eAppendix

TITLE

Time-related bias in administrative health database studies of disease incidence

OBJECTIVE

To illustrate the bias resulting from temporal aspects of administrative health databases and to quantify the impact of such bias on inflammatory bowel disease (IBD) incidence estimates in Québec, Canada.

METHODS

Study design and source population

The present population-based study was conducted using the administrative health data of the Régie de l'Assurance Maladie du Québec (RAMQ). We retrieved information regarding all Québec residents with at least one physician claim or hospitalization with an IBD diagnostic code during the 1996 to 2009 period. Data regarding medical and pharmaceutical services as well as hospitalization discharge summaries were obtained for all subjects. The International Classification of Diseases (ICD) versions 9 and 10 were used for both Crohn's disease (CD) (ICD 9 555.xx, ICD 10 K50.xx) and ulcerative colitis (UC) (ICD 9 556.xx, ICD 10 K51.xx) to assess IBD occurrence.

Case identification

IBD cases were identified using three different case definitions to illustrate the impact of time-related bias. We used case definitions requiring increasing numbers of diagnostic codes for either

CD or UC, namely two, four and six IBD contacts occurring on distinct dates to classify a subject as an IBD case.

Data analysis

IBD incidence rates were computed using as denominator the population of Québec eligible for the provincial health plan, available from the RAMQ website[1]. The rates were standardized by age and sex using the Statistics Canada 2001 Québec Census population[2].

Assessment of bias

The factors that can lead to biased incidence estimates are the time and number of physician visits required to fulfill the diagnosis criteria, the timing of disease diagnosis and the time span of the database observation period.

To illustrate the impact of the time needed to fulfill the case definition criteria we ran two separate analyses:

a) allowing the criteria to be fulfilled at any time during the observation period, without any restrictions, such as for Subjects 1 and 3 in the diagram (see eFigure 1, diagram of different scenarios that can be encountered when administrative databases are used to assess disease incidence);
b) requiring that all criteria are fulfilled within a specified two-year long interval (cluster), as for Subjects 2,4 and 5 in the diagram (eFigure 1).

In each of these two analyses we illustrated the impact of the remaining factors that may induce bias:

- the timing of disease diagnosis – considered to occur at the first IBD contact in the series of visits required by the case definition (Subjects 1 and 2) or at the case-defining IBD contact (Subjects 3,4 and 5);

- the number of physician visits required to fulfill the criteria – by using three different case definitions with varying number of IBD contacts required;
- the span of the database observation period – IBD incidence was calculated first using 1996-2004 as observation period (9 years) and then using the full 1996-2009 dataset as observation period (14 years) to show the impact of adding five more years of data on incidence estimates.

When case diagnosis is considered at the case-defining contact, all other IBD contacts required to meet the case definition criteria have to occur prior to the date of diagnosis. Therefore, we chose to compute incidence rates for 2001, to allow sufficient time from the beginning of the study period (1996) to fulfill the criteria. We also computed IBD incidence for 2002 through 2006 to show how the value of the estimates may vary in time (see eFigure 2 and eFigure 3). No rates were computed after 2006 to allow sufficient time to fulfill case definition criteria prior to the end of the study period (2009).

To differentiate between incident and prevalent cases, we computed IBD incidence rates using a mandatory two year disease-free period prior to the first IBD contact in the dataset or cluster, a period in which a subject was covered by the health plan but did not have any IBD contacts.

RESULTS

A total of 240,937 Quebec residents had at least one diagnosis code of IBD during 1996-2009. The overall number of IBD cases identified using the three different case definitions varied between 40,537 and 85,313, depending on how restrictive the case definitions were in terms of number of contacts required to meet the criteria.

Quantification of bias

a) No specified time period to meet criteria

The diagram in eFigure 1 shows that subjects 1 and 2 would be counted as incident in 2001 if IBD diagnosis is considered to occur at the first IBD contact in the series of visits required by the case definition. If, on the contrary, IBD diagnosis is timed at the case-defining contact, when all criteria were fulfilled, subjects 3, 4 and 5 would be considered as incident in 2001. In the Table we showed the impact of bias from timing of disease diagnosis and span of observation period on the standardized incidence rate for IBD. Timing of disease diagnosis caused biases of up to 53% depending on the time span of observation period and the number of visits (IBD contacts) required by the case definition. The difference in estimates due to timing of IBD diagnosis was relatively constant throughout 2001-2006 for all three case definitions used (see eFigure 2).
As illustrated in the Table, the span of the observation period caused biases of up to 21% in incidence estimates when IBD diagnosis was timed at the first IBD contact in the series of visits required by the case definition. The diagram (eFigure 1) shows that only subject 2 would be counted as incident in 2001 if the observation period was 9 years long (1996-2004), but both subject 1 and 2 would be counted as incident in 2001 if the longer observation period was used (1996-2009). The span of observation period had no impact on disease occurrence when the disease diagnosis was timed at the case-defining IBD contact (Table). Subjects 3, 4 and 5 would be counted as incident in 2001 regardless of the span of observation period.

b) Specified time period to meet criteria

The timing of disease diagnosis has an impact on disease occurrence even if the case definition criteria need to be fulfilled in a specified amount of time. If the criteria are to be fulfilled in 2 years, subject 1 in the diagram (eFigure 1) would be counted as incident in 2001 when IBD

diagnosis is considered at the first IBD contact, whereas subjects 4 and 5 would be considered as incident in 2001 when IBD diagnosis is timed at the case-defining contact.

The bias in IBD incidence resulting from timing of IBD diagnosis was up to 5% when IBD was considered as diagnosed at the first IBD contact in the specified two-year cluster period (see eTable 1). We illustrated the impact of bias resulting from timing of IBD diagnosis for the 2001-2006 period (see eFigure 3). The difference in estimates decreased in time and during 2004-2006 IBD incidence calculated using the first IBD contact in the cluster was slightly higher than when using the case-defining contact.

When using a two-year cluster period to fulfill the case definition criteria the span of database observation period had no impact on IBD incidence estimates, regardless of the case definition used or the timing of IBD diagnosis (see eTable 1).

Special considerations related to differentiation of incident and prevalent cases

When no information is available prior to the first contact of the disease of interest or when the first contact occurs a short period of time after the start of coverage in the health plan, such as the case of Subjects 3 and 5 in the diagram, it is difficult to assess if the cases are incident or prevalent. A mandatory period of disease free coverage may help reduce overestimation of incident rates.

We assessed the bias in IBD estimates resulting from the span of database observation period when imposing a two-year mandatory IBD free period of enrollment in the health plan prior to the first IBD contact in the dataset or in the two-year specified cluster period to fulfill the case definition criteria (see eTable 2 and eTable 3). Even though overall IBD incidence was lower given that fewer subjects satisfied the additional requirement, the impact of bias resulting from span of database observation period was between 9 and 21% when IBD diagnosis was timed at

the first IBD contact in the dataset (eTable 2), the same as that observed without implementing the two-year IBD free period (Table). There was no change in IBD incidence due to the span of database observation period when the timing of IBD diagnosis was considered at the case-defining contact (eTable 2) or when using a two-year cluster requirement to fulfill the case definition criteria (eTable 3).

DISCUSSION

In this population-based retrospective cohort study we illustrated the bias due to temporal aspects of administrative health databases and quantified its impact on estimates of IBD occurrence using the provincial health databases of the RAMQ. Even though IBD was used as outcome for the purpose of this study, time-related bias may occur during the data analysis process in any study using administrative databases to assess disease occurrence, as long as there is a requirement that a certain diagnosis code appears at least twice in the dataset before a subject can be classified as a case. The factors that lead to bias are the time and number of physician visits needed to fulfill the criteria, the timing of disease diagnosis and the span of the database observation period. Such factors are strictly related to the methodological approach of assessing disease occurrence once a case definition was chosen and validated in the source population. The magnitude of the impact on estimates of disease occurrence may vary from dataset to dataset and increases with the number of physician visits included in the criteria.

We have shown that the timing of disease diagnosis caused biases in IBD incidence of up to 53% depending on the number of IBD contacts required by the case definition and the span of observation period.

The number of cases identified as incident in any given year using a small observation period may be very different from that resulted from a larger dataset because more subjects have the opportunity to meet the number of contacts required. We found that bias associated with the span of observation period varied from 9 to 22% depending on the case definition.

When the date of diagnosis is considered at the time a subject meets the criteria for a case definition, subsequent contacts would not change this date of diagnosis; the span of observation period could not impact estimates of disease occurrence in this case. Indeed, we found that, when IBD diagnosis was timed at the case-defining contact, there was no impact on incidence estimates resulting from the span of observation period, regardless of the number of physician visits required by the case definition. This approach differs from the clinical perspective, when the case diagnosis is considered at the earliest time a diagnosis was made, and will result in a shift of the "true" incidence date by a few years. However, a subject cannot be considered an incident case prior to meeting all criteria for being classified as a case. The time it takes to fulfill the criteria needs to be considered as unexposed[3].

Timing the case diagnosis at the case-defining contact also avoids inducing immortal time bias in prognostic studies using cohorts identified with the use of administrative databases[4]. The time needed to fulfill the criteria can be considered as immortal, as every subject classified as a case, by definition, has survived long enough to meet the criteria. This topic has been thoroughly characterised in the context of pharmacoepidemiology[4-6].

Bias from timing of disease diagnosis was reduced to 5% when the criteria were required to be fulfilled in a two-year cluster period. Furthermore, there was no bias resulting from the span of observation period regardless of the case definition.

When working with administrative databases it can be difficult to differentiate incident from prevalent cases, especially at the beginning of the study period, given that information prior to the start of the study period is not available. Therefore the use of a disease free period prior to the first disease contact helps avoid an overestimation of incidence rates. Even though the use of such disease-free or "washout" or "look-back" periods in IBD research is not common practice, many administrative database studies in other chronic diseases have used washout periods to increase the accuracy of incidence estimates[7-13]. Griffiths et al. found that a minimum of 2 years was necessary to exclude pre-existing disease cases when using Medicare claims data[14]. It is very important to apply the disease free period at the individual level when information regarding the start of coverage in the health plan is not available. In our analyses, the two year disease free period had a small impact on the incidence rates. Bias in incidence estimates resulting from timing of case definition or the span of database observation period was, however, still present. In conclusion, we have shown that, when using administrative health databases in research, the methodological approach itself may lead to biased estimates of disease occurrence. Our results show that time-related bias can be minimized when case diagnosis is considered at the moment when all criteria are met, regardless of the case definition used. Implementing a disease free period is recommended to avoid overestimating the incidence rates. A uniform methodological approach that addresses these aspects would facilitate the comparison of estimates of disease occurrence between different populations.

REFERENCES

1. RAMQ. Nombre de personnes inscrites et admissibles au régime d'assurance maladie du Québec selon le sexe, le groupe d'âge et la région sociosanitaire 2013; https://www.prod.ramq.gouv.qc.ca/IST/CD/CDF_DifsnInfoStats/CDF1_CnsulInfoStatsCNC_iut/DifsnInfoStats.aspx?ETAPE_COUR=3&IdPatronRapp=8&Annee=1998&Per=0&LANGUE=fr-CA, 2013.

2. Statistics Canada. Age and Sex for Population, for Canada, Provinces, Territories and Federal Electoral Districts http://www12.statcan.gc.ca/english/census01/products/standard/themes/Rp-eng.cfm?TABID=1&LANG=E&A=R&APATH=3&DETAIL=1&DIM=0&FL=A&FREE=0&GC=24&GID=361266&GK=1&GRP=1&O=D&PID=55436&PRID=0&PTYPE=55430,53293,55440,55496,71090&S=0&SHOWALL=0&SUB=0&Temporal=2001&THEME=37&VID=0&VNAMEE=&VNAMEF=&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0.

3. Rothman KJ, Greenland, S. Section II Study design and conduct. In: Rothman KJ, Greenland, S., Lash, T.L., ed. *Modern epidemiology*. 3rd ed. Philadelphia, USA: Lippincott Williams&Wilkins; 2008.

4. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiology and drug safety.* 2007;16(3):241-249.

5. Suissa S. Immeasurable Time Bias in Observational Studies of Drug Effects on Mortality. *American journal of epidemiology.* August 1, 2008 2008;168(3):329-335.

6. Suissa S. Immortal Time Bias in Pharmacoepidemiology. *American journal of epidemiology.* February 15, 2008 2008;167(4):492-499.

7. Roifman I, Therrien J, Ionescu-Ittu R, et al. Coarctation of the aorta and coronary artery disease: fact or fiction? *Circulation.* Jul 3 2012;126(1):16-21.

8. St Germaine-Smith C, Liu M, Quan H, Wiebe S, Jette N. Development of an epilepsy-specific risk adjustment comorbidity index. *Epilepsia.* Dec 2011;52(12):2161-2167.

9. Tu K, Campbell NR, Chen ZL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Medicine.* 2007;1(1):e18-26.

10. Bernier M, Marelli AJ, Pilote L, et al. Atrial arrhythmias in adult patients with right- versus left-sided congenital heart disease anomalies. *American Journal of Cardiology.* 2010;106(4):547-551.

11. Rezai MR, MacLagan LC, Guo H, Donovan LR, Tu JV. Cardiovascular risk factors and hospitalizations in visible minority neighbourhoods in urban Ontario. *Canadian Journal of Cardiology.* September-October 2012;1):S223-S224.

12. Marrie RA, Yu BN, Leung S, et al. The incidence and prevalence of fibromyalgia are higher in multiple sclerosis than the general population: A population-based study. *Multiple Sclerosis and Related Disorders.* 2012;1(4):162-167.

13. Asghari S, Courteau J, Carpentier A, Vanasse A. Optimal strategy to identify incidence of diagnostic of diabetes using administrative data. *BMC Medical Research Methodology.* 2009;9(1):62.

14. Griffiths RI, O'Malley CD, Herbert RJ, Danese MD. Misclassification of incident conditions using claims data: impact of varying the period used to exclude pre-existing disease. *BMC Med Res Methodol.* 2013;13:32.

eTable 1. Estimating IBD incidence (cases/100,000 person-years) for the year 2001 in Québec: Effect of timing of diagnosis and duration of database observation period on time-related bias for a case definition involving an increasing number of IBD contacts, met in a specified two-year period.

Effect of timing of diagnosis						
Case definition	9 year observation time span			14 year observation time span		
	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)
2 IBD contacts	66.74	69.24	3.75	66.74	69.24	3.75
4 IBD contacts	37.83	39.81	5.23	37.83	39.81	5.23
6 IBD contacts	28.77	29.85	3.75	28.77	29.85	3.75
Effect of observation time span						
Case definition	Diagnosis at 1st IBD contact			Diagnosis at case-defining IBD contact		
	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)
2 IBD contacts	66.74	66.74	0.00	69.24	69.24	0.00
4 IBD contacts	37.83	37.83	0.00	39.81	39.81	0.00
6 IBD contacts	28.77	28.77	0.00	29.85	29.85	0.00

IBD=inflammatory bowel disease.

eTable 2. Estimating IBD incidence (cases/100,000 person-years) for the year 2001 in Québec: Effect of timing of diagnosis and duration of database observation period on time-related bias for a case definition involving an increasing number of IBD contacts, mandatory two year disease free period prior to first IBD contact.

Effect of timing of diagnosis						
Case definition	9 year observation time span			14 year observation time span		
	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)
2 IBD contacts	63.81	66.23	4.58	69.88	66.23	-5.22
4 IBD contacts	33.19	34.98	5.39	38.73	34.98	-9.68
6 IBD contacts	24.41	25.73	5.41	29.73	25.73	-13.45
Effect of observation time span						
Case definition	Diagnosis at 1st IBD contact			Diagnosis at case-defining IBD contact		
	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)
2 IBD contacts	63.81	69.88	9.51	66.23	66.23	0.00
4 IBD contacts	33.19	38.73	16.69	34.98	34.98	0.00
6 IBD contacts	24.41	29.73	21.79	25.73	25.73	0.00

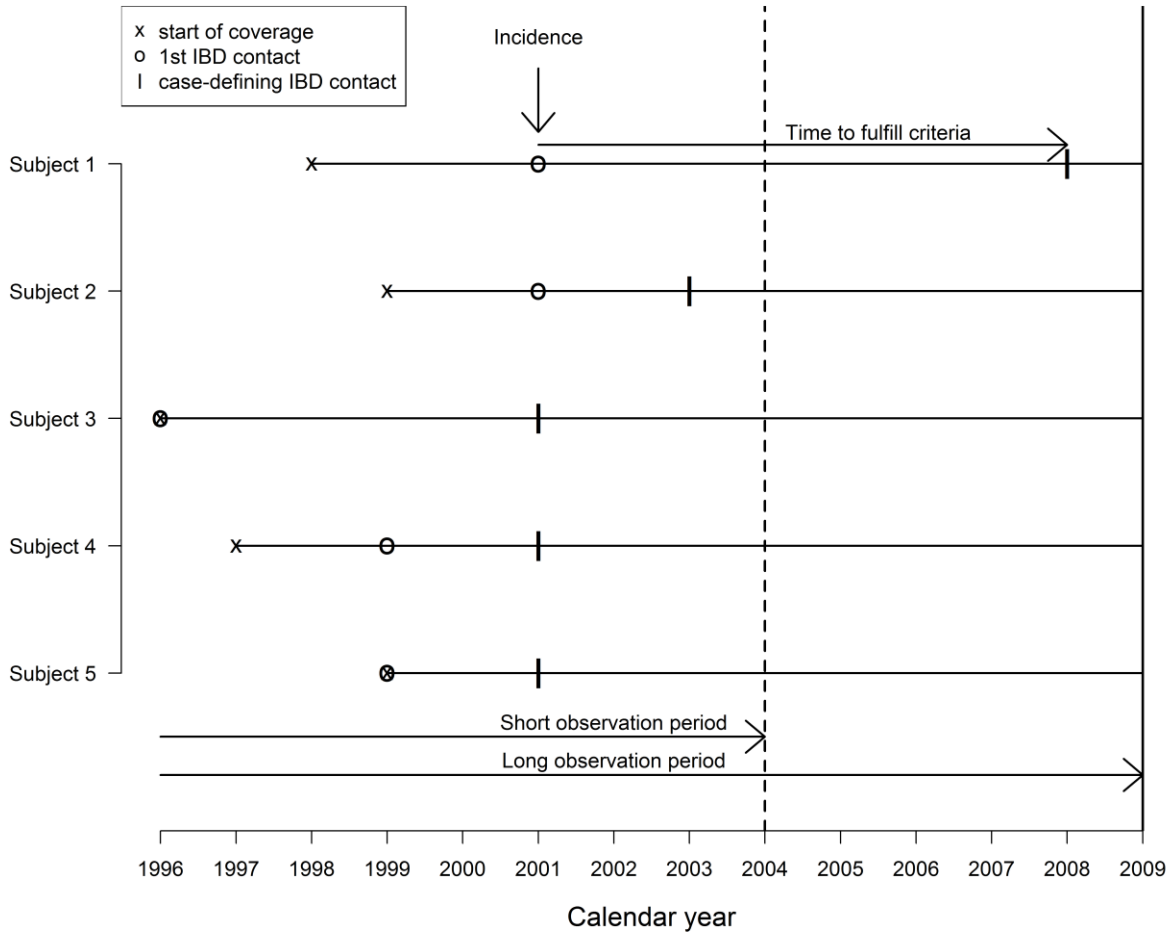
IBD=inflammatory bowel disease.

eTable 3. Estimating IBD incidence (cases/100,000 person-years) for the year 2001 in Québec: Effect of timing of diagnosis and duration of database observation period on time-related bias for a case definition involving an increasing number of IBD contacts, met in a specified two-year period, mandatory two year disease free period prior to first IBD contact.

Effect of timing of diagnosis						
Case definition	9 year observation time span			14 year observation time span		
	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)	Diagnosis at 1st IBD contact	Diagnosis at case-defining IBD contact	Bias (% change)
2 IBD contacts	65.41	67.36	2.98	65.41	67.36	2.98
4 IBD contacts	36.93	38.62	4.58	36.93	38.62	4.58
6 IBD contacts	28.07	29.17	4.58	28.07	29.17	4.58
Effect of observation time span						
Case definition	Diagnosis at 1st IBD contact			Diagnosis at case-defining IBD contact		
	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)	Observation period 1996-2004 (9 years)	Observation period 1996-2009 (14 years)	Bias (% change)
2 IBD contacts	65.41	65.41	0.00	67.36	67.36	0.00
4 IBD contacts	36.93	36.93	0.00	38.62	38.62	0.00
6 IBD contacts	28.07	28.07	0.00	29.17	29.17	0.00

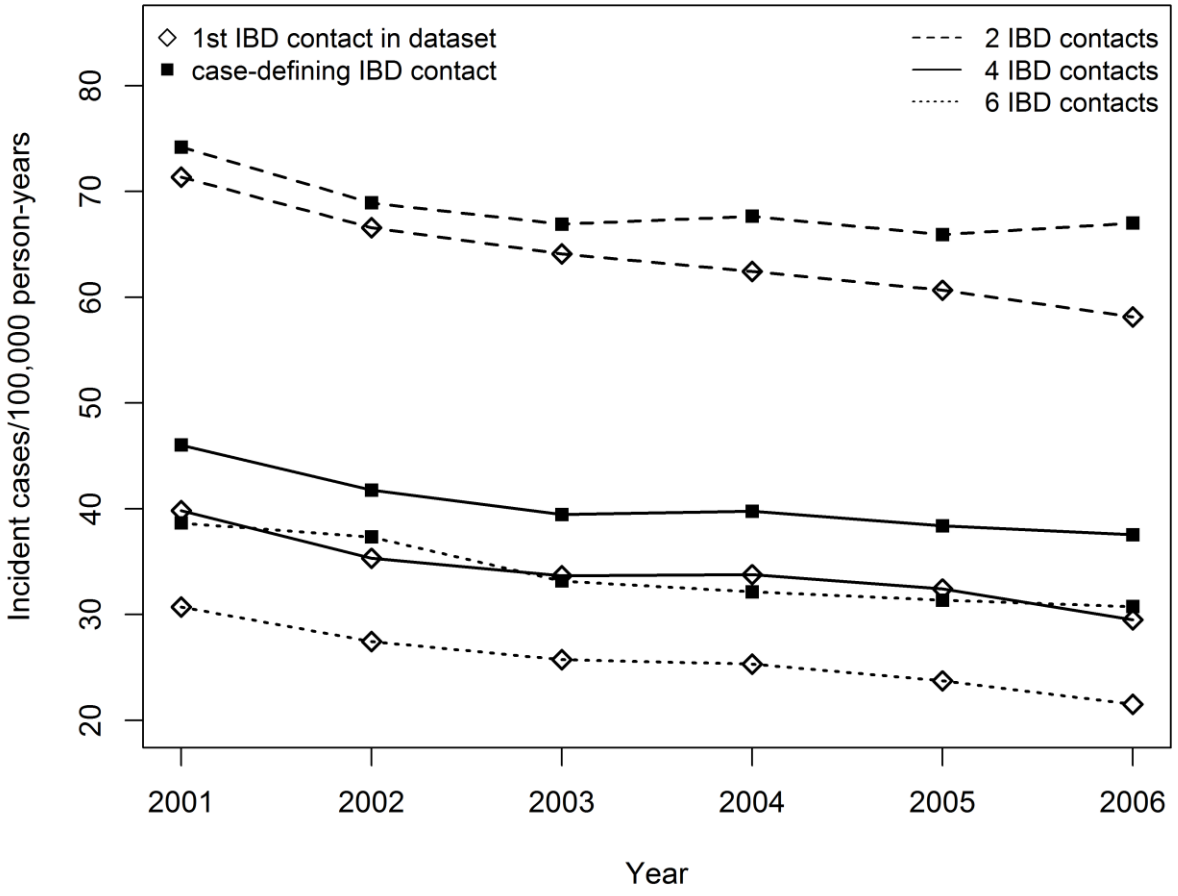
IBD=inflammatory bowel disease.

eFigure 1. Diagram of subjects to be counted as incident in 2001 depending on the methodological approach.



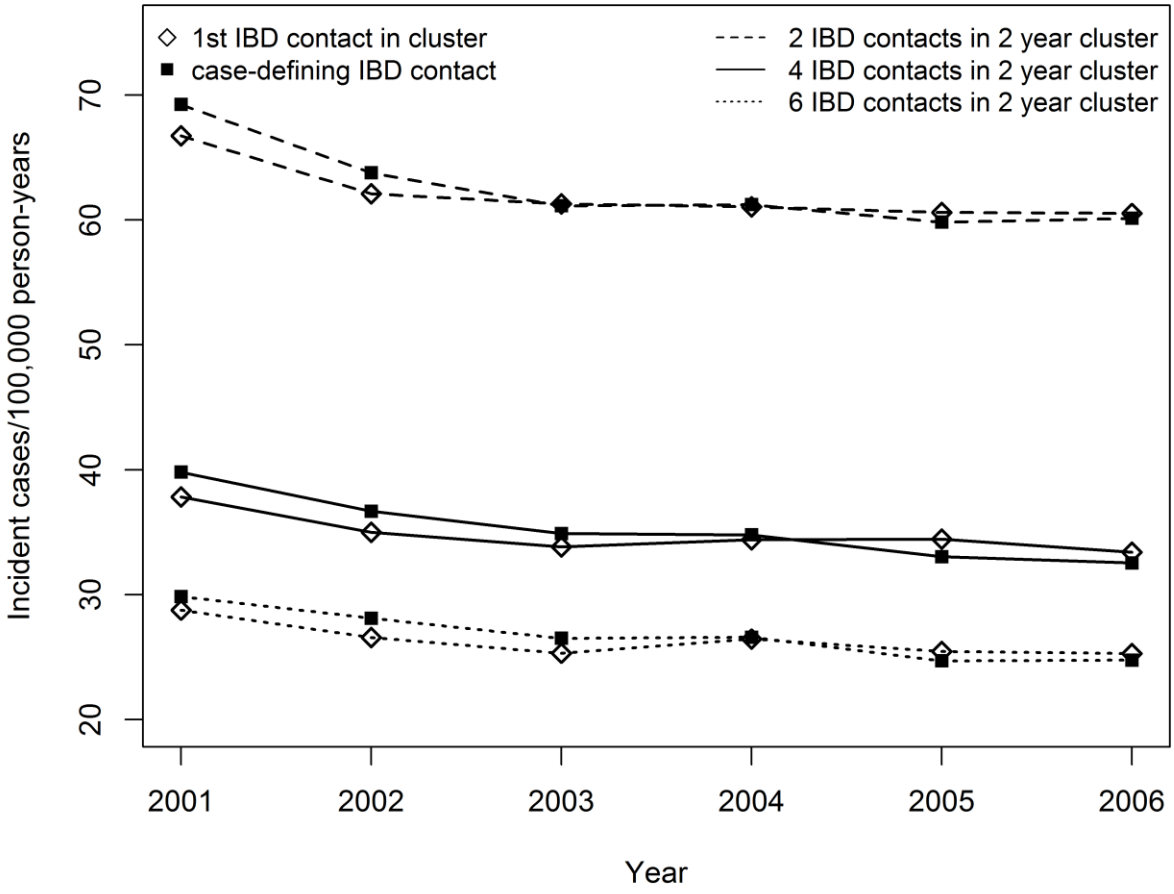
IBD=inflammatory bowel disease; start of coverage=start of coverage in the health plan; 1st IBD contact=first occurrence of a physician claim or hospitalization with an IBD diagnostic code (either for Crohn's disease or ulcerative colitis); case-defining IBD contact=IBD contact that completes the requirements of the case definition criteria.

eFigure 2. Estimating annual IBD incidence (cases/100,000 person-years) for 2001-2006 in Québec: Effect of timing of diagnosis for a case definition involving an increasing number of IBD contacts.



IBD=inflammatory bowel disease; hosp=hospitalization; claims=physician claims.
Note: Incidence rates were calculated using the Québec RAMQ insured population at July 1st, 2001. The rates were standardized by age and sex using the 2001 Québec population as available from the Institut de Statistique du Québec.

eFigure 3. Estimating annual IBD incidence (cases/100,000 person-years) for 2001-2006 in Québec: Effect of timing of diagnosis for a case definition involving an increasing number of IBD contacts, met in a specified two-year period.



IBD=inflammatory bowel disease; hosp=hospitalization; claims=physician claims.
Note: Incidence rates were calculated using the Québec RAMQ insured population at July 1st, 2001. The rates were standardized by age and sex using the 2001 Québec population as available from the Institut de Statistique du Québec.