ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada

Eric I. Benchimol[a,b,c,d,*], Astrid Guttmann[a,e,f], David R. Mack[b,c], Geoffrey C. Nguyen[a,g,h], John K. Marshall[i], James C. Gregor[j], Jenna Wong[a], Alan J. Forster[a,k,l], Douglas G. Manuel[a,d,l,m]

[a]*Institute for Clinical Evaluative Sciences, 2075 Bayview Avenue, Toronto, Ontario, Canada, M4N 3M5*
[b]*CHEO Inflammatory Bowel Disease Centre, Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Eastern Ontario, 401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1*
[c]*Department of Pediatrics, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5*
[d]*Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5*
[e]*Department of Paediatrics, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada, M5S 1A8*
[f]*Institute of Health Policy, Management and Evaluation, 155 College Street, University of Toronto, Toronto, Ontario, Canada, M5T 3M7*
[g]*Department of Medicine, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada, M5S 1A8*
[h]*Centre for Inflammatory Bowel Disease, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada, M5G 1X5*
[i]*Department of Medicine, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1*
[j]*Department of Medicine, London Health Sciences Centre, University of Western Ontario, 339 Windermere Road, London, Ontario, Canada, N6G 2V4*
[k]*Department of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5*
[l]*Ottawa Hospital Research Institute, 725 Parkdale Ave., Ottawa, Ontario, Canada, K1Y 4E9*
[m]*Department of Family Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5*

Accepted 28 February 2014; Published online xxxx

Abstract

Objective: Health administrative databases can be used to track disease incidence, outcomes, and care quality. Case validation is necessary to ensure accurate disease ascertainment using these databases. In this study, we aimed to validate adult-onset inflammatory bowel disease (IBD) identification algorithms.

Study Design and Setting: We used two large cohorts of incident patients from Ontario, Canada to validate algorithms. We linked information extracted from charts to health administrative data and compared the accuracy of various algorithms. In addition, we validated an algorithm to distinguish patients with Crohn's from those with ulcerative colitis and assessed the adequate look-back period to distinguish incident from prevalent cases.

Results: Over 5,000 algorithms were tested. The most accurate algorithm to identify patients 18 to 64 years at diagnosis was five physician contacts or hospitalizations within 4 years (sensitivity, 76.8%; specificity, 96.2%; positive predictive value (PPV), 81.4%; negative predictive value (NPV), 95.0%). In patients ≥65 years at diagnosis, adding a pharmacy claim for an IBD-related medication improved accuracy.

Conclusion: Patients with adult-onset incident IBD can be accurately identified from within health administrative data. The validated algorithms will be applied to administrative data to expand the Ontario Crohn's and Colitis Cohort to all patients with IBD in the province of Ontario.

Keywords: Inflammatory bowel disease; Crohn's; Ulcerative colitis; Epidemiology; Health administrative data; Routinely collected health data; Validation

1. Introduction

Inflammatory bowel disease (IBD) is rising worldwide, with particularly high incidence in developed nations [1]. Canada has among the highest incidence of IBD in the world [1,2]. Ontario is Canada's most populous province, and its single-payer health system entitles legal residents to universal access to all health-care services. Ontario's health administrative databases are a large repository of all health-care encounters for every legal resident, and these data have been used to develop surveillance programs for multiple chronic diseases including diabetes [3,4] and asthma [5]. These data represent a unique opportunity to conduct population-based surveillance of patients with

Conflict of interest: The authors have no conflicts of interest to disclose.

* Corresponding author. Tel.: +1-416-737-7600x1516; fax: +1-416-738-4854.
E-mail address: ebenchimol@cheo.on.ca (E.I. Benchimol).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.02.019

What is new?

- Algorithms to identify adults and elderly patients with inflammatory bowel disease (IBD) from within Ontario health administrative data have been validated, including algorithms to classify IBD subtype and to determine the look-back period required to distinguish incident from prevalent cases.
- Although some previously reported international IBD identification algorithms (such as the Manitoba algorithm) function well in Ontario, further refinement has improved accuracy of classification of incident cases of IBD.
- Algorithm validation should be an ongoing process. Validation, improvements, and refinement are essential when applying algorithms to new administrative databases, jurisdictions, time periods, and patient age groups.

IBD within a large jurisdiction. Administrative database–derived cohorts have been used to assess epidemiology, health services use, and outcomes in IBD in other jurisdictions as well [6–12]. Critical to the accuracy of such research, however, is the ability to accurately identify individuals with IBD using a rigorously validated algorithm comprising the best combination of health administrative data codes [13].

The Ontario Crohn's and Colitis Cohort (OCCC), derived from health administrative data, is the largest ongoing population-based surveillance cohort of pediatric-onset IBD in the world, comprising all children living with IBD in Ontario, Canada [14]. An identification algorithm validated specifically in the pediatric age group was used to identify cases, and the cohort demonstrated exceptionally high incidence of pediatric IBD in Ontario [14]. Additionally, a study from the Canadian IBD Epidemiology database used health administrative data to report on Canada's high incidence in five provinces [15]. This report excluded Ontario and Quebec, Canada's most populous provinces and used an identification algorithm validated in Manitoba only. IBD identification algorithms have been validated for other jurisdictions [16–19], each different from the Manitoba and Ontario algorithms. This indicated the importance of validation in the jurisdiction of the administrative data to which the algorithm will be applied. In addition, the validity of these algorithms in separate age groups (such as patients with adult-onset or elderly-onset IBD) was not conducted, nor was re-validation in multiple cohorts, all identified as important aspects of algorithm validation by a recent systematic review [13].

The aims of this study were to (1) develop an algorithm to identify individuals with incident adult-onset IBD using data from charts in Ottawa, Ontario, Canada; (2) validate and apply the case identification algorithm to the entire Ontario adult population to expand the OCCC to include adult patients with IBD. We also assessed the accuracy of other internationally validated algorithms in two distinct Ontario cohorts using chart review as the reference standard.

2. Methods

2.1. Ethical issues

This study was approved by the research ethics boards of the Children's Hospital of Eastern Ontario, The Ottawa Hospital, Hamilton Health Sciences, the London Health Sciences Centre, Mount Sinai Hospital (Toronto), and North York General Hospital. Ontario health administrative data is housed at the Institute for Clinical Evaluative Sciences (ICES; Toronto, Ontario, Canada) designated a prescribed entity under Section 45 of Ontario's Personal Health Information Protection Act. Under this Section, prescribed entities are permitted to share personal health information with other prescribed entities (such as hospitals) or linked across databases without obtaining informed consent [20]. Privacy of personal health data is regulated by the Information and Privacy Commissioner of Ontario, and approval of ICES activities were most recently reviewed in 2011 [21]. All projects linking to or using ICES data are evaluated and approved by the ICES privacy officer.

2.2. Administrative data sources

The databases used in this study included hospital discharge abstract data (DAD) mandatorily collected from all hospitals and reported to the Canadian Institute for Health Information (CIHI-DAD), billing claims for all physician services provided from the Ontario Health Insurance Plan (OHIP), and the Registered Persons Database (demographic data including region of residence). Hospital data before 2002 and all physician billing claims have diagnoses associated using codes from the International Classification of Disease (ICD)-9 [22]. Hospitalizations after 2002 used ICD-10 codes [23]. The Ontario Drug Benefits database was used to obtain complete prescription information for patients $\geq$65 years at diagnosis. Prescription records for patients $<$65 years were only available for those on social assistance, and therefore, was not used for this study.

The Ottawa Hospital is a multi-campus medicare facility located in Ottawa, Ontario, Canada and serves as the largest adult referral center in Ontario for a population of 1.1 million. The Ottawa Hospital Data Warehouse (OHDW) was used to generate a reference standard for the algorithm development sample. This database contains administrative data on hospitalization, outpatient visits, emergency