 clinical manifestations and/or diagnosis recorded in the medical records. These individuals would not have been flagged in our search of the EMR. Despite these limitations, we found that our EMR search strategy performs well to identify individuals with a clinical diagnosis of NF1.

The generalizability of this strategy needs to be assessed and validated in other settings, for example in databases that use a different EMR vendor. The use of a free-text search makes generalizability more likely across vendors, because it can identify the search terms regardless of where they are stored in the EMR or how the EMR is organized. However, this strategy will still need to be validated in a different EMR. It is also possible that some clinicians use other terms or abbreviations (e.g CALMs for cafe-au-lait macules) that may need to be incorporated in the search terms when validating this search strategy in other settings. Other limitations to generalizability include language, as our search strategy was done in English, therefore language and cultural-specific search terms will be needed in different regions. Finally, we do acknowledge that electronic medical records will be limited or non-existent in low income regions, where this approach may not be feasible. However, we believe that any effort to identify cohorts of patients with rare diseases, such as NF1, where clinical questions can be answered, will eventually be useful to other patients, including those without access to electronic medical records.

In summary, we found that a simple EMR search is highly accurate in identifying individuals with NF1 from EMRs, and can be used for future studies. Administrative data algorithms within Ontario had poor performance therefore province-wide identification of NF1 using administrative data is not possible. However, linkage of registry data or EMR data to administrative databases can still help study health utilization and outcomes of individuals with NF1 in the province, where our estimate of NF prevalence is in keeping with published literature.

Declarations:

1. **Ethics and consent:** ICES is a prescribed entity under Ontario's Personal Health Information Protection Act (PHIPA). Section 45 of PHIPA authorizes ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management of, evaluation or monitoring of, the allocation of resources to or planning for all or part of the health system. Projects that use data collected by ICES under section 45 of PHIPA, and use no other data, are exempt from REB review. The use of the data in this project is authorized under section 45 and approved by ICES' Privacy and Legal Office.

2. **Consent for publication:** All authors agree with the final version and consent for publication.

3. **Availability of data and materials:** The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet pre-specified criteria for confidential access, available at www.ices.on.ca/DAS (email: das@ices.on.ca). The full dataset creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification.

4. **Funding:** This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This study also received funding from the US Department of Defense, award number: W81XWH-19-1-0177, NF180027. The analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended

or should be inferred.

5. Competing interests:

- C. Barnett has received grant support from US Department of Defence, MGNet, Muscular Dystrophy Canada, Grifols and Octapharma. She has been a member of advisory board and consultant for Alexion, Sanofi and Argenx.
- E. Candido reports no competing interests.
- B. Chen reports no competing interests.
- P. Pequeno reports no competing interests.
- P. Parkin has received grant support from the Hospital for Sick Children Foundation and Canadian Institutes of Health Research.
- K. Tu reports no competing interests.

6. Contributions:

- C. Barnett designed and conceptualized the study, performed chart validation, data analysis and wrote the manuscript.
- E. Candido helped with study design, data analysis and edited the manuscript for content.
- B. Chen conducted the EMR search, identified charts for abstraction, and developed the EMR algorithm. He edited the manuscript for content.
- P. Pequeno conducted the billing algorithm analyses; she edited the manuscript for content.
- P. Parkin helped with study design and edited the manuscript for content.
- K. Tu helped with study design, data analysis and edited the manuscript for content.

7. Acknowledgements: Not applicable

References:

1. Uusitalo E, Leppävirta J, Koffert A et al. Incidence and Mortality of Neurofibromatosis: A Total Population Study in Finland. Journal of Investigative Dermatology. 2015;135:904-906.

2. Ferner RE, Huson SM, Thomas N et al. Guidelines for the diagnosis and management of individuals with neurofibromatosis 1. Journal of Medical Genetics. 2006;44:81-88.

3. Wolkenstein P, Zeller J, Revuz J, Ecosse E, Leplège A. Quality-of-life impairment in neurofibromatosis type 1: a cross-sectional study of 128 cases. Archives of Dermatology. 2001;137:1421-1425.

4. Mautner VF, Asuagbor FA, Dombi E et al. Assessment of benign tumor burden by whole-body MRI in patients with neurofibromatosis 1. Neuro Oncol. 2008;10:593-598.

5. McCaughan JA, Holloway SM, Davidson R, Lam WW. Further evidence of the increased risk for malignant peripheral nerve sheath tumour from a Scottish cohort of patients with neurofibromatosis type 1.[letter]. J Med Genet 2007;44(7):463-466.

6. Seminog OO, Goldacre MJ. Risk of benign tumours of nervous system, and of malignant neoplasms, in people with neurofibromatosis: population-based record-linkage study. British journal of cancer. 2012;108:193-198.

7. Terry AR, Barker FG, Leffert L, Bateman BT, Souter I, Plotkin SR. Neurofibromatosis type 1 and pregnancy complications: a population-based study. American journal of obstetrics and gynecology. 2013;209:46.e1-8.

8. Pöyhönen M, Kytola S, Leisti J. Epidemiology of neurofibromatosis type 1 (NF1) in northern Finland. Journal of Medical Genetics. 2000;37:632-636.

9. Breiner A, Young J, Green D et al. Canadian Administrative Health Data Can Identify Patients with Myasthenia Gravis. Neuroepidemiology. 2015;44:108-113.

10. widdifield J, Bernatsky S, Paterson JM et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: A validation study using the medical records of rheumatologists. Arthritis Care & Research. 2013n/a-n/a.

11. Tu K, Widdifield J, Young J et al. Are family physicians comprehensively using electronic medical records such that the data can be used for secondary purposes? A Canadian perspective. BMC Med Inform Decis Mak. 2015;15:67.

12. Stumpf D. Consensus development conference of neurofibromatosis. Arch Neurol. 1988;45:575-578.

13. Carley S. Simple nomograms to calculate sample size in diagnostic studies. Emergency Medicine Journal. 2005;22:180-181.

14. Garty BZ, Laor A, Danon YL. Neurofibromatosis type 1 in Israel: survey of young adults. Journal of Medical Genetics. 1994;31:853-857.

15. Orraca M, Morejón G, Cabrera N, Menéndez R, Orraca O. Neurofibromatosis 1 prevalence in children aged 9-11 years, Pinar del Río Province, Cuba. MEDICC review. 2014;16:22-26.

16. Tu K SS, Kidd MR, Grunfeld E, Ji C, et al. The University of Toronto Family Medicine Report: Caring for our Diverse Populations. Toronto: Department of Family and Community Medicine.; 2020