UNIVERSITY OF CALGARY

DEVELOPMENT OF AN ICD CODING DEFINITION FOR INFLAMMATORY BOWEL DISEASE

by

ALI REZAIE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

FEBRUARY 2007

© Ali Rezaie 2007

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "DEVELOPMENT OF AN ICD CODING DEFINITION FOR INFLAMMATORY BOWEL DISEASE" submitted by Ali Rezaie in partial fulfilment of the requirements of the degree of Master Of Science.

Supervisor, Dr. Robert Hilsden
Departments of Medicine and Community Health Sciences

Dr. Remo Panaccione
Department of Medicine

Dr. Hude Quan
Department of Community Health Sciences

Dr. John Marshall
Department of Medicine, McMaster University

Date

Abstract

INTRODUCTION: To facilitate the conduction of population-based epidemiologic studies, this study was performed to develop a criterion for capturing IBD patients through comprehensive administrative databases in the province of Alberta, Canada. METHODS: Reports of endoscopy procedures performed within the Calgary Health Region are housed in a computerized database (EndoPro). Computerized and manual search strategies were conducted on EndoPro to categorize the patients undergone a colonoscopy into definite IBD, possible IBD, and definite non-IBD patients. Between 1995-2004, the medical service history (lCD diagnostic codes) was tracked by a unique personal health identification number in physician claims, hospital abstracts and Ambulatory Care Classification System (ACCS) databases. ICD-9 and lCD-10 codes for Crohn's disease and ulcerative colitis were extracted. Over 150 administrative definitions were developed for IBD patients depending on the number and periodicity of IBD codes in each database. Complementary chart reviews were done in case of disagreement between administrative criteria and EndoPro. RESULTS: 1,399 IBD patients and 15,439 definite non-IBD individuals were identified through EndoPro database search and chart reviews. The favorable definition classifies individuals as having IBD only if they have four physician claims OR two ACCS claims OR a hospitalization with an ICD diagnostic code of IBD. The definition must be met within a maximum time period of three years (Specificity, 99.75%, 95%CI, 99.66%-99.82%; Sensitivity, 85.70%, 95%CI, 83.76%-87.50%). CONCLUSION: Through a novel methodology we have developed a reliable lCD coding definition for IBD, which could be used in future population-based studies.

Acknowledgements

While I enjoyed many rewarding relationships throughout my training at U of C, I would especially like to thank my advisor, Dr.Robert Hilsden, for advising and mentoring me through the training. I would like to extend my gratitude and appreciation for the guidance provided to me by my dedicated thesis committee members, Dr.Hude Quan and Dr.Remo Panaccione.

I would like to thank the staff members in the Calgary Health Region especially Quality, Safety and Health Information portfolio.

I would like to thank all my friends particularly, Robin Walker and Deidre Hennessy for their wise advice and patient consultation.

Finally, I would like to thank my sister, her husband, my nephew and my little niece for being my greatest source of strength and renewal. I want to give special thanks to my parents for their unwavering love and support. I feel blessed to have such a family for teaching me how to embrace what life has to give.

Dedication

To my parents and my sister who made what I am.

Table of Contents

Approval Page ii
Abstract iii
Acknowledgements iv
Dedication v
Table of Contents vi
List of Tables ix
List of Figures and Diagrams x
List of Abbreviations xi
Epigraph xii

CHAPTER ONE: INTRODUCTION 1

CHAPTER TWO: BACKGROUND 3
2.1 Inflammatory Bowel disease 3
2.1.1 Crohn's disease 3
2.1.2 Ulcerative colitis 4
2.1.3 Diagnosis of Inflammatory Bowel disease 5
2.1.4 Endoscopic findings in Inflammatory Bowel disease 5
2.1.5 Epidemiology of IBD 6
2.2 Administrative databases 9
2.2.1 Hospital Discharge Abstract Database (Inpatient database) 10
2.2.2 Physician claims services database (PC database) 12
2.2.3 Ambulatory Care classification system (ACCS) 13
2.2.4 EndoPro database 13
2.3 International Classification of Diseases (ICD) 13
2.3.1 International Classification of Diseases, 9th revision, clinical modification (ICD-9-CM) 14
2.3.2 International Classification of Diseases, 10th revision (ICD-10) 14

CHAPTER THREE: SIGNIFICANCE AND OBJECTIVES 15
3.1 Significance 15
3.2 Objectives 16
3.2.1 Primary objective 16
3.2.2 Secondary objectives 16

CHAPTER FOUR: METHODOLOGY AND DATA ANALYSIS 17
4.1 Summarized design 17
4.2 Data sources 18
4.2.1 EndoPro database 18
4.2.2 Alberta Health Care Insurance Plan registry (AHCIP) 18
4.2.3 Physician claims database (PC) 18
4.2.4 Ambulatory Care Classification System (ACCS) database 19
4.2.5 Hospital Discharge Abstracts database (Inpatient database (IP)) 19
4.2.6 Chart reviews 19
4.3 Classification of endoscopic procedures within EndoPro database 19

4.3.1 Computerized search....19
4.3.2 Manual search....20
4.3.3 Final classification....20
4.4 Retrieving the personal health numbers through linkage with AHCIP....21
4.5 Classification of the subjects....21
4.6 Linkage with administrative databases....22
4.7 Preliminary development of the administrative definitions for IBD....23
4.7.1 Sensitivity....27
4.7.2 Specificity....27
4.7.3 Kappa....27
4.7.4 Predictive values....28
4.8 Chart reviews....28
4.8.1 False positives....29
4.8.2 False negatives....30
4.8.3 True positives....30
4.9 Integration of charts reviews with EndoPro search....30
4.10 Selection of the final administrative definition for IBD....30
4.11 Ethical considerations....31

CHAPTER FIVE: RESULTS....32
5.1 Classification of colonoscopies within EndoPro database....32
5.2 Linkage with the registry....34
5.3 Classification of the subjects....34
5.4 Linkage to the Calgary Health Region administrative data....34
5.4.1 IP database (FY 1994 – FY 2004)....35
5.4.2 PC database (FY 1995 – FY 2004)....37
5.4.3 ACCS database (FY 1996 – FY 2004)....38
5.5 Preliminary administrative definitions for IBD....40
5.6 Chart reviews....42
5.6.1 False positive subjects....42
5.6.2 False negative patients....47
5.6.3 True positive patients....47
5.7 Final classification of the subjects....48
5.8 Consistency of the final definition over age and gender....49

CHAPTER SIX: DISCUSSION....50
6.1 Study subjects....50
6.2 Separate linkage to administrative databases....51
6.2.1 IP database....51
6.2.2 PC database....52
6.2.3 ACCS database....53
6.3 Reasoning behind the selection of the final definition....54
6.3.1 One contact in IP database....55
6.3.2 Four claims in PC database....56
6.3.3 Two contacts in ACCS database....57
6.3.4 Periodicity (3 years)....57
6.3.5 Alternative definition for health regions with no access to ACCS database....58

6.4 Comparison of our administrative data and definition with other IBD administrative databases....58
6.5 Limitations....65
6.6 Conclusion....66

CHAPTER SEVEN: REFERENCES....68

APPENDIX 1. INTERNATIONAL CLASSIFICATION OF DISEASES, NINTH AND TENTH REVISION....71

APPENDIX 2. SEARCH TERMS IN ENDOPRO DATABASE....77

APPENDIX 3. ABSTRACTION FORM FOR INPATIENT AND OUTPATIENT CHARTS....78

APPENDIX 4. ETHICAL APPROVAL....81

APPENDIX 5. SENSITIVITIES AND SPECIFICITIES FOR 100 PRELIMINARY DEFINITIONS....83

APPENDIX 6. SENSITIVITIES AND SPECIFICITIES FOR 50 PRELIMINARY....93

DEFINITIONS THAT DID NOT CONSIDER ACCS DATABASE....93

APPENDIX 7. 2×2 TABLES AND TEST CHARACTERISTICS PREFERABLE PRELIMINARY ADMINISTRATIVE DEFINITIONS OUT OF 150 PRELIMINARY DEFINITIONS....98

APPENDIX 8. HEALTH SERVICE RECORDS OF OUR PATIENTS (AFTER CHART REVIEWS AND FINAL CLASSIFICATION) IN IP, PC AND ACCS DATABASES....107

APPENDIX 9. SENSITIVITIES AND SPECIFICITIES FOR 100 DEFINITIONS INCLUSIVE OF ACCS DATABASE (FINAL CLASSIFICATION)....111

APPENDIX 10. SENSITIVITIES AND SPECIFICITIES FOR 50 DEFINITIONS THAT DID NOT INCLUDE ACCS DATABASE , FINAL CLASSIFICATION 121

APPENDIX 11. 2×2 TABLES AND TEST CHARACTERISTICS FOR PREFERABLE ADMINISTRATIVE DEFINITIONS OUT OF 150 DEFINITIONS....126

List of Tables

Table1. Incidence and prevalence rates per 100,000 persons of IBD in different geographic regions 6

Table 1. 2×2 contingency table used to determine the administrative definition (administrative definition) characteristics 26

Table 2. Cumulative percentage and percentage of the top 11 procedures that had a diagnostic code for IBD in ACCS database. 40

Table 3. 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥1 *or* ACCS≥1 within 10 years" 41

Table 4. 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" 41

Table 5. . 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥5 *or* ACCS (not considered) within 2 years" 42

Table 6. Diagnoses made by chart reviews for 58 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were proven not to have IBD 43

Table 7. Diagnoses made by chart reviews for 21 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were shown that might have IBD 45

Table 8. Diagnoses made by chart reviews for 46 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were shown that do have IBD. 46

Table 9. Chart diagnoses for 32 random false negative patients 47

Table 10.Demographic characteristics of our IBD and non-IBD subjects 48

List of Figures and Diagrams

Figure 1. Distribution of the procedures that were considered as IBD among indication, findings, and impression fields....33

Figure 2. . Distribution of the procedures that were considered as "possible IBD" among indication, findings, and impression fields....33

Figure 3. Annual increase in the number of admissions among our cohort of patients irrespective of diagnosis....35

Figure 4. Annual increase in the number of admissions with a diagnosis of IBD....35

Diagram 1. Number of admissions and IBD diagnostic codes in IBD and non-IBD patients....36

Diagram 2. Frequency of physician claims and IBD diagnostic codes among IBD patients....37

Diagram 3. Frequency of physician claims and IBD diagnostic codes among non-IBD patients....38

Diagram 4. Frequency of ACCS contacts and IBD diagnostic codes among IBD and non-IBD patients....39

Diagram 5. Number of hospitalizations and IBD diagnostic codes in IBD and non-IBD patients after final classification....107

Diagram 6. Frequency of physician claims and IBD diagnostic codes among IBD patients after final classification....108

Diagram 7. Frequency of physician claims and IBD diagnostic codes among non-IBD patients after final classification....109

Diagram 8. Frequency of ACCS contacts and IBD diagnostic codes among IBD and non-IBD patients after final classification....110

List of Abbreviations

Symbol	Definition
ACCS	Ambulatory Care classification system database
AHCIP	Alberta health care insurance plan
CD	Crohn's disease
FN	False negative
FP	False positive
FY	Fiscal year
IBD	Inflammatory bowel disease
IC	Indeterminate colitis
IP	Hospital (inpatient) discharge abstract database
NPV	Negative predictive value
PC	Physician claims database
PHN	Personal health number
PPV	Positive predictive value
Sn	Sensitivity
Sp	Specificity
TN	True negative
TP	True positive
UC	Ulcerative colitis

Epigraph

"No knowledge is acquired prudent through the study of its causes and beginnings, if it has had causes and beginnings, nor completed except by knowledge of its effects and associated essentials"

Ibn Sina (Avicenna) Iranian Physician and philosopher (980–1037 AD): Iranian Culture Journal

Chapter One: INTRODUCTION

Crohn's disease (CD) and ulcerative colitis (UC), known as inflammatory bowel diseases (IBD), are chronic inflammatory conditions of the gastrointestinal tract. Although several hypotheses suggest a combination of environmental factors, genetic predisposition and dysfunctional immunoregulation for the etiology of IBD, the exact etiology of IBD remains uncertain. The highest incidence and prevalence of IBD is reported in Canada and about 0.5% of Canadians are estimated to have IBD [1]. Quality of life is negatively affected with loss of productivity in school or at work, and patients require continuous treatment and long-term follow-up.

One approach to achieve valuable clues to the etiology, or the social or economic burden of IBD is through meticulous population-based epidemiologic studies. Epidemiologic studies describe, understand, interpret and predict the impact of a disease on public health to a wide variety of audiences including caregivers, patients, decision makers, government agencies, and service providers.

However, establishing a universal registry for IBD patients in a health region is not an easy task. More and more IBD cases are being diagnosed and managed in outpatient clinics [2], so that hospital based data would not be adequate to identify IBD patients. In Canada there is a unique ability to access health care contacts through comprehensive administrative databases, which can be utilized to identify IBD patients.

Although Alberta benefits from a rich and sophisticated health care data collection system, there are no indigenous administrative criteria to extract IBD cases from the administrative databases. This study was conducted to develop an administrative

definition for identification of IBD patients in the Calgary Health Region to facilitate epidemiologic and health services research.

Chapter Two: BACKGROUND

2.1 Inflammatory Bowel disease

CD and UC are chronic inflammatory conditions of the gastrointestinal tract and share several pathologic and clinical presentations. Occasionally these two diseases are not clinically or histologically distinguishable, a condition called indeterminate colitis (IC). Although several hypotheses suggest a combination of environmental factors, genetic predisposition and dysfunctional immunoregulation for the etiology of IBD, the exact etiology of IBD remains uncertain.

2.1.1 Crohn's disease

CD or regional enteritis is a chronic relapsing granulomatous condition that involves mainly the terminal ileum; however, the entire gastrointestinal tract from mouth to anus can be affected by this disease. Inflammation extends through the intestinal wall from mucosa to serosa. Compared to UC, CD is a more complex and diverse entity. Patients experience a variety of manifestations and a fluctuating clinical course [3].

Anatomic distribution of the disease has three major patterns: (1) In 40% of the cases disease is present both in the ileum and the cecum (ileocolitis); (2) 30% have disease limited to the small intestine and (3) in 25% of the cases disease is confined to colon. The predominant symptoms are diarrhea (100% of ileocolitis cases), abdominal pain (62%), weight loss (19%), and perianal disease (38%) [4]. In addition, several debilitating exteraintestinal manifestations such as arthritis, ankylosing spondylitis, and hepatic complications can emerge in CD patients [5].

Without treatment, half of the quiescent cases of CD will experience a relapse in a period of two years. Sixty percent of the patients undergo surgery within 10 years of diagnosis, of which, half of them will experience a clinical relapse within 4 years of surgery. Eventually, 45% of CD patients will require another operation [6]. CD patients suffer a plethora of disabilities in their social life such as career change, job absence, and decreased work hours [7].

2.1.2 Ulcerative colitis

Similar to CD, ulcerative colitis is a chronic inflammatory condition of digestive tract. UC habitually involves the rectal mucosa; however it can extend to involve the left colon or the entire colon, a condition called "pancolitis". UC can appear at any age but classically happens in early adulthood. The disease is slightly more common in women and 15% of patients have a family history of IBD [3].

The majority of the patients have proctitis or proctosigmoiditis; therefore, blood in the stool is often the first manifestation of the disease. Apart from tenesmus and cramping, most patients have frequent bowel movements containing pus, mucus, and blood. As more of the colon becomes involved, diarrhea happens to be the prominent feature of the disease.

As with CD, UC patients may develop extraintestinal complications such as skin eruptions, arthritis and eye problems. The most important hepatobiliary complication of UC is sclerosing cholangitis, which is characterized by inflammation, destruction and fibrosis of the intrahepatic and extrahepatic bile ducts. Mild diffuse fatty liver changes

are usually apparent. Potentially lethal complications like toxic megacolon (i.e. marked dilation of the colon) or colonic perforation rarely occur.

2.1.3 Diagnosis of Inflammatory Bowel disease

There is no single gold standard test to diagnose IBD, but the prime test remains colonoscopy with mucosal biopsy.

Laboratory findings (e.g. anemia, mild leukocytosis, increased acute phase reactants, and nutritional deficiencies) of CD are mostly non-specific and are mainly used for prognosis rather than diagnosis of the disease [3].

2.1.4 Endoscopic findings in Inflammatory Bowel disease

Endoscopy plays a significant role in diagnosis of IBD, distinguishing between CD and UC, determining disease activity, and evaluating the response to treatment. Moreover, colonoscopy can be used as a surveillance tool for colon cancer in IBD patients due to their higher risk for developing colon dysplasia [8].

The earliest endoscopic presentation of CD is an apthous ulcer, a small discrete ulcer surrounded by a thin erythematous halo. Apthous ulcers usually occur in groups and can grow into large and deep ulcers. Ulcerated areas are usually surrounded by normal mucosa (i.e. skip lesions). Erythema, granularity, pseudopolyps, stenosis and edema may also be present [9].

Loss of vascular pattern and development of diffuse erythema in the rectum are usually the first endoscopic manifestations of UC. Typically, mucosal edema is present that is

manifested by loss of normal vasculature, blunting of the haustral pattern and granularity. A yellow mucosal exudate and inflammatory pseudopolyps might be present. Inflamed mucosa is friable, meaning that the intestinal wall bleeds easily following minor trauma [10].

2.1.5 Epidemiology of IBD

The literature surrounding the epidemiology of IBD is very rich [11]. Table 1. summarizes some of the incidence and prevalence rates reported in population-based studies for CD and UC.

Table1. Incidence and prevalence rates per 100,000 persons of IBD in different geographic regions

Region	Year	UC incidence	UC prevalence	CD incidence	CD prevalence
North America					
Northern Alberta[12]	1981	6	37.5	10	44.4
Manitoba[2]	1994	14.3	169.7	14.6	198.5
Olmsted, MN[13]	1993	8.3	229	6.9	144.1
Europe					
Denmark[14]	1987	9.2	161.2	4.2	54
Denmark[15]	1999	20.3	-	3.6	-
UK[16]	1995	13.9	243	8.3	144
Italy[17]	1992	9.6	121	3.4	40
Croatia[18, 19]	1989	1.5	21.4	0.7	8.3
East Asia					
Japan[20]	1991	1.9	18.1	0.5	5.8

Incidence and prevalence rates of IBD are dynamic and differ enormously based on geographic region. The highest reported rates for CD are from Manitoba [2]. For UC the highest rates have been reported from Manitoba [2], US [13] and Denmark [14]. In general the incidence and prevalence of IBD seems to be higher in northern regions [11] and also western locales (North to South, West to East gradient).

In addition to the incidence and prevalence of IBD being different all around the world, it seems that the rates are not constant in a given region. In the mid 1980s and 1990s, explosive increases were reported from Europe, Africa and Latin America [21].

Although some reports have mentioned a plateau in developed countries, this increase still continues in developing countries [13].

Three population-based epidemiologic studies have been performed on IBD patients in Canada:

1-Northern Alberta: (Population=1,295,360)

Pinchbeck et al. [12] identified all the patients discharged from northern Alberta hospitals with the diagnosis of IBD from 1977 to 1981, along with the records of IBD patients being managed by Edmonton gastroenterologists in case they were not hospitalized within that period. All of eligible 2,419 charts were reviewed and only those with clinical plus radiological and/or pathological criteria were included (n=1716). Results of this study are shown in Table 1.

2-Manitoba: (Population≈1,140,000)

Bernstein et al. [2] used Manitoba health insurance databases to identify IBD patients. Physicians billing claims and hospital abstracts were searched for *International Classification of Diseases, 9th revision (ICD-9)* diagnostic codes for CD and UC between 1984 and 1998.

Those with ≥3 health care contacts from 1984-1993 and at least one contact from 1993-1995 were included in the study. A questionnaire was sent to the patients (60% response rate) and chart reviews were done on 16% of the respondents to establish and validate a definition for the identify IBD patients. Considering equal sensitivity and specificity of 90%, the following criteria were developed:

- For individuals who had been a resident or registered with the insurance system for at least 2 years, cases were those with at least 5 health care contacts (physician claim or hospitalization) with diagnostic codes of IBD.
- For residents who had been insured for less than 2 years, 3 health care contacts were required for the diagnosis of IBD.
- Those who fulfilled both CD and UC criteria were classified as CD or UC based on the majority diagnosis of the 9 most recent coantacts.

The annual mid-year population for incidence and prevalence rates was obtained from Manitoba Health registry. The date of the first medical contact was considered as the date of diagnosis and the incidence rates were calculated for the cases with the first medical

contact in 1989 or later, to allow for a hiatus of at least 5 years. Results of this study are shown in Table 1.

3-Five provinces of Canada

Bernstein et al. used the same criteria for the five provinces of Alberta, British Columbia, Nova Scotia, Manitoba and Saskatchewan [1]. In 1998-2000 the incidence rate of CD per 100,000 were 16.4-17 (AB), 12.2-14.1 (SK), 14.8-15.7 (MB), 8.1-9.6 (BC), and 18.4-22.5 (NS). The incidence rate for UC were 10-12.4 (AB), 8.6-11.1 (SK), 14.8-15.7 (MB), 7.6-11.7 (BC), and 17.9-21.2 (NS). The prevalence rates for CD were 281-287 (AB), 231-250 (SK), 249-265 (MB), 159-164 (BC), and 299-325 (NS). Prevalence rates for UC were 180-188 (AB), 212-225 (SK), 226-246 (MB), 160-167 (BC), and 222-255 (NS). Extrapolation of these results suggests that approximately 0.5% of the Canadian population has IBD.

2.2 Administrative databases

Over the last few decades, the use of computers has improved the process of data storage and has allowed large volumes of information to be analysed. Administrative datasets are secondary collected datasets that are less costly to acquire, troublesome to peruse and cumbersome to update than health records data. Moreover, administrative datasets include a defined population, allowing accurate estimates of rare diseases.

Available information in administrative databases can be classified into four domains. The first is the *unique identifiers* for the patients and physicians. The second domain is

demographic data such as age, gender and postal code for place of residence. The third domain is *disease*, which includes *most responsible diagnosis, co-morbidities* and *complications.* The fourth domain is the medical and surgical *procedures.*

Although administrative data offers practical advantages, investigators have no control over the quality of the data collection; therefore, accuracy and completeness of these datasets has always been a source of concern [22-25]. The best solution to overcome the concern of validity is to link several independently collected databases [26].

Although administrative data may be accompanied by a trade-off in credibility, this is often counterbalanced by the feasibility and comprehensiveness.

In Canada, administrative databases were implemented by provincial governments after the introduction of universal medical insurance. The structure of the administrative databases used in this study is described below:

2.2.1 Hospital Discharge Abstract Database (Inpatient database)

The Discharge Abstract Database is one of the key database holdings in the Calgary Health Region. It is completed according to instructions outlined in the *Canadian Institute of Health Information* (*CIHI)* Abstract Manual [27] after patients leave the hospital (acute care, chronic care, and rehabilitation) to reflect diagnoses and procedures performed during the period of hospitalization [28]. The database records up to sixteen diagnostic (ICD-9-CM) and ten procedure codes (ICD-9-CM and Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures), demographic data (e.g. age and gender), health care number, and admission and discharge dates in the system. The most responsible diagnosis in CIHI is defined as "the one diagnosis which describes the most

significant condition of the patient which causes his/her *stay* in hospital". That is in contrast with US databases with emphasis on patient admission diagnosis [29].

Hospitals forward discharge abstracts to CIHI every month with a one year delay between CIHI receiving the data and abstracts becoming available. Data is coded in ICD-9-CM until fiscal year 2002/2003, when a transition was made towards International Statistical Classification of Diseases and related Health problems, Tenth Revision, Canadian modification (ICD-10-CA) [30].

The accuracy of CIHI data in comparison with hospital discharge abstracts has been evaluated in several studies. The Ontario Hospital Association (OHA) conducted a re-abstraction study in 1991, reviewing 3,000 randomly selected charts from 43 hospitals [31]. The study revealed that the rate of agreement on demographic data was greater than 95% and the rate of agreement for principal and secondary diagnosis was 81% and 37%, respectively. In another study an almost perfect agreement was found for demographic and procedure codes for 175 inpatients undergoing hip replacement [32].

Acute myocardial infarction coding was examined in CIHI hospital abstracts in Ontario [29]. A total of 58,816 patients participating in the Fastrak II Acute Coronary Syndromes registry were included in this study. The records obtained from Fastrak registry were linked to IP database. The PPV, sensitivity and specificity for the main diagnosis of AMI were 88.5%, 88.8% and 92.8%, respectively. The specificities for main diagnoses of arrhythmia, congestive heart failure, chest pain and unstable angina were all at least 93.9%; however, the sensitivities and PPVs were no greater than 60.7% and 80.8%, respectively.

"The national data quality study of Canadian Discharge Abstract Database" is an ongoing study to ensure the data quality of IP database. Preliminary findings from the first year (fiscal year 1999/2000) have now been released. Its target population consists of approximately 2.5 million inpatients in 550 acute care hospitals. Injury hospitalization, vaginal births after cesarean section, hospitalization due to pneumonia and influenza, coronary artery graft, and total hip replacement have been considered. The highest number of false positive and false negative is reported for cases of pneumonia (i.e. 7.1% and 12.8%, respectively) and the lowest percentages are documented for delivery (0 %) [28].

2.2.2 Physician claims services database (PC database)

This database contains information obtained from physicians' payment claims. For every eligible visit or procedure, a physician files a payment claim, in which the diagnosis and service codes are recorded along with patient (age, gender and health card number) and physician information.

Unlike hospital discharge abstracts, accuracy of physician claims databases have been given less attention. Reported agreements between physician claims and hospital discharge abstracts for schizophrenia and depression in Saskatchewan have been 89% and 94%, respectively [33]. Two other studies that considered self-reports as gold standard found 96%, 87%, and 85% agreement for myocardial infarction, diabetes and hypertension [34, 35].

Young et al.[36] found that agreement is 95% when comparing hospital discharge abstracts with physician claims for diagnosis of diabetes. The three studies discussed above all have been conducted in Manitoba.

2.2.3 Ambulatory Care classification system (ACCS)

Implemented in 1997, ACCS is a made-in-Alberta system that collects information on facility-based ambulatory care (i.e. day surgeries and procedures, and ER visits). Recorded data consist of patient demographic data, service dates, diagnostic codes (6 diagnostic codes (ICD-9) until FY 2001/2002; 10 diagnostic codes (ICD-10-CA) from FY 2002/2003), and procedure codes [37].

2.2.4 EndoPro database

Pentax's endoPRO® is a management tool for endoscopy procedures used in the Calgary Health Region. Physicians performing the endoscopy procedure generate a procedure report that contains patient's demographic information, physical exam, brief medical history, indication (the reason why the patients underwent colonoscopy), endoscopy findings, impressions (gastroenterologist's diagnosis) and recommendations. This program has been implemented in Calgary since May 2000. Each procedure has a unique identifier.

2.3 International Classification of Diseases (ICD)

ICD is a system to classify medical diagnoses and procedures into universal medical code numbers. It also provides codes to classify signs, symptoms, external sources of injury and complaints. ICD is used for a variety of applications in medicine and bioinformatics

such as epidemiological and statistical analysis of morbidities and mortalities, reimbursements, and surveillance. The variety and extensiveness of the ICD codes enable us to draw a fair picture of the patient's medical status. The International Classification of Diseases is published and maintained by the World Health Organization and is revised periodically [38].

2.3.1 International Classification of Diseases, 9th revision, clinical modification (ICD-9-CM)

In 1977 ICD-9 was published by the WHO. This version was extended with extra morbidity codes and a section of procedure codes. The extension was called ICD-9-CM, which has over 14,000 codes than can be shown with a maximum of 5 digits based on the level of specificity. (Appendix 1)

2.3.2 International Classification of Diseases, 10th revision (ICD-10)

Development of ICD-10 began in 1983 and was completed in 1992; however, transition from the ninth to tenth revision of ICD has not been implemented in several health care databases such as physician claims database.

Canada has introduced an enhanced version of ICD-10 called ICD-10-CA. It benefits from an alpha-numeric format ranging from 3 to 6 characters [39] (Appendix1).

Chapter Three: Significance AND OBJECTIVES

3.1 Significance

- This study bears important theoretical implications, as studying the epidemiology of IBD remains critically important to search for clues on its pathophysiology and etiology.
- Another problem that has risen in recent years due to improvements in the treatment of IBD is that many IBD patients are never admitted to the hospital; therefore, population indices are not achievable exclusively through hospital records.
- Although Alberta benefits from a rich and sophisticated health care data collection system, we do not have any indigenous administrative definition to extract IBD cases from the administrative databases.
- Different health insurers and consequently the lack of all-inclusive databases makes population-based epidemiologic studies extremely difficult in US; however, Canadian provinces have unique characteristic of "single insurer system", which allow us to establish a population-based database of IBD patients.
- The development of an administrative IBD definition will lay the groundwork for future population-based studies in IBD patients. Moreover, it will facilitate determining the incidence and prevalence of IBD, and also will provide valuable information on the economic burden of the disease for policy makers and health care planners.

- Genetic predispositions and environmental factors both play potential etiologic roles in IBD; furthermore, the epidemiology of IBD changes over time. Hence, establishing a database for IBD patients in the Calgary Health Region is essential to obtain valuable information about the direction and determinants of the dynamic process of IBD epidemiology in this region.
- This study will provide valuable information regarding the accuracy of Alberta physician claims, and endoscopy and hospital discharge abstract databases.

3.2 Objectives

3.2.1 Primary objective

- To develop an ICD coding definition to identify IBD patients using IP, PC and ACCS administrative databases in the Calgary Health Region.
- To determine the sensitivity and specificity of the administrative definition

3.2.2 Secondary objectives

- To develop an alternative ICD coding definition using administrative databases of IP and PC for health regions that do not have access to ACCS database
- To describe the distribution of IBD diagnostic codes in the Calgary Health Region's administrative databases

Chapter Four: METHODOLOGY AND DATA ANALYSIS

4.1 Summarized design

Computerized and manual search strategies were conducted on EndoPro database to categorize the patients who had undergone a colonoscopy within the period of May 16th, 2000 through April 1st, 2004 into definite IBD, possible IBD, and definite non-IBD patients. Records were linked to Alberta Health Care Insurance Plan population registry to retrieve unique personal health identification numbers (PHNs). PHN was then used to identify any records for these patients in physician claims, hospital discharge abstracts and Ambulatory Care Classification System (ACCS) databases form April 1995 through April 2004 that included a diagnostic code for CD or UC. Over 150 administrative definitions were developed to classify patients as having or not having IBD depending on the number and periodicity of IBD codes in each database. Chart reviews were completed on false positive, false negative and true positive cases to increase the accuracy of our definitions. The EndoPro and chart diagnosis was then used as the 'gold standard' test to determine if the classification by 'administrative definitions' was correct. The most practical definition was chosen to have a high specificity and an acceptable sensitivity.

4.2 Data sources

4.2.1 EndoPro database

Endoscopy reports containing a unique procedure identification number, hospital chart number, procedure date, patient's demographic information, physical exam, brief medical history, indication, endoscopic findings, impressions, and recommendations were examined within the time period of May 16th, 2000 through April 1st, 2004. Each procedure can have multiple indications, findings or impressions, which the endoscopist entered either from a drop-down menu or as free text.

4.2.2 Alberta Health Care Insurance Plan registry (AHCIP)

This database is housed in the Calgary Health Region and contains PHN, patient's demographic data, and activity of the patient in the health region within each fiscal year. Calgary Health Region's access to information is limited to the individuals who are or have had been residents of the Calgary Health Region.

4.2.3 Physician claims database (PC)

Calgary Health Region supplied the data containing PHN, patient's demographic data, ICD-9-CM diagnostic codes and schedule of medical benefits' service codes within the time period of April 1995 through April 2004.

4.2.4 Ambulatory Care Classification System (ACCS) database

Calgary Health Region supplied the data containing PHN, patient's demographic data, and ICD-10-CA procedure and diagnostic codes within the time period of April 1996 through April 2004.

4.2.5 Hospital Discharge Abstracts database (Inpatient database (IP))

Calgary Health Region supplied the data containing PHN, patient's demographic data, and procedure and diagnostic codes within the time period of April 1996 through April 2004. Data is coded in ICD-9-CM until fiscal year 2002/2003 when a transition was made to ICD-10-CA diagnostic codes (International Classification of Disease- Canadian enhancement).

4.2.6 Chart reviews

Reviews on inpatient and outpatient charts were performed for selected subjects. All the available chart volumes after 1995 in Calgary hospitals (i.e. Rocky View General Hospital, Foothills Medical Centre, and Peter Lougheed center) were reviewed.

4.3 Classification of endoscopic procedures within EndoPro database

4.3.1 Computerized search

Using IBD-related terms, three separate search strategies in EndoPro database were developed for indication, findings, and impression fields. Each search strategy assigned the procedures into IBD, non-IBD, and possible IBD categories (Appendix 2).

4.3.2 Manual search

As some physicians prefer entering procedure information in the text fields of the endoscopy reports rather than using the drop-down options, indication, findings and impression text fields were also screened manually and procedures were categorized into IBD, non-IBD, and possible IBD.

4.3.3 Final classification

Computerized search strategies and manual searches were combined to develop the final classification of each endoscopic procedure:

- A procedure was considered as IBD if it had been categorized as: IBD in *Indication* field computerized or manual search **OR** IBD in *Findings* field computerized or manual search **OR** IBD in *Impression* field computerized or manual search.
- If the procedure was not considered as IBD, it would be considered as possible IBD if it had been categorized as: Possible IBD in *Indication* field computerized or manual search **OR** possible IBD in *Findings* field computerized or manual search **OR** possible IBD in *Impression* field computerized or manual search.
- A procedure was considered as non-IBD if it had been categorized as: Non- IBD in *Indication* field computerized or manual search **AND** non-IBD in *Findings* field computerized or manual search **AND** non-IBD in *Impression* field computerized or manual search.

4.4 Retrieving the personal health numbers through linkage with AHCIP

A data file containing the EndoPro search status, unique procedure identification numbers, hospital chart numbers, and patients' demographic information was developed and sent to the Quality, Safety & Health Information department of the Calgary Health Region for linkage.

Data was linked with the Alberta Health Care registry using last name, first name, date of birth and gender. A personal health number was assigned to each linkable record. Linkage was done by the Calgary Health Region personnel. If two or more of these variables were missing for any procedure, the procedure was considered as unlinkable and was removed from the linkage process.

4.5 Classification of the subjects

As some patients have several endoscopic procedures within the time period of May 16th, 2000 through April 1st, 2004 and consequently had several records in the EndoPro database; we classified patients into IBD, possible IBD and IBD with the following algorithm:

- A patient was considered to have IBD if she/he had had at least one endoscopic procedure classified as IBD.
- A patient was considered to possibly have IBD if she/he had had at least one endoscopic procedure classified as possible IBD **AND** no endoscopic procedure classified as IBD.
- A patient was considered not to have IBD if she/he had had no endoscopic procedure classified as IBD **OR** possible IBD.

4.6 Linkage with administrative databases

Our cohort of IBD, non-IBD, and possible IBD patients was linked to the Calgary Health Region administrative databases using their unique PHNs. Linkage was done separately to PC (FY1995-FY2004), IP (FY1995-FY2004) and ACCS (FY1996-FY2004) databases. All the data files were transformed from SAS into STATA datasets and all of the analysis was done with STATA software version 9.0 (Stata Corp, College Station, Tex).

In each database, any record with an IBD diagnostic ICD code in any diagnosis field was identified. A preliminary analysis was done regarding the frequency and periodicity of IBD codes in each database. In addition the procedures that led to a diagnosis of IBD were retrieved in the ACCS database. The status of the IBD diagnostic code in IP database was also determined (i.e. most responsible diagnosis or not).

A pooled dataset was developed out of IP, ACCS, and PC databases with the following variables:

- PHNs
- IBD diagnostic code(s) in IP database for each PHN
- Assignment date(s) of the IBD code(s) in IP database IBD for each PHN
- IBD diagnostic code(s) in ACCS database for each PHN
- Assignment date(s) of the IBD diagnostic code(s) in IP database for each PHN
- IBD diagnostic code(s) in PC database for each PHN
- Assignment date(s) of the IBD diagnostic code(s) in PC database for each PHN
- IBD status for each PHN (retrieved from the EndoPro search)

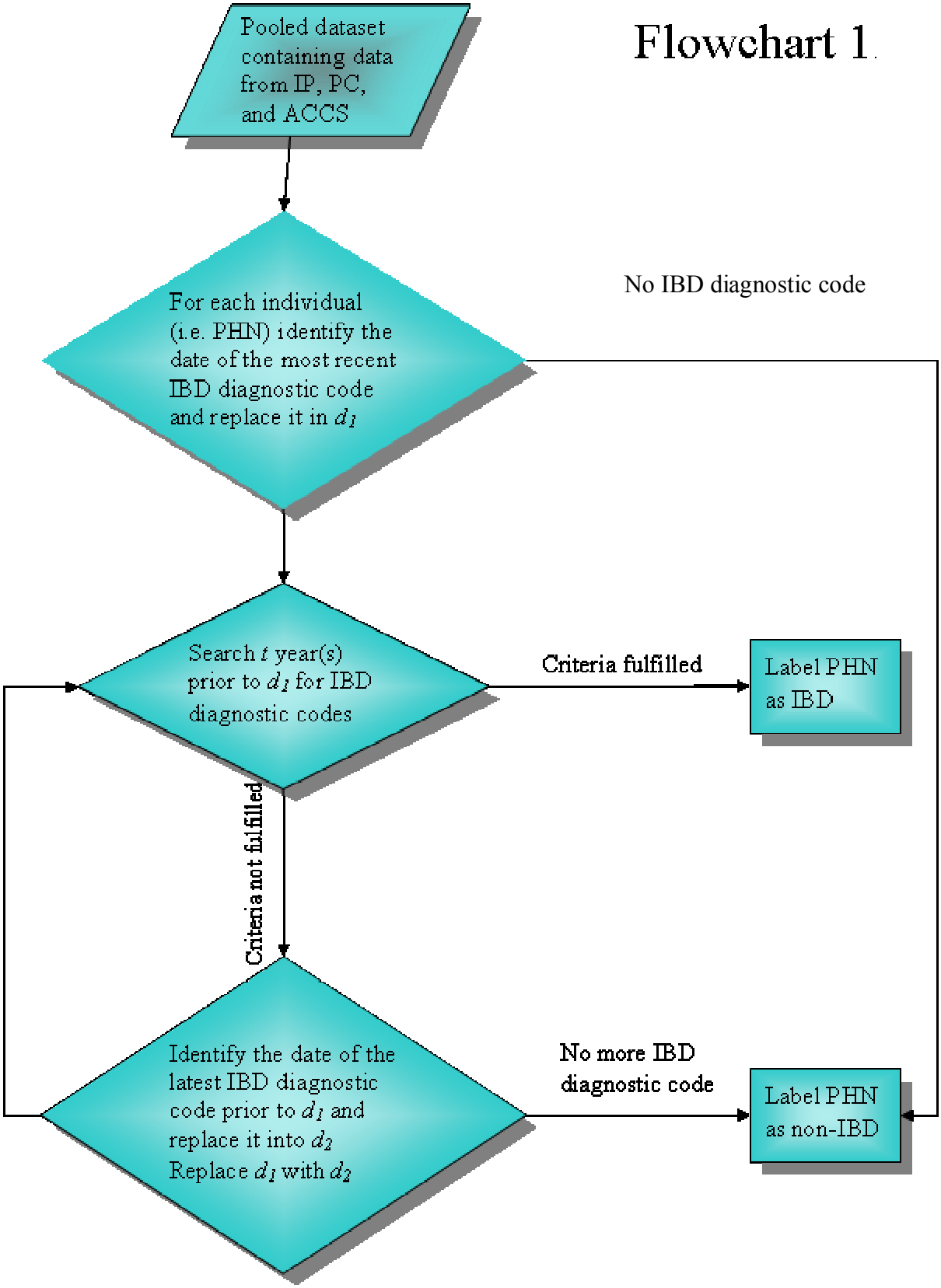
4.7 Preliminary development of the administrative definitions for IBD

Depending on the number of the IBD diagnostic codes in the administrative databases of IP, PC, and ACCS, we developed several administrative definitions within different time frames:

Condition {(*x* IBD codes in IP within *t* years) **OR** (*y* IBD codes in PC within *t* years) **OR** (*z* IBD codes in ACCS within *t* years)} = True $\Longleftrightarrow$ Patient has IBD

A computerized algorithm was developed to process the data based on Flowchart 1.

Flowchart 1


Pooled dataset containing data from IP, PC, and ACCS
For each individual (i.e. PHN) identify the date of the most recent IBD diagnostic code and replace it in d_1
No IBD diagnostic code
Search t year(s) prior to d_1 for IBD diagnostic codes
Criteria fulfilled
Label PHN as IBD
Criteria not fulfilled
Identify the date of the latest IBD diagnostic code prior to d_1 and replace it into d_2 Replace d_1 with d_2
No more IBD diagnostic code
Label PHN as non-IBD

Each time frame[1] (i.e. *'t' in flowchart 1)* was set to be started with an IBD diagnostic code and was searched for the desired numbers of IBD diagnostic codes in each administrative database. For each individual, several iterations of the search loop could have been completed depending on the cumulative number of IBD diagnostic codes. Loop breakers were implemented in the computerized algorithm to enhance the computation pace.

Based on the results of the preliminary analysis done in section 4.6; the following combinations of time frames and IBD diagnostic codes frequencies were incorporated into the computerized algorithm (i.e. a total of 100 administrative definitions):

- IP database: 1 and 2 IBD diagnostic code(s)
- PC database: 2 to 6 IBD diagnostic codes
- ACCS database: 2 to 3 IBD diagnostic codes
- Different time frames of 1 to 5 year(s)

As mentioned previously, ACCS is only available in the province of Alberta; therefore, to facilitate the use of our **definition** in other provinces of Canada, we conducted a second analysis in which the ACCS database was not included (i.e. a total of 50 administrative definitions):

- IP database: 1 and 2 IBD diagnostic code(s)
- PC database: 2 to 6 IBD diagnostic codes
- ACCS database: Not included in the analysis

---

[1] Can accept non-integer values (e.g. 0.5 year)

- Different time frames of 1 to 5 year(s)

To assess the empirical association of each definition with 'gold standard' diagnostic test (here, EndoPro search), 2×2 contingency tables were developed (Table 1). As we were examining disease prediction in our study with two discrete categories of "diseased" and "non-diseased"; it was more appropriate to assess the accuracy of our administrative definitions by 2×2 tables rather than correlation coefficients [40].

Table 1. 2×2 contingency table used to determine the administrative definition characteristics

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	True positive	False negative
	Non-IBD	False positive	True negative

Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each definition using the corresponding 2×2 table. For each test characteristic, 95% confidence intervals (CI) were calculated using exact method [41].

Two possible errors were *false positives* (i.e. Individuals who were considered not to have IBD by the gold standard test (EndoPro) but fulfilled the administrative definition for IBD) and *false negatives* (i.e. an IBD patient that the administrative definition failed to identify). Depending on where the dividing line was placed one error would have been increased and the other would have been decreased. Considering that we aimed to identify cases of a rare disease, the percentage of *FPs* should have been far less than *FNs*.

4.7.1 Sensitivity

Sensitivity (true positive rate) is one of the measures of the validity of the test; it is defined as the probability that an IBD diagnosis within the specified group was asserted by the administrative definition given that it was actually present in the gold standard. In other words, sensitivity is the probability of testing positive if the event is truly present. Sensitivity was estimated as:

$$\text{Se} = \frac{\textit{True}\ \text{Positive}}{\text{True Positive} + \text{False Negative}}$$

High rates of sensitivity indicate that an administrative definition is not substantially underestimating the number of patients who fulfilled the eligibility criteria in EndoPro search.

4.7.2 Specificity

Specificity is the probability that an IBD diagnosis was not asserted by the administrative definition, given that it was not present in the gold standard. Specificity was defined as:

$$\text{Sp} = \frac{\textit{True}\ \text{Negative}}{\text{True Negative} + \text{False Positive}}$$

4.7.3 Kappa

The kappa statistic (κ) is another measure of overall agreement that is frequently used in the health services literature to summarize agreement between data sources. The kappa agreement is appealing because it is a single index of agreement that considers chance. That is, κ statistics adjusts the agreement for rare conditions by using the potential agreement beyond chance as a baseline. The definition of kappa is the n the agreement

obtained beyond that which would be expected by chance compared to the maximum possible agreement that could be obtained , simply,

$$K=\frac{observed.agreement - chance.agreement}{1 - chance.agreement}$$

However, interpretation of κ is not always straightforward because this statistic is affected by prevalence. For example, high levels of agreement between claims and hospital discharge summary data may emerge with low values of κ if the prevalence of the event of interest is low. In general, at extremes of prevalence κ tends to decrease for a fixed sensitivity and specificity.

A wide confidence interval for κ should serve as cautionary note to the reader and is often associated with disparate values of positive and negative e agreement proportions.

4.7.4 Predictive values

The probabilities of IBD and non-IBD patients on the administrative definition having or not having IBD were referred to as PPV or NPV, respectively. They were estimated as:

$$PPV = \frac{True\ \text{Positive}}{\text{True Positive} + \text{False Positive}} \qquad NPV = \frac{True\ \text{Negative}}{\text{True negative} + \text{False Negative}}$$

4.8 Chart reviews

As a complementary component for the study, we performed a comprehensive inpatient and outpatient chart review on selected individuals within the time period of 1995-2006. As each patient could have been admitted or seen in an outpatient clinic of different Calgary Health Region hospitals, all the charts within different Calgary Health Region

hospitals were reviewed for each patient. Within each hospital all the inpatient and outpatient chart volumes were reviewed. Charts were abstracted in a comprehensive abstraction form (Appendix 3).

Charts were reviewed for clinical, radiologic and pathologic features of IBD and diagnosis of UC or CD was based on standard Lennard-Jones criteria [42]. Patients were assigned into three major groups: 1) Definite case of IBD 2) Possible case of IBD 3) Definitely not IBD. Moreover, definite cases of IBD were also classified into 1) Definite CD 2) Definite UC 3) Indeterminate colitis.

We also documented the most responsible diagnosis, comorbidities and, if possible, the date of the diagnosis that allowed us to elaborate the circumstances surrounding discrepancies between EndoPro search and chart reviews.

Chart reviews were done on the following clusters of patients (Abstractor (AR) was blinded from which cluster the chart was drawn):

4.8.1 False positives

Considering the sensitivity and specificity of our different administrative definitions, the definition with the lowest acceptable specificity was chosen. Hospital chart numbers for FPs of the selected definition were identified and their charts were reviewed. The advantage of this approach was that the definition with the lowest acceptable specificity includes the FPs of the more specific definitions as well; minimizing the possibility to do further chart reviews in future.

4.8.2 False negatives

Considering the sensitivity and specificity of our different sets of administrative definitions, the definition with the lowest acceptable sensitivity was chosen. A random number of 35 PHNs were selected from the total number of FNs. Corresponding hospital chart numbers were extracted from EndoPro database and charts were reviewed.

The advantage of this approach was that the definition with the lowest acceptable sensitivity includes the FNs of the more sensitive definition as well.

4.8.3 True positives

To assess the reliability of EndoPro search in capturing IBD patients, we performed a random number of chart reviews (i.e. 33 patients) in the true positive group.

4.9 Integration of charts reviews with EndoPro search

In case there was a discrepancy in IBD status between EndoPro search and chart reviews, we considered the results of the chart reviews as the final decisive diagnostic tool. EndoPro IBD status was kept for patients whose charts were unavailable.

Number of discrepant cases in each cluster and the reason for the discrepancy were reported. A final data file containing the IBD status of our subjects was developed based on the final gold standard diagnostic tool (i.e. EndoPro search + Chart review).

4.10 Selection of the final administrative definition for IBD

Using the data file developed in section 4.9, all the analysis in section 4.7 was repeated and definition characteristics were calculated through the same methods. Selection of our

final two definitions was founded on a balance among specificity, sensitivity, results of chart reviews, and periodicity of the administrative definition (one definition does not integrate the ACSS database).

4.11 Ethical considerations

Ethical approval was received from the Office of Medical Bioethics at the University of Calgary (Appendix 4). This research was carried out according to Tri-council policy statement: the ethical conduct for research involving humans [43].

Data were presented in aggregate form and individual patients were not identified. Individual patients were never contacted and their care was not influenced by any means by this research.

Chapter Five: RESULTS

5.1 Classification of colonoscopies within EndoPro database

In the time period of May 16th, 2000 through April 1st, 2004, a total of 23,527 colonoscopies were conducted in the Calgary Health Region and recorded in EndoPro. As shown in section 4.3, a computerized and a manual search were conducted to classify procedures:

- Definite IBD n = 2,875
- Possible IBD n = 2,469
- Non-IBD n = 18,181

As shown in Figure 1, indication fields were more helpful to screen procedures for a definite diagnosis of IBD. Moreover, as expected there was considerable overlap among indication, findings, and impression screenings. Of 2,875 definite IBD, procedures 292 had findings or an impression of IBD without any indication of IBD, these might represent procedures for incident cases.

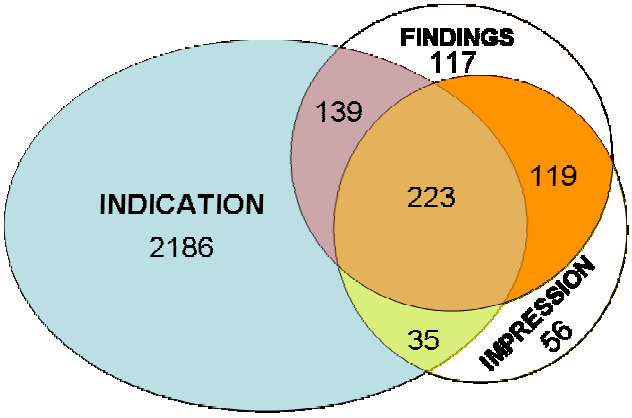


Figure 1. Distribution of the procedures that were considered as IBD among indication, findings, and impression fields. Numbers correspond to the zones with the same color. Total number of the procedures considered as IBD by indication, findings, and impression were 2583, 598 and 433, respectively.

Distribution of the procedures that were considered as "possible IBD" was different from "definite IBD" procedures; only a small number of procedures were captured through indication search (n=11) while findings (n=2,164) and impression (n=2,294) fields contributed the most.(Figure 2)

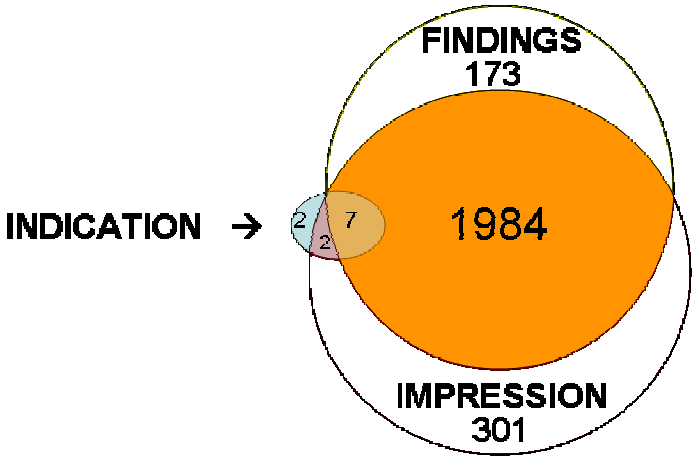


Figure 2. Distribution of the procedures that were considered as "possible IBD" among indication, findings, and impression fields. Numbers correspond to the zones with the same color.

5.2 Linkage with the registry

A data file with the necessary variables (refer to section 4.4) was developed for the EndoPro procedures (i.e. 23,527 colonoscopies for 21,193 patients). As two or more of the four linkage variables (i.e. first and last name, date of birth, and gender) were missing in some procedure records; they were removed from the linkage process (i.e. 4.5% of the procedures). 17,699 subjects were successfully assigned a valid PHN in the Calgary Health Region registry (83.5%). Of 17,699 subjects, 16,717 (94.45%) were tagged as active in the registry in fiscal year 2003/04.

5.3 Classification of the subjects

Based on the criteria presented in section 4.5 , we classified each patient (i.e. each PHN) into one of the following categories:

- Definite IBD n = 1,353
- Possible IBD n = 840
- Non-IBD n = 15,506

Of 15,506 non-IBD subjects, 5,204 underwent colonoscopy for colorectal cancer screening (33.6%).

5.4 Linkage to the Calgary Health Region administrative data

Using the PHNs, health service records of our patients were traced in IP, PC and ACCS databases.

5.4.1 IP database (FY 1994 – FY 2004)

In total 29,886 hospitalization records were found for 9,162 patients. Annual number of admissions showed a dramatic increase since 1994 (Figure 3). The same pattern was observed when only hospitalizations with a diagnosis of IBD were considered. (Figure 4)

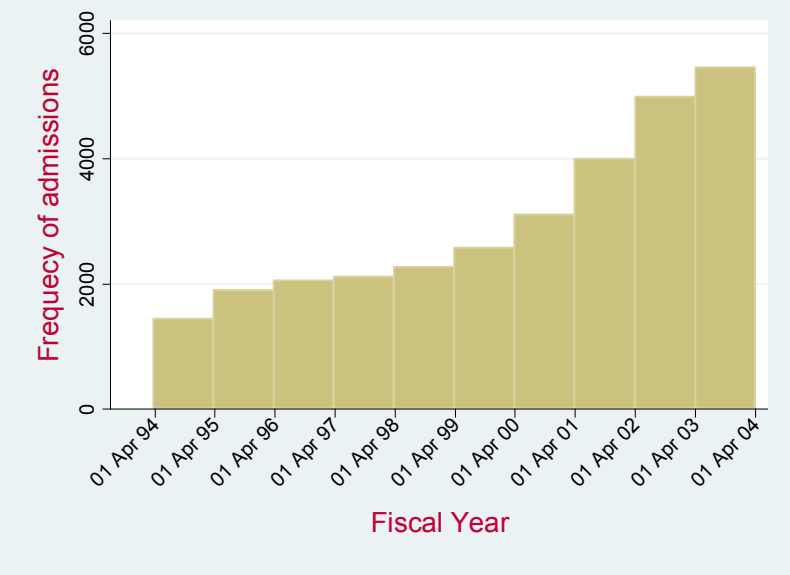


Figure 3. Annual increase in the number of admissions among our cohort of patients irrespective of diagnosis.

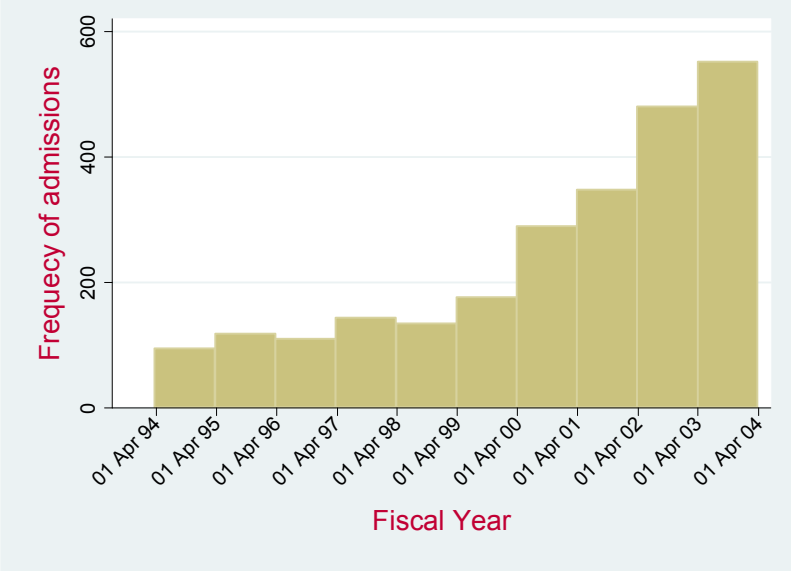


Figure 4. Annual increase in the number of admissions with a diagnosis of IBD

Number of subjects with at least one hospitalization and frequency of IBD diagnostic codes in hospital discharge abstracts among IBD and non-IBD patients are shown in Diagram 1.

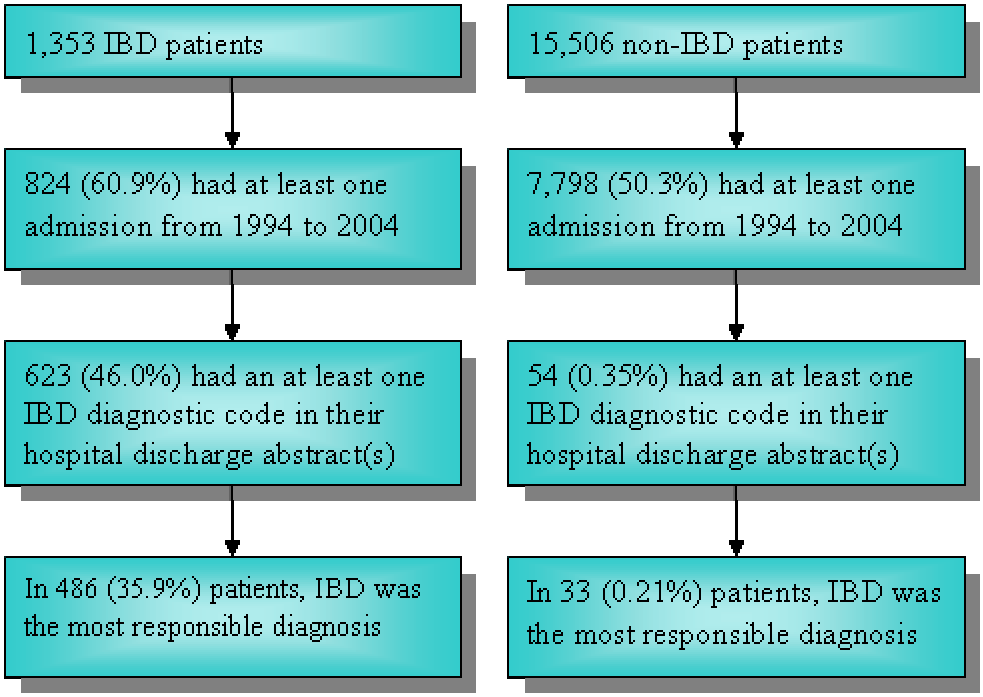


Diagram 1. Number of admissions and IBD diagnostic codes in IBD and non-IBD patients.

IBD patients had a median number of two hospitalizations with a diagnostic code of IBD within the period of 1994-2004 (Mean, 2.5±3.1; range, 0-33). Within the same period IBD patients had a median number of 1 hospitalization with a most responsible diagnosis of IBD (Mean, 1.4±2.1; range, 0-20).

5.4.2 PC database (FY 1995 – FY 2004)

In total 2,386,552 claims were found for the cohort of patients from 1995 to 2004. Of 17,699 subjects, 570 did not have any physician claims. Of 2,386,552 claims, 310,990 claims did not have diagnostic codes and were removed from the analysis (i.e. radiology claims). In total, 28,586 claims were identified with an ICD code of IBD (24,444 claims in IBD patients, 2,575 in possible IBD cases, and 1,567 claims in non-IBD patients) (Diagram 2 and Diagram 3).

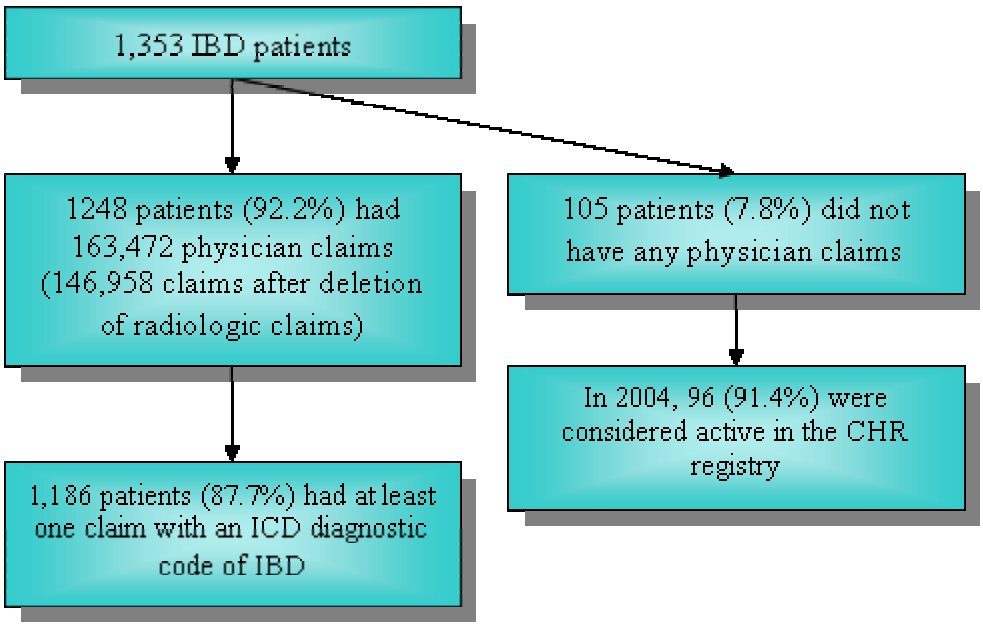


Diagram 2. Frequency of physician claims and IBD diagnostic codes among IBD patients.

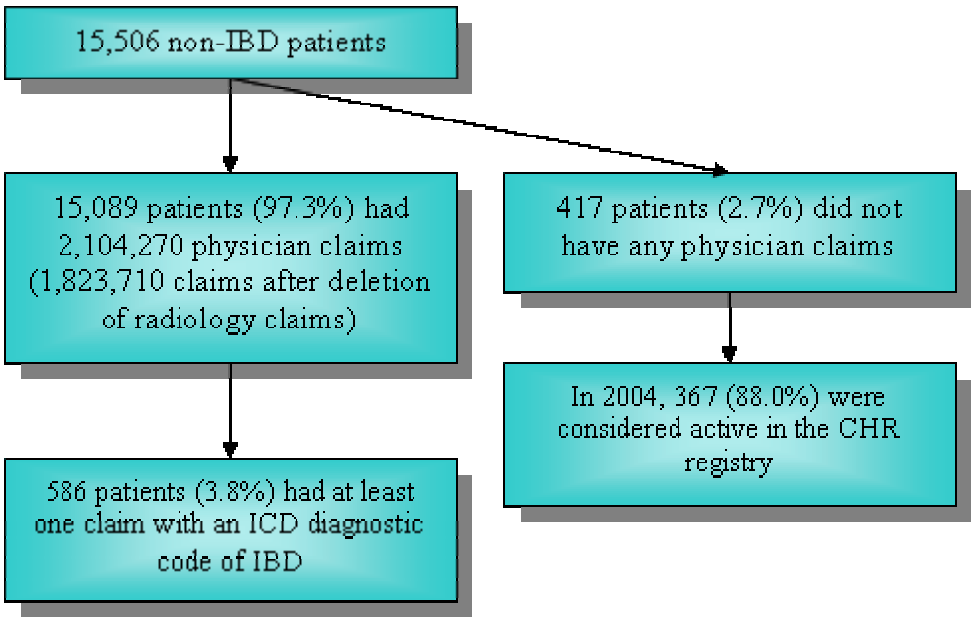


Diagram 3. Frequency of physician claims and IBD diagnostic codes among non-IBD patients.

IBD patients had a median number of 12 claims with a diagnostic code of IBD within the period of 1995-2004 (Mean, 19.4±23.4; range, 0-240). Within the same period, non-IBD patients had a median number of 0 claims (95^{th} percentile = 0) with a diagnostic code of IBD (mean, 0.1±1.5; range, 0-138).

5.4.3 ACCS database (FY 1996 – FY 2004)

In total 240,142 ACCS contacts were found for 17,112 (96.7%) patients. 7,246 contacts had a diagnostic code of IBD. Distribution of ACCS contacts is shown in Diagram 4.

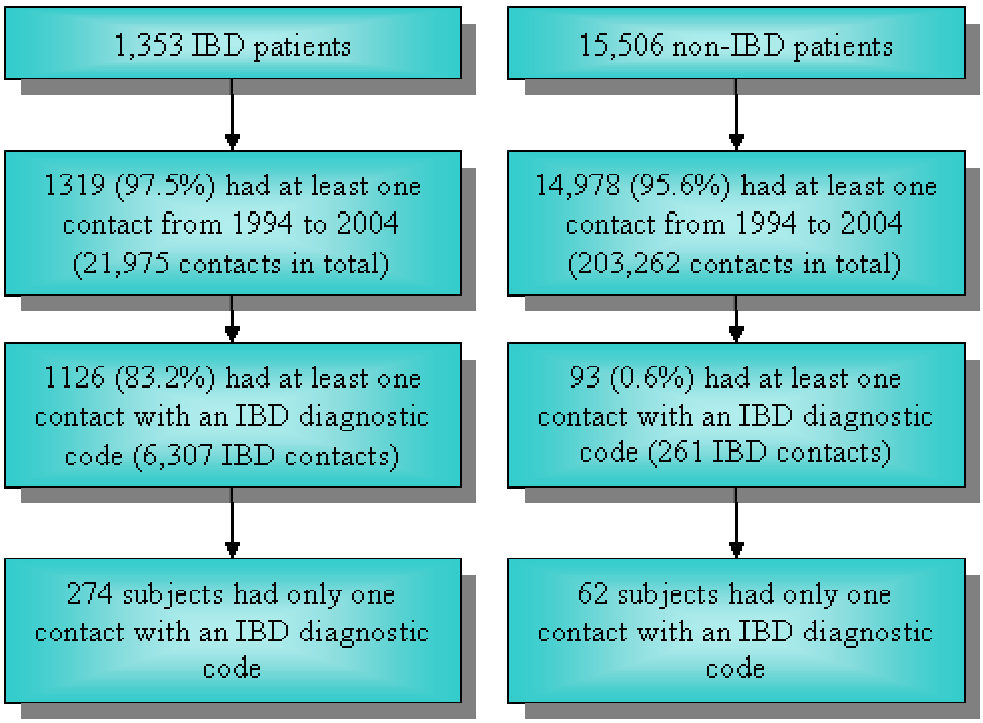


Diagram 4. Frequency of ACCS contacts and IBD diagnostic codes among IBD and non-IBD patients.

Although all of our patients had undergone colonoscopy, 3.3% of them did not have a contact in the ACCS database. This might be due to colonoscopies performed during the hospitalization periods, which are not recorded in ACCS database. IBD patients had a median number of two ACCS contacts with a diagnostic code of IBD within the period of 1996-2004 (Mean, 4.8±7.6; range, 0-87). Within the same period, non-IBD patients had a median number of 0 claims (99^{th} percentile = 0) with a diagnostic code of IBD (Mean, 0.005±0.1; range, 0-5).

The top eleven procedures that held a diagnosis of IBD are shown in Table 2.

Table 2. Cumulative percentage and percentage of the top 11 procedures that had a diagnostic code for IBD in ACCS database.

Procedure	Percentage	Cumulative percentage
Closed endoscopic biopsy of large intestine	11.52	11.52
Endoscopy via orifice or stoma with biopsy	10.41	21.93
Assessment, examination consult	10.12	32.05
Missing (no procedure code)	9.37	41.42
Endoscopy via orifice or stoma	7.87	49.28
Injection or infusion of other therapeutic or prophylactic substances	5.48	54.76
Intravenous line	3.74	58.50
Endoscopy small intestine with biopsy	3.38	61.88
Colonoscopy	2.99	64.88
Endoscopy with rectal biopsy	2.43	67.31
Intravenous iron injection	1.96	69.27

5.5 Preliminary administrative definitions for IBD

Based on the results in section 5.4, 100 definitions including ACCS database and 50 definitions without ACCS were developed. Sensitivity and specificity of these definitions are summarized in appendices 5 and 6.

Definition "IP≥1 *or* PC≥1 *or* ACCS≥1 within 10 years" (i.e. any IBD diagnosis in any database at any time equals being an IBD patient) was the most sensitive definition we could define (refer to Table 3); meaning that the highest sensitivity that we could achieve with our data was 96%. (Table 3)

Table 3. 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥1 *or* ACCS≥1 within 10 years"

		Administrative definition		
		IBD	Non-IBD	
EndoPro search (Gold standard test)	IBD	1,343	56	Sn: 96.00% (95%CI=94.83,96.96)
	Non-IBD	576	14,863	Sp: 96.27% (95% CI=95.96,96.56)
				PPV: 69.98%(95%CI=67.88,72.03)
				NPV:99.63%(95%CI=99.51,99.72)

Considering the definition characteristics, definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" (refer to Table 4) was considered to have the lowest acceptable specificity and definition "IP≥1 *or* PC≥5 *or* ACCS (not considered) within 2 years" (refer to Table 5) was considered to have the lowest acceptable sensitivity.

Table 4. 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years"

		Administrative definition		
		IBD	Non-IBD	
EndoPro search (Gold standard test)	IBD	1,212	141	Sn: 89.58% (95%CI=87.83,91.16)
	Non-IBD	137	15,369	Sp: 99.12% (95% CI=98.96,99.26)
				PPV: 89.84%(95%CI=88.11,91.41)
				NPV:99.09%(95%CI=98.93,99.23)

Table 5. . 2×2 contingency table and characteristics for definition "IP≥1 *or* PC≥5 *or* ACCS (not considered) within 2 years"

		Administrative definition		
		IBD	Non-IBD	Sn: 72.65% (95%CI=70.19,75.01)
EndoPro search (Gold standard test)	IBD	983	370	Sp: 99.63% (95% CI=99.52,99.72)
	Non-IBD	57	15,449	PPV: 94.52%(95%CI=92.96,95.82)
				NPV:97.66%(95%CI=97.41,97.89)

Corresponding 2×2 tables for 23 definitions that were more sensitive than "IP≥1 *or* PC≥5 *or* ACCS≥2 within 2 years" and also more specific than "IP≥1 *or* PC≥3 *or* ACCS≥2 within 2 years" are shown in Appendix 7. In other words, FPs for all these 24 definitions fell under the FPs for the definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" and all the FNs fell under the FNs of definition "IP≥1 *or* PC≥5 *or* ACCS (not considered) within 2 years".

Twelve of these definitions considered ACCS database (i.e. IP≥1 *or* PC≥3-5 *or* ACCS≥2 within 2-5 years) while the other 12 did not consider ACCS database (IP≥1 *or* PC≥3-5 within 2-5 years).

5.6 Chart reviews

5.6.1 False positive subjects

Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" had the lowest specificity (i.e. highest FPs) among the 24 preferable preliminary definitions (Appendix 7); thus, FPs for the other 23 definitions are a subgroup of FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2

within 5 years". Of 137 false positive subjects, inpatient and clinic charts for 125 (91.2%) subjects could be retrieved[2].

Of 125 subjects, in agreement with EndoPro diagnosis, 58 patients did not have IBD. These were generally patients with GI problems other than IBD who were miscoded as IBD in administrative databases. However, non-GI patients and even healthy individuals started to show up as the definition became more non-specific (refer to Table 6).

Of these 58 subjects, 21 and 33 were considered as IBD by definitions "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" and "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years", respectively.

Table 6. Diagnoses made by chart reviews for 58 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were proven not to have IBD. Whether or not theses cases are considered as IBD by more specific definitions (i.e. definition A "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" or definition B "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years") is also shown.

Chart diagnosis	Frequency	Considered IBD by definition A (21 patients)	Considered IBD by definition B (33 patients)
Colorectal cancer ± bowel resection	4	+	+
Irritable Bowel Syndrome	2	+	+
Diverticulitis + bowel resection	2	+	+
Diverticulosis	2	+	+
Diverticulosis + Irritable Bowel Syndrome	1	+	+
Microscopic colitis	1	+	+
Clostridium difficile colitis + Overflow diarrhea due to ALS (UC was in differential diagnosis at the beginning)	1	+	+

[2] Unavailable charts were coded as CF (Can't find), which is based on Health Records Services experience is usually due to mis-shelving.

Bladder cancer with vaginal fistula + large bowel obstruction due to adhesions	1	+	+
ERCP induced duodenal perforation and pancreatitis + small bowel obstruction due to adhesions	1	+	+
Collagenous sprue	1	+	+
Chronic abdominal pain syndrome	1	+	+
Undiagnosed chronic diarrhea (IBD ruled out)	1	+	+
Gastric ulcer + iron deficiency anemia	1	+	+
Diabetes type II + adenovillous polyp + phimosis	1	+	+
Ethanol abuse	1	+	+
Radiation proctitis	4	-	+
Irritable Bowel Syndrome	3	-	+
MALT type B cell lymphoma of colon	1	-	+
Idiopathic chronic pancreatitis	1	-	+
Ischemic colitis	1	-	+
Adenomatous polyp + alcoholic gastritis	1	-	+
Adenomatous polyp	1	-	+
Appendiceal phlegmon due to actinomyces	1	-	-
Clostridium difficile colitis	2	-	-
Infectious colitis with yersinia	1	-	-
Radiation proctitis	1	-	-
Colorectal cancer	1	-	-
Undiagnosed diarrhea and abdominal pain (IBD ruled out)	2	-	-
Collagenous colitis	1	-	-
Celiac	1	-	-
Irritable Bowel Syndrome	2	-	-
Anemia + diverticulosis (IBD ruled out)	1	-	-
GI bleeding due to peptic ulcer disease	1	-	-
NSAID induced GI bleeding	1	-	-
Erosive duodonitis + polyposis + + Hepatitis C + coronary artery disease + interstitial cystitis + low GFR	1	-	-
Ethanol induced cirrhosis	1	-	-
Rectal bleeding due to internal hemorrhoids	3	-	-
Rectal bleeding due to anal fissure	1	-	-
familial adenomatous polyposis	1	-	-
Adenomatous polyp	2	-	-
Healthy individual	1	-	-

Of 125 FPs, 21 were diagnosed as possible IBD patients. Theses were the cases that IBD was never ruled out. Of these 21 subjects, 18 were considered as IBD by definitions

"IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" and "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years" (refer Table 7).

Table 7. Diagnoses made by chart reviews for 21 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were shown that might have IBD. Whether or not theses cases are considered as IBD by more specific definitions (i.e. definition A "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" or definition B "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years") is also shown.

Chart diagnosis	Frequency	Considered IBD by definition A (18 patients)	Considered IBD by definition B (18 patients)
Undiagnosed chronic diarrhea and abdominal pain	5	+	+
Pending diagnosis between IBD and IBS	4	+	+
Undiagnosed colitis	3	+	+
Nonspecific chronic inflammation in pathology + abdominal pain + normal colonoscopy	2	+	+
Pending diagnosis between IBD and microscopic colitis	2	+	+
Possible right sided CD but colonoscopy unsuccessful	1	+	+
Clinical IBD + abdominal CT and small bowel follow through suggestive of CD + normal colonoscopy	1	+	+
Abnormal liver function tests + Possible primary sclerosing cholangitis on MRI + visible lymphoid hyperplasia in terminal ileum with normal pathology	1	-	-
Chronic rectal fistula + recurrent anal abscess + inflammatory polyp in 2001+ history of tuberculosis	1	-	-
Abdominal pain + blood in stool (terminal ileum could not be intubated)	1	-	-

Of 125 FPs, 46 patients were found to have IBD in their charts. The majority (n=43) were IBD cases who underwent colonoscopy or pouchoscopy for colorectal cancer screening or disease activity with normal colonoscopy results, but unfortunately the colonoscopist failed to mention the previous diagnosis of IBD in the indication field of EndoPro

database. In one case, the patient had a normal colonoscopy, despite clinical manifestation of CD and later pathology revealed CD (tagged by * in Table 8). Three cases were wrongfully classified as non-IBD, during the manual search on EndoPro database (Section 4.3.2).

Of these 46 subjects, 36 and 42 were considered as IBD by definitions "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" and "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years", respectively. (Table 8)

Table 8. Diagnoses made by chart reviews for 46 FPs for definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" who were shown that do have IBD. Whether or not theses cases are considered as IBD by more specific definitions (i.e. definition A "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years" or definition B "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years") is also shown.

Chart diagnosis	Frequency	Considered IBD by definition A (36 patients)	Considered IBD by definition B (42 patients)
CD	9	+	+
UC	8	+	+
CD + bowel resection	8	+	+
UC + bowel resection	2	+	+
IC	5	+	+
UC + colorectal cancer + bowel resection	1	+	+
CD$^{\phi}$	1	+	+
UC (Mistakenly classified as non-IBD in manual EndoPro search)	1	+	+
CD (Mistakenly classified as non-IBD in manual EndoPro search)	1	+	+
CD + bowel resection	2	-	+
UC	2	-	+
UC (Mistakenly classified as non-IBD in manual EndoPro search)	1	-	+
CD	1	-	+
UC	4	-	-

$^{\phi}$ Normal colonoscopy, despite clinical manifestation of CD and later pathology revealed CD

5.6.2 False negative patients

Definition "IP≥1 *or* PC≥5 within 2 years" had the lowest sensitivity (i.e. highest FNs) among the 24 preferable preliminary definitions; therefore, FNs for the other 23 definitions are a subgroup of FNs for definition "IP≥1 *or* PC≥5 within 2 years". From 370 FNs for the definition "IP≥1 *or* PC≥5 within 2 years", a random sample of 35 (9.5%) patients was drawn. Inpatient and clinic charts for 32 (91.4%) patients could be retrieved and abstracted. All the patients were found to have IBD in their charts. The majority of patients (87.5%) were old cases of IBD with no serious flares within the period of 1995-2004; thus, they lacked health care contacts in the administrative databases. 12.5% of patients were newly diagnosed cases of IBD that did not have enough time to have enough administrative contacts to meet our definition.

Table 9. Chart diagnoses for 32 random false negative patients.

Chart diagnosis	Frequency	Percentage
Old case of UC	14	43.75
Old case of CD	11	34.37
New case of UC	2	6.25
New case of CD	2	6.25
Old case of IC	2	6.25
Old case of UC with bowel resection	1	3.13

5.6.3 True positive patients

Among our 24 preferable preliminary definitions, definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" is the most sensitive definition (highest TP). A random sample of 32

patients was drawn from 1,212 TPs. Charts for 29 patients could be retrieved, which all of them had the diagnosis of IBD (8 CD, 8 CD + bowel resection, 8 UC and 5 UC + bowel resection).

If we add up the results of chart reviews for TPs and FNs, 61 out of 61 (29+32) patients with a definite IBD diagnosis through EndoPro search actually had IBD.

5.7 Final classification of the subjects

Using the results of the chart reviews, subjects were reclassified as described in section 4.9:

- Definite IBD n = 1,399
- Possible IBD n = 861
- Non-IBD n = 15,439

Age and gender of the IBD and non-IBD patients are shown in Table 10.

Table 10. Demographic characteristics of our IBD and non-IBD subjects

	Number of subjects	Gender	Age ± SD (As in Apr 2004)
Definite IBD patients	1,399	F = 741 (53.0%) M = 658 (47.0%)	45.00±14.67 (17-89)
Definite non IBD individuals	15,439	F = 8,374 (54.2%) M = 7,065 (45.8%)	58.95±14.24 (17-99)

All the analysis in section 4.6 was repeated with the final classification. An update is provided in Appendix 8..

2×2 contingency tables and test characteristics were again calculated for 150 different definitions through our computerized algorithm. (Appendices 5 to 7) The following

definitions were selected as our final administrative definition to capture IBD patients through administrative databases:

1. IP≥1 *or* PC≥4 *or* ACCS≥2 within 3 years
2. IP≥1 *or* PC≥4 within 3 years

5.8 Consistency of the final definition over age and gender

Characteristics of the final definition were examined on different age and gender groups (Table 11).

Table 11. Final definition characteristics over gender and age

	Sensitivity	Specificity	PPV	NPV
Sex		11.52		
Male	85.3 (82.3-87.9)	99.7 (99.6-99.8)	96.9 (95.1-98.1)	98.6 (98.3-98.9)
Female	86.1 (83.4-88.5)	99.7 (99.6-99.8)	96.8 (95.2-98.0)	98.8 (98.5-99.0)
Age				
18-34	90.1 (86.9-92.8)	99.4 (98.7-99.8)	98.7 (97.0-99.6)	95.4 (93.9-96.7)
35-60	85.4 (82.8-87.8)	99.8 (99.7-99.9)	97.4 (95.9-98.5)	98.6 (98.4-98.9)
+60	76.8 (70.0-82.6)	99.7 (99.6-99.9)	89.9 (84.1-94.1)	99.3 (99.1-99.5)

Chapter Six: DISCUSSION

Health care administrative databases are being used in many epidemiologic studies including studies of resource consumption, incidence and prevalence estimations, and outcome assessment. Previous studies have evaluated the accuracy of the data elements within several Canadian administrative databases by assessing the agreement with data from charts or other electronic databases. These studies showed that demographic information, primary diagnoses, and some procedural codes were accurately recorded [31, 44].

Other than one study conducted by Bernstein et al.[2], that used administrative databases of inpatient hospital abstracts (IP) and physician claims (PC), no other study has developed a population-based database for IBD patients in Canada. In Alberta, we have the advantage of having an endoscopy database and a database for hospital ambulatory care clinics (ACCS). Using these two databases in addition to IP and PC databases, we have developed an administrative definition to identify IBD patients.

6.1 Study subjects

Of 21,193 patients categorized by EndoPro search, 3494 patients (16.5%) did not receive a valid PHN from AHCIP registry (refer to section 5.2). We speculated the following reasons for our inability to link these patients with the registry; however, further studies should be conducted to elaborate this issue:

- Incomplete documentation of linkage variables (i.e. first and last name, date of birth, and gender) in EndoPro database (Approximately responsible for a quarter of the cases). Demographic data is entered by hand by clerks, nurses and sometimes physicians in case of emergency cases.
- Colonoscopies done on patients residing outside of the Calgary Health Region.
- Data linkage limitations due to use of non-unique linkage variables.

Our IBD patients had a mean age of 45 years, which is in agreement with previous studies that reported a peak prevalence occurring among those aged 30-49 years [2, 45, 46]. Non-IBD patients were older than IBD patients (mean age of 59 years) probably due to the fact that one-third of our non-IBD patients underwent colonoscopy for colorectal cancer screening which is usually done in patients older than fifty years of age [47].

Female to male ratio of our IBD patients was 1.13, which is again with agreement with previous studies that reported a slight female predominance in IBD [8]. Higher number of females in our non-IBD subjects can be explained by higher life expectancy of women in Canada [48].

6.2 Separate linkage to administrative databases

6.2.1 IP database

Our data showed a dramatic increase in the number of the hospitalizations since 1994 (Figure 3).This phenomenon could have been due to the following reasons:

- High growth rate of the Calgary Health Region [49].

- We captured the patients, both incident and prevalent cases, through the period of 2000 to 2004; therefore, we had fewer hospitalizations prior to 2000 as our incident cases did not have clinical IBD at that time.

As the time of the diagnosis of IBD is not clear in our IBD patients, we cannot comment on admission rates or risk of hospitalization in our patients or compare them to other studies.

IP database proved to be exceptionally specific to capture IBD patients (Appendix 8). Only 13 out of 7,778 hospitalizations in non-IBD patients were wrongfully coded as IBD. However, as expected the sensitivity was not acceptable because not all IBD patients are hospitalized within a 10 year period. Interestingly, only three fourth of IBD patients who were hospitalized were coded as IBD (644 out of 854) which undermines the sensitivity of IP database to capture IBD patients. A proportion of this deficiency in sensitivity could have been due to the hospitalizations that occurred prior to diagnosis of IBD.

6.2.2 PC database

Surprisingly, 7.5% of IBD patients had no physician claims although we knew they are/have been residents of the Calgary Health Region and had at least one colonoscopy within the period of 2000 to 2004 (Appendix 8). This might have been due to the following reasons:

- A flaw in AHCIP registry that led to inclusion of non-residents of the Calgary Health Region in our data, so that we cannot trace their claims in the Calgary Health Region's PC database. However, such a flaw can be rejected because

94.5% of our subjects were regarded as active in fiscal year 2003/04 and of those IBD patients who did not have any claims 91.4% were active in this year.

- Shadow billing (physician practices that are reimbursed through alternative payment plan). When physicians are paid automatically regardless of the performed procedure, inaccurate and incomplete claims might be submitted.

Moreover, 4.6% of IBD patients did not have any claims with an IBD diagnostic code even though they had claims in PC database. This can be explained by non-specific coding by physicians and shadow billing. A study in Ontario has shown that 8.6% of the physician who see IBD patients do not use specific ICD-9 codes of 555 (CD) or 556 (UC) [50].

The fact that approximately 12.1% of IBD patients did not have any claim or a claim with an ICD code of IBD underscores the inability of PC database, if used alone, to detect IBD cases.

Of 15,439 non-IBD patients, 524 had at least one claim with an IBD diagnostic code (3.4%) showing that ICD codes for IBD are sometimes used to submit a claim although the patient has no documented IBD (Appendix 8). Therefore, relying on PC database as the sole source to capture IBD patients might yield a large number of false positives. To further investigate the reason behind these miscodings we recommend additional chart reviews.

6.2.3 ACCS database

ACCS database was the most sensitive of our databases (96.7%). This was expected as all of our patients had undergone colonoscopy and the 3.3% gap might have been due to the

colonoscopies performed within the hospitalization periods which do not get recorded in ACCS database.

82.3% of IBD patients had at least one contact with an IBD diagnostic code (62% had more than one contact with an IBD code). Considering the fact that these codes are generated from generally robust diagnostic procedures (Table 2), ACCS database proves to be a valuable complementary database to capture IBD patients.

Of 15,439 non-IBD patients, 57 had a contact with diagnosis of IBD but only 10 had two or more contacts coded as IBD (all proved to be wrongfully coded as IBD through chart reviews). This was the reason why we did not consider a patient with a single IBD code in ACCS to have IBD.

6.3 Reasoning behind the selection of the final definition

Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 3 years" was selected as our final administrative definition based on its sensitivity, specificity, and the results of the chart reviews. Sensitivity and specificity were preferred over predictive values to compare our definitions because:

- Unlike sensitivity and specificity, predictive values are dependent on the pre-test probability (i.e. prevalence) of the disease[3]. There is no exact estimate about the

---

[3] $\text{PPV}=\frac{Sn\times \text{Pr}}{Sn\times PR+(1-Sp)\times(1-\text{Pr})}$, $\text{NPV}=\frac{Sp\times(1-\text{Pr})}{Sp\times(1-\text{Pr})+(1-Sn)\times \text{Pr}}$

prevalence of IBD in the Calgary Health Region but applying Bernstein's criteria to the province of Alberta yields an approximate prevalence of 0.5% [1]. The prevalence of the disease in our study sample is 8.3% ($\frac{1399}{1399+15439}$) because our subjects have undergone colonoscopy and consequently have a higher chance of GI disorders. This means that our reported PPVs are overestimated and NPVs are underestimated when they are generalized to the population.

- The goal of this study is to develop a definition that identifies IBD patients. In other words we need a definition that includes IBD patients (Sn) and excludes non-IBD subjects (Sp). Post-test predictive measures cannot contribute to these two situations.

Level of agreement between gold standard test and administrative definition proved to be significantly high using kappa statistics (i.e. 0.9); however, it should be noted that application of kappa statistics might not be rational because kappa is conventionally used to compare non-gold standard tests.

6.3.1 One contact in IP database

Comparison of sensitivities and specificities of the definitions that only differ in the number of the IBD diagnostic codes in IP database (e.g. Page 3 vs. 8 in Appendix 9) shows that when we increase the number of the IP contacts from 1 to 2, specificity

increases by less than 0.05% at the expense of more than a 1% drop in sensitivity. In addition we have shown that IP database, if used alone, is 99.9% specific for IBD (Appendix 8), so increasing the ICD code limit for capturing IBD patients will not substantially enhance the validity of our definition.

6.3.2 Four claims in PC database

Generally, sensitivity of our administrative definitions fell to unacceptable levels (<80%) with 6 claims; conversely; specificities fell short of acceptable levels (<99%) with 2 claims.

Definitions that incorporated 3, 4, or 5 claims yielded reasonable characteristics; however, we rejected definitions with 3 claims for the following reasons:

- Table 6 shows the actual medical conditions for subjects who were mistakenly considered as IBD by our administrative definitions (FPs). More than half of FPs exclusive to definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" were patients with GI bleeding or abdominal pain unrelated to colitis (e.g. haemorrhoids), or even healthy individuals. The validity of our definition could be seriously affected if common diseases that share certain common symptoms with IBD are considered as IBD in our database.
- Table 8 shows 46 of our preliminary FPs actually had IBD. Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years" captures 42 of them while definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years" only adds four extra cases.
- The only disadvantage of moving from 3 to 4 claims is that we might sacrifice some sensitivity but in Table 9 we have shown that our FNs are old IBD cases

with no flare in the last 10 years or new cases with limited time to fulfill our definition. As most of the old cases will eventually flare and new cases will have more IBD-related health care contacts, by updating the database we will be able to identify some of these patients and compensate the sacrificed sensitivity. In contrast, future updates will slightly decrease or at best maintain the specificity of our definition; therefore, to be on the safe side it is logical to select the more specific definition (i.e.4 claims).

We also rejected the definitions with 5 claims because of the sharp drop in their sensitivity ($\approx$4%) and limited in specificity ($\approx$0.07%). In addition chart reviews failed to show a remarkable advantage for 5 over 4 claims. (Table 6-9)

6.3.3 Two contacts in ACCS database

The logic not to consider a subject with a single ACCS contact is discussed in section 6.2.3. Comparison of sensitivities and specificities of definitions that only differ in the number of the IBD diagnostic codes in ACCS database reveals that when we raise the number of the ACCS contacts from 2 to 3, specificity increases by less than 0.03% at the expense of 1-2% decrease in sensitivity.

6.3.4 Periodicity (3 years)

During the analysis, a range of 1 to 5 years was chosen to develop the definitions. Although our computerized algorithm could accept non-integer inputs (e.g. 0.5 year), a year was considered as the lower limit. Periods less than a year might miss well-controlled IBD cases and also be dependent on periodical flares in IBD patients. Periods

longer than 5 years were not practically feasible as the availability of the administrative data is sometimes limited for long periods of time in different health regions.

Referring to Appendix 9, when the definitions were defined within a period of 1 or 2 years they missed a significant number of patients relative to 3-5 years. Moreover, the gain in specificity was not remarkable.

There was not a noteworthy difference among 3, 4 and 5-year periods in terms of specificity and sensitivity. Therefore, we chose 3 years as the final time frame because it was shorter than 4 and 5-year periods. In addition, we showed that the majority of physician claims for IBD are usually less than one year apart (Page 37); knowing the fact that our time periods start with an IBD code, 4 physician claims of our final definition can conveniently fit within a 3-year period.

6.3.5 Alternative definition for health regions with no access to ACCS database

ACCS is not available outside the province of Alberta; therefore we propose the definition "IP≥1 or PC≥4 within 3 years" for non-Albertan health regions. Sensitivity of this definition is 5.57% less than "IP≥1 *or* PC≥4 *or* ACCS≥2 within 3 years" and the specificity is slightly higher by 0.04%.

6.4 Comparison of our administrative data and definition with other IBD administrative databases

As mentioned before, there is only one other study reporting the development of an administrative IBD database in North America [2]; however, several limitations are evident in this study that we tried to overcome in ours:

- At least two independent and separately-collected data sources are required to develop an administrative definition. One database (could be a non-administrative database) for collecting a significant number of healthy and diseased subjects and at least one other administrative database for investigation of the health care history of the subjects. In addition a third separate and independent database (could be a non-administrative database) is needed to validate the administrative definition. In the Manitoba study, only one set of administrative data (two dependent databases of IP and PC) was used for capturing the primary cohort of IBD patients, development of administrative criteria, and criteria validation. In other words, the ICD coding definition was developed on patients who all had IBD diagnostic ICD codes in their health service records. In our study we had the opportunity to capture IBD patients through an independent colonoscopy database, and then develop our definition based on administrative databases.
- The Manitoba study used mailed self-administered questionnaires (response rate of 60%) plus random chart reviews on 16% (448 charts) of questionnaire respondents as their gold standard. Apart from the general biases accompanied with such surveys; to our knowledge, the validity for self-reported questionnaires in diagnosis of IBD has not been described yet. In our study we used colonoscopy reports plus complementary chart reviews as our gold standard test that provide more accurate and reliable information regarding the diagnosis of IBD.

- In the Manitoba study random chart reviews were done only on questionnaire respondents who had a history of hospitalization (had a record in IP database): 1-This was not a representative sample of questionnaire respondents as they have a more severe course of disease relative to those who did not have a history of hospitalization. 2-Those who get hospitalized are more aware of their disease status (recall bias), so they might have achieved a falsely high agreement between questionnaires and chart reviews (the agreement percentage is not reported). 3-This stipulates the fact that the value of the physician claims database was not verified by the chart reviews and only IP database contacts could have been re-evaluated. In our study we had access to both hospital and outpatient clinic charts.
- As IBD is considered a rare disease, a large number of documented non-IBD patients (TNs) are needed to estimate the specificity of the administrative definition. The Manitoba study had only access to about 250 non-IBD patients that were not representative of non-IBD subjects because they had at least 3 IBD contacts (184 subjects self reported that they do not have IBD and number of the non-IBD subjects found through chart reviews was not reported but could not have been more than 50 based on the sample false positive rate of 10%). Therefore, at best PPV (i.e. $\frac{TP}{TP+FP}$) or *sample false positive rate* of the administrative criteria could have been calculated and there would not have been enough information to calculate the specificity (i.e. $\frac{TN}{TN+FP}$) nor false positive

rate (i.e. 1-specificity or $\frac{FP}{TN+FP}$). The 99.975% specificity reported in their recent study [1] (that as described above is theoretically unattainable), is an extrapolation of their "false positive rate among those with three or more IBD contacts" to the whole population of Manitoba; however, this extrapolation is conditional on an invalid assumption that all those without an IBD health-care contact do not have IBD and are *truly negative*. Our findings show that 10.4% of IBD patients do not have a contact in IP or PC databases within a period of 10 years[4]. This error is not present in our study as we have documented non-IBD patients (true negatives) through EndoPro database, thus we were able to calculate the specificity of our definition.

- The 90% sensitivity of the Manitoba study, is reliant on the assumption that whoever has less than three IBD claims or hospitalization does not have IBD; however, based on our study findings this assumption is not valid as approximately 16.2% of IBD patients fall in that category[5], suggesting that the sensitivity of the Manitoban study may have been overestimated. Applying a set of criteria similar to their administrative criteria on our cohort of IBD patients yields a sensitivity of 70.91% (95%CI: 68.45%-73.28%)[6]. No primary assumptions for IBD contacts were needed in our study because EndoPro database is abstracted separately from administrative databases.

---

[4] In our study, definition "IP≥1 **OR** PC≥1 in 10 years" fails to capture 145 of our IBD patients (10.4%).

[5] In our study, definition "IP+PC≥3 in 10 years" fails to capture 226 of our IBD patients.

[6] IP+PC≥5 in 10 years

- Bernstein et al. used Youden's index [51], which gives equal weight to the sensitivity and specificity[7]. A far larger weight should be given to specificity when we are dealing with rare diseases. As an example if we have 5000 IBD patients in a population of 1,000,000; a 1% decrease in specificity will falsely classify 10,000 healthy individuals as IBD patients whereas a 1 % decrease in sensitivity will only fail to capture 50 IBD patients. In our study, our primary goal was to achieve a high specificity.
- In the Manitoba study the same weight was applied to health care contacts in IP and PC database. Our study shows that IP and PC databases, if used alone, have significantly different sensitivity and specificity (refer to section 5.4). For example, one hospitalization with a diagnostic code of IBD means that 98% of the times the patient has IBD but this number for one claim in PC database is 70%. As applying similar weights to contacts in different administrative databases is unjustifiable, we developed our definition based on dissimilar weights on ACCS, IP and PC databases.
- There is no specific ICD code for indeterminate colitis. Diagnosis of UC or CD was assigned to all the IBD patients; therefore, a percentage of CD or UC patients reported in the Manitoba study are actually IC patients. In our study we only captured IBD patients and we did not distinguish between CD and UC. In future studies we will develop additional definitions to identify UC, CD and IC patients. This will also give us an advantage to track UC patients who become

[7] Youden's index calculated from sensitivity+specificity+1

CD, IC patients who become CD or UC, and determination of the course of disease in IC patients.

- The Manitoba study allowed a hiatus of 5 years with no health care contacts to distinguish between incident and prevalent cases. However, our study shows that 5 years is not an adequate amount of time for discrimination between incident and prevalent cases, the Manitoba study might have overestimated the number of incident cases. Random chart reviews on false negatives showed that 87.5% (95%CI: 71%-96.5%) of IBD patients who were not considered as IBD by our definitions are old cases of IBD (Table 9). In the other words, about 23.1% (95%CI: 18.8%-25.5%) of prevalent IBD patients might not fulfill our criteria in a period of 10 years. In the Calgary Health Region, it is possible to identify incident cases by searching EndoPro. Comparison of the health care records of these patients with controls (i.e. prevalent cases) will recognize a reasonable time hiatus to distinguish between incident and prevalent cases.
- The administrative criteria in the Manitoba study was developed on a fixed time period of 1984-1995, adding more years of data to this period without reassessment of the criteria characteristics is unjustifiable. There is no guarantee that health care contacts after 1995 follow the same trend as 1984-95. That is why that conventionally a fixed time period is fractionated into smaller fixed periods so that updating the database in the future could be justifiable (e.g. Hux et al. [52] divided the time period between 1991 through 1999 into 2-year time periods to develop an administrative definition for diabetes in Ontario). In addition, application of the criteria on different 10 year periods (e.g. 1994 to

2005) is again unjustifiable as the number and trend of physician claims and hospitalizations significantly change through time [53] (Figure 3, page 35). Another remarkable flaw in extrapolation of the Manitoba criteria to longer periods of time is the accumulation of the false positive cases. For example applying the criteria on the administrative databases in the province of Manitoba from 1984-2006 will at least double the number of false positives (we are assuming that the number of health contacts in the period of 1995 to 2006 is equal to 1984-1995, even though most probably it is higher and there would be more false positives). Another example is the application of the Manitoba criteria to the province of Alberta on a 17-year period (1985-2002) [1]. The Alberta population is about three times that of Manitoba; therefore, the number of the reported false positives are less than one fifth of the actual number of the FPs ($1.7\times3=5.1$). In our study we used a novel floating algorithm to identify IBD patients. In some instances hundreds of 3-year periods were searched within our databases that make the updating process justifiable. Moreover, in the future by defining a washout period for patients who barely fulfilled the definition without further repetition of IBD contacts; we will try to minimize the accumulation of false positives.

Another advantage of our floating algorithm is its potential ability to detect flares of the disease by entering decimal number of years (e.g.0.08 year$\approx$ 1 month). Moreover it is possible to use different administrative databases with different available time frames (e.g. ACCS 1996-2004 with PC 1995-2004).

6.5 Limitations

When extrapolating our administrative definition to the population of the Calgary Health Region, 85.7% sensitivity of our definition might be an overestimation due to the fact that our primary cohort of patients was drawn from a colonoscopy database, which could contain more severe cases of IBD. However, as shown in Figure 1, 2583 procedures were classified as IBD through indication search (meaning that they were already known cases of IBD). Of 2,583 colonoscopy procedure, 2186 (84.6%) were not classified as IBD in findings and impressions search signifying that they were probably mild or quiescent cases.

Linkage of administrative databases using non-unique identifiers such as date of birth, gender, and name does not have a perfect accuracy. This could be could be a limitation as we linked our EndoPro records to registry using date of birth, gender, first name and last name.

It is acknowledged that there is no perfectly sensitive and specific "standard diagnostic test" for IBD. The same rule applies to endoscopy reports and chart reviews, which were considered as our "reference standard" in our study.

Severity of a disease is a complex and fluctuating construct that is not easy, if not impossible, to extract from administrative databases; therefore, our definition is not capable of determination of disease severity in our patients.

Another potential source of bias in our study is the miscodings present in the administrative databases; however, the percentage have been shown to be relatively small in several studies [30, 35-38].

We did not review any charts from true negative cases, the reason was the feasibility of performing a reasonable number of chart reviews on 15,000 individuals. Although it is unlikely that an IBD patient has at least one normal colonoscopy report and also not enough health service records to fulfill our administrative criteria; we can not rule out that possibility. We believe that extending the time period in which we searched the administrative databases (in our study 10 years) will minimize this error.

In our study, some of the patients were classified as "possible IBD", because there was not enough evidence to rule-in or rule-out IBD in their endoscopy reports or charts. We did not include these patients in our definition development process, which can be debated as a flaw in the generalizability of our definition. However, it is not uncommon in clinical practice to have patients with pending diagnosis of IBD; in fact long delays in diagnosis have been reported since the onset of the clinical manifestations in IBD patients [54]. Obviously, there is no feasible way to classify these patients before the confirmation of the diagnosis.

6.6 Conclusion

Through a novel method, we developed an administrative definition to identify IBD patients using *physician claims, hospital abstracts, and ambulatory care system* databases. The definition considers a subject as an IBD patient if four physician claims or two ambulatory care contacts, or one hospitalization with diagnostic ICD codes of IBD exist within a period of 3 years. Sensitivity and specificity of our administrative definition are 85.70% and 99.75% respectively. We also propose the definition *"IP≥1 or PC≥4 within 3 years"* with a sensitivity of 80.13% and a specificity of 99.79% for health

regions that do not have access to the ACCS database. The results of this study can pave the way for future population-based studies on IBD.

Chapter Seven: REFERENCES

1. Bernstein, C.N., et al., *The epidemiology of inflammatory bowel disease in Canada: a population-based study.* Am J Gastroenterol, 2006. **101**(7): p. 1559-68.
2. Bernstein, C.N., et al., *Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study.* Am J Epidemiol, 1999. **149**(10): p. 916-24.
3. Yamada, T. and D.H. Alpers, *Textbook of gastroenterology*. 4th ed. 2003, Philadelphia: Lippincott Williams & Wilkins. 2 v. (xxv, 3309).
4. Farmer, R.G., W.A. Hawk, and R.B. Turnbull, Jr., *Clinical patterns in Crohn's disease: a statistical study of 615 cases.* Gastroenterology, 1975. **68**(4 Pt 1): p. 627-35.
5. Greenstein, A.J., H.D. Janowitz, and D.B. Sachar, *The extra-intestinal complications of Crohn's disease and ulcerative colitis: a study of 700 patients.* Medicine (Baltimore), 1976. **55**(5): p. 401-12.
6. Meyers, S. and H.D. Janowitz, *"Natural history" of Crohn's disease. An analytic review of the placebo lesson.* Gastroenterology, 1984. **87**(5): p. 1189-92.
7. Wyke, R.J., F.C. Edwards, and R.N. Allan, *Employment problems and prospects for patients with inflammatory bowel disease.* Gut, 1988. **29**(9): p. 1229-35.
8. Syngal, S., P. Bandipalliam, and C.R. Boland, *Surveillance of patients at high risk for colorectal cancer.* Med Clin North Am, 2005. **89**(1): p. 61-84, vii-viii.
9. Rubesin, S.E., et al., *Radiologic and endoscopic diagnosis of Crohn's disease.* Surg Clin North Am, 2001. **81**(1): p. 39-70, viii.
10. Quinn, P.G., D.G. Binion, and P.J. Connors, *The role of endoscopy in inflammatory bowel disease.* Med Clin North Am, 1994. **78**(6): p. 1331-52.
11. Loftus, E.V., Jr., *Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences.* Gastroenterology, 2004. **126**(6): p. 1504-17.
12. Pinchbeck, B.R., J. Kirdeikis, and A.B. Thomson, *Inflammatory bowel disease in northern Alberta. An epidemiologic study.* J Clin Gastroenterol, 1988. **10**(5): p. 505-15.
13. Loftus, E.V., Jr., et al., *Crohn's disease in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival.* Gastroenterology, 1998. **114**(6): p. 1161-8.
14. Munkholm, P., et al., *Incidence and prevalence of Crohn's disease in the county of Copenhagen, 1962-87: a sixfold increase in incidence.* Scand J Gastroenterol, 1992. **27**(7): p. 609-14.
15. Langholz, E., et al., *Incidence and prevalence of ulcerative colitis in Copenhagen county from 1962 to 1987.* Scand J Gastroenterol, 1991. **26**(12): p. 1247-56.
16. Rubin, G.P., et al., *Inflammatory bowel disease: epidemiology and management in an English general practice population.* Aliment Pharmacol Ther, 2000. **14**(12): p. 1553-9.

17. Trallori, G., et al., *A population-based study of inflammatory bowel disease in Florence over 15 years (1978-92).* Scand J Gastroenterol, 1996. **31**(9): p. 892-9.
18. Vucelic, B., et al., *Epidemiology of Crohn's disease in Zagreb, Yugoslavia: a ten-year prospective study.* Int J Epidemiol, 1991. **20**(1): p. 216-20.
19. Vucelic, B., et al., *Ulcerative colitis in Zagreb, Yugoslavia: incidence and prevalence 1980-1989.* Int J Epidemiol, 1991. **20**(4): p. 1043-7.
20. Morita, N., et al., *Incidence and prevalence of inflammatory bowel disease in Japan: nationwide epidemiological survey during the year 1991.* J Gastroenterol, 1995. **30 Suppl 8**: p. 1-4.
21. Loftus, E.V., Jr. and W.J. Sandborn, *Epidemiology of inflammatory bowel disease.* Gastroenterol Clin North Am, 2002. **31**(1): p. 1-20.
22. Roos, L.L., et al., *Registries and administrative data: organization and accuracy.* Med Care, 1993. **31**(3): p. 201-12.
23. Lawthers, A.G., et al., *Identification of in-hospital complications from claims data. Is it valid?* Med Care, 2000. **38**(8): p. 785-95.
24. Iezzoni, L.I., *Assessing quality using administrative data.* Ann Intern Med, 1997. **127**(8 Pt 2): p. 666-74.
25. Romano, P.S. and D.H. Mark, *Bias in the coding of hospital discharge data and its implications for quality assessment.* Med Care, 1994. **32**(1): p. 81-90.
26. Romano, P.S., et al., *A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Ischemic Heart Disease Patient Outcomes Research Team.* J Clin Epidemiol, 1994. **47**(3): p. 249-60.
27. *Abstracting manual. Ottawa, Ontarion, Canada: Canadian institute for health information; 1999.*
28. *Richards J, Brown A, Homan C. The data quality study of the Canadian discharge abstract database. Proceeding statistics Canada Symposium 2001.*
29. Austin, P.C., P.A. Daly, and J.V. Tu, *A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario.* Am Heart J, 2002. **144**(2): p. 290-6.
30. Mitchell S, B.A. *Is it fit for use? Data quality initiatives at CIHI.* [cited 2005 Nov]; Available from: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=services_dad_e#quality.
31. *Goel V, Williams JI, Anderson GM, Blackstein-Hirch P, Fooks C, Naylor CD. Patterns of Health care in Ontario. The ICES Practice Atlas 2nd ed., Ottawa: Canadian Medical Association; 54-64; 1996.*
32. Hawker, G.A., et al., *Accuracy of administrative data for assessing outcomes after knee replacement surgery.* J Clin Epidemiol, 1997. **50**(3): p. 265-73.
33. Rawson, N.S., E. Malcolm, and C. D'Arcy, *Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles.* Soc Psychiatry Psychiatr Epidemiol, 1997. **32**(4): p. 191-9.
34. Muhajarine, N., et al., *Comparison of survey and physician claims data for detecting hypertension.* J Clin Epidemiol, 1997. **50**(6): p. 711-8.
35. Robinson, J.R., et al., *Estimating the burden of disease. Comparing administrative data and self-reports.* Med Care, 1997. **35**(9): p. 932-47.

36. Young, T.K., N.P. Roos, and K.M. Hammerstrand, *Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: a pilot study.* Cmaj, 1991. **144**(3): p. 318-24.
37. *Alberta Ambulatory care reporting manual 2002 (Accessed on Nov 2005)* http://www.health.gov.ab.ca/public/document/ACCS_manual_02/index.htm.
38. *(Accessed on Nov 2005)* http://www.who.int/classifications/icd/en/http://www.who.int/classifications/icd
39. http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=codingclass_icd10_e. [cited 2007 Jan 10].
40. *Robert F. DeVellis. Scale Development: Theory and Applications.2nd Edition. Sage publication Inc. P-51.*
41. Bolboaca, S. and L. Jantschi, *Binomial distribution sample confidence intervals estimation for positive and negative likelihood ratio medical key parameters.* AMIA Annu Symp Proc, 2005: p. 66-70.
42. Lennard-Jones, J.E., *Classification of inflammatory bowel disease.* Scand J Gastroenterol Suppl, 1989. **170**: p. 2-6; discussion 16-9.
43. *MRC, NSERC, SSHRC. Tri-council policy statement: the ethical conduct for research involving humans.* http://www.pre.ethics.gc.ca/english/policystatement/policystatement.cfm *(Accessed in March 2006).*
44. Delfino, R.J., M.R. Becklake, and J.A. Hanley, *Reliability of hospital data for population-based studies of air pollution.* Arch Environ Health, 1993. **48**(3): p. 140-6.
45. Calkins, B.M. and A.I. Mendeloff, *Epidemiology of inflammatory bowel disease.* Epidemiol Rev, 1986. **8**: p. 60-91.
46. Russel, M.G. and R.W. Stockbrugger, *Epidemiology of inflammatory bowel disease: an update.* Scand J Gastroenterol, 1996. **31**(5): p. 417-27.
47. *Colorectal cancer screening. Recommendation statement from the Canadian Task Force on Preventive Health Care.* Cmaj, 2001. **165**(2): p. 206-8.
48. http://www40.statcan.ca/l01/cst01/demo10a.htm *(Accessed on Jan 11, 2007).*
49. *(Accessed on Jan 8, 2007)* http://www.calgaryhealthregion.ca/communications/regionhealth/health_region_report_06.pdf#start.
50. Farrokhyar, F., K. McHugh, and E.J. Irvine, *Self-reported awareness and use of the International Classification of Diseases coding of inflammatory bowel disease services by Ontario physicians.* Can J Gastroenterol, 2002. **16**(8): p. 519-26.
51. *Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3: 32-35.*
52. Hux, J.E., et al., *Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm.* Diabetes Care, 2002. **25**(3): p. 512-6.
53. Nabalamba, A., C.N. Bernstein, and C. Seko, *Inflammatory bowel disease--hospitalization.* Health Rep, 2004. **15**(4): p. 25-40.
54. Pimentel, M., et al., *Identification of a prodromal period in Crohn's disease but not ulcerative colitis.* Am J Gastroenterol, 2000. **95**(12): p. 3458-62.

APPENDIX 1. INTERNATIONAL CLASSIFICATION OF DISEASES, NINTH AND TENTH REVISION

- ICD-9-CM

NONINFECTIOUS ENTERITIS AND COLITIS (555-558)

555 Regional enteritis

Includes: Crohn's disease

Granulomatous enteritis

Excludes: Ulcerative colitis (556)

555.0 Small intestine

Ileitis:

regional

segmental

terminal

Regional enteritis or Crohn's disease of:

duodenum

ileum

jejunum

555.1 Large intestine

Colitis:

granulomatous

regional

transmural

Regional enteritis or Crohn's disease of:

colon

large bowel

rectum

555.2 Small intestine with large intestine

Regional ileocolitis

555.9 Unspecified site

Crohn's disease NOS

Regional enteritis NOS

556 Ulcerative colitis

556.0 Ulcerative (chronic) enterocolitis

556.1 Ulcerative (chronic) ileocolitis

556.2 Ulcerative (chronic) proctitis

556.3 Ulcerative (chronic) proctosigmoiditis

556.4 Pseudopolyposis of colon

556.5 Left-sided ulcerative (chronic) colitis

556.6 Universal ulcerative (chronic) colitis

Pancolitis

556.8 Other ulcerative colitis

556.9 Ulcerative colitis, unspecified

Ulcerative enteritis NOS

557 Vascular insufficiency of intestine

Excludes: necrotizing enterocolitis of the newborn (777.5)

557.0 Acute vascular insufficiency of intestine

Acute:

hemorrhagic enterocolitis

ischemic colitis, enteritis, or enterocolitis

massive necrosis of intestine

Bowel infarction

Embolism of mesenteric artery

Fulminant enterocolitis

Hemorrhagic necrosis of intestine

Infarction of appendices epiploicae

Intestinal gangrene

Intestinal infarction (acute) (agnogenic) (hemorrhagic) (nonocclusive)

Mesenteric infarction (embolic) (thrombotic)

Necrosis of intestine

Terminal hemorrhagic enteropathy

Thrombosis of mesenteric artery

557.1 Chronic vascular insufficiency of intestine

Angina, abdominal

Chronic ischemic colitis, enteritis, or enterocolitis

Ischemic stricture of intestine

Mesenteric:

angina

artery syndrome (superior)

vascular insufficiency

557.9 Unspecified vascular insufficiency of intestine

Alimentary pain due to vascular insufficiency

Ischemic colitis, enteritis, or enterocolitis NOS

558 Other and unspecified noninfectious gastroenteritis and colitis

Excludes: Infectious:

colitis, enteritis, or gastroenteritis (009.0-009.1)

diarrhea (009.2-009.3)

558.1 Gastroenteritis and colitis due to radiation

Radiation enterocolitis

558.2 Toxic gastroenteritis and colitis

558.3 Allergic gastroenteritis and colitis

558.9 Other and unspecified noninfectious gastroenteritis and colitis

Colitis, NOS, dietetic, or noninfectious

Enteritis, NOS, dietetic, or noninfectious

Gastroenteritis, NOS, dietetic, or noninfectious

Ileitis, NOS, dietetic or noninfectious

Jejunitis, NOS, dietetic, or noninfectious

Sigmoiditis, NOS, dietetic, or noninfectious

- ICD-10-CA

NONINFECTIOUS ENTERITIS AND COLITIS (555-558)

Includes: Inflammatory bowel disease

Excludes: Irritable bowel syndrome (K58), megacolon (K59.3)

K50 Crohn's disease

K50.0 Crohn's disease of small intestine

K50.1 Crohn's disease of large intestine

K50.8 Other Crohn's disease

Includes: Crohn's disease of both small and large intestine

K50.9 Crohn's disease, unspecified

K51 Ulcerative colitis

K51.0 Ulcerative (chronic) enterocolitis

K51.1 Ulcerative (chronic) ileocolitis

K51.2 Ulcerative (chronic) proctitis

K51.3 Ulcerative (chronic) rectosigmoiditis

K51.4 Pseudopolyposis of colon

K51.5 Mucosal proctocolitis

K51.8 Other Ulcerative colitis

K51.9 Ulcerative colitis, unspecified

K52 Other non-infective gastroenteritis

K52.0 Gastroenteritis and colitis due to radiation

K52.1 Toxic gastroenteritis and colitis

K52.2 Allergic and diatetic gastroenteritis and colitis

K52.8 Other specified non-infective gastroenteritis and colitis

Includes: Eosinophilic gastritis and gastroenteritis

K52.9 Non-infective gastroenteritis and colitis, unspecified

APPENDIX 2. SEARCH TERMS IN ENDOPRO DATABASE

SEARCH TERMS USED TO SCREEN ENDOPRO DATABSE INDICATION FIELD:

- **Definite IBD:** “crohn”, “ulcerative colitis”, “ulcerative proctitis”, “uc”, “cd”, “activity”, “ibd” , NOT “family”
- **Possible IBD:** “colitis”, “proctitis” AND not the terms for definite IBD

SEARCH TERMS USED TO SCREEN ENDOPRO DATABSE FINDINGS FIELD:

- **Possible IBD:** "ulcer", "proctitis", "loss of vascular ", "inflammation", "granular", "fistula", "diffuse ulceration", "colitis", "blunting of vascular", "apthous ulcers", "patchy, erythema"
- **Definite IBD:** “crohn”, “ulcerative colitis”, “ulcerative proctitis”, “uc”, “cd”, “ibd” , “ileitis”

SEARCH TERMS USED TO SCREEN ENDOPRO DATABSE IMPRESSION FIELD:

- **Possible IBD:** "ulcer", "pseudopolyp", "proctitis", "loss of vascular", "inflammation", "ileitis", "granular", "fistula", "diffuse ulceration", "colitis", "blunting", "apthous", “patchy erythema”
- **Definite IBD: “crohn”, “ulcerative colitis”, “ulcerative proctitis”, “uc”, “cd”, “ibd”**

APPENDIX 3. ABSTRACTION FORM FOR INPATIENT AND OUTPATIENT CHARTS.

IBD CRITERIA STUDY

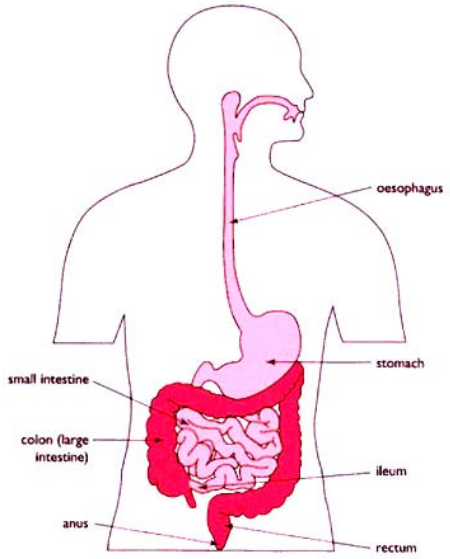
CHART ABSTRATCTION FORM

All the information contained in this form are strictly confidential.

Name (Last, First, M.I.):	☐ M ☐ F	**DOB:**
Chart No:	**Abstraction date:**	**Record No:**
Hospital name:	**Date of hospitalization:**	

SPECIFIC FINDINGS IN HISTORY AND PHYSICAL EXAMINATION				
	☐ Complex anal lesion		☐ Skin (EN or PG)	☐ Diarrhea/constipation
	☐ Fistula		☐ Abdominal mass in RLQ	☐ Blood in stool
	☐ Buccal apthous		☐ Abdominal pain	☐ Hx of bowel resection

RADIOLOGIC FINDINGS		
	☐ Suggestive of IBD	☐ Abscess
	☐ Stenosis	☐ Fistula
	☐ Mass in terminal ileum	☐ Thickening of the intestinal wall

ENDOSCOPIC FINDINGS			
	INVOLVEMENT Shade the corresponding area(s)		
	☐ Ulceration	☐ Continuous	☐ Skip lesions
	☐ Erythema	☐ Patchy	☐ Diffuse
	☐ Stenosis		
	☐ Apthous ulcers		

PATHOLOGIC FINDINGS		
	☐ Chronic inflammation	☐ Crypt abscess
	☐ Inflammation in mucosa	☐ Transmural inflammation
	☐ Granuloma	☐ Non-specific inflammation

HAVE THE FOLLOWING DISEASES BEEN EXCLUDED?					
	Radiation Colitis (Hx)	☐	Yes	☐	No
	Infectious colitis	☐	Yes	☐	No
	Microscopic colitis (Pathology and Hx)	☐	Yes	☐	No
	Lymphoma/carcinoma (Hx and suggestive radiologic findings)	☐	Yes	☐	No
	Ischemic colitis (Hx, predisposing factors, pathology, distribution of disease)	☐	Yes	☐	No

IBD Related Medication:

Past medical History:

FINAL DIAGNOSIS			
IBD DIAGNOSIS WAS MENTIONED IN: ☐ PHYSICIAN NOTE ☐ ENDOSCOPY REPORT ☐ PATHOLOGY REPORT ☐ RADIOLOGY REPORT			
☐ Inflammatory Bowel disease	☐ Crohn's disease	☐ Ulcerative colitis	☐ Indeterminate colitis
☐ Possible IBD			
☐ Definitely not IBD			

APPENDIX 4. ETHICAL APPROVAL

AUG 14,2006 08:32 403 283 8524 Page
TOTAL P.01

FACULTY OF MEDICINE | UNIVERSITY OF CALGARY

OFFICE OF MEDICAL BIOETHICS
Room 93, Heritage Medical Research Bldg
3330 Hospital Drive NW
Calgary, AB, Canada T2N 4N1
Telephone: (403) 220-7990
Fax: (403) 283-8524
Email: omb@ucalgary.ca

2006-08-10

Dr. Robert Hilsden
Department of Medicine
Room 1703 - HSC
University of Calgary
Calgary, Alberta

Dear Dr. R. Hilsden:

RE: Development and Validation of an ICD Coding Definition for Inflammatory Bowel Disease

Ethics ID: E-20322

Student: Ali Rezaie

The above-noted proposal including the Research Proposal has been submitted for Board review and found to be ethically acceptable.

Please note that this approval is subject to the following conditions:
(1) consent for access to personal identified health information in retrospective chart review is not required on grounds considered under Section 50 of the Health Information Act;
(2) a copy of the informed consent form must have been given to each research subject, if required for this study;
(3) a Progress Report must be submitted by August 10, 2007, containing the following information:
- i) the number of subjects recruited;
- ii) a description of any protocol modification;
- iii) any unusual and/or severe complications, adverse events or unanticipated problems involving risks to subjects or others, withdrawal of subjects from the research, or complaints about the research;
- iv) a summary of any recent literature, finding, or other relevant information, especially information about risks associated with the research;
- v) a copy of the current informed consent form;
- vi) the expected date of termination of this project.

4) a Final Report must be submitted at the termination of the project.

Please note that you have been named as the principal collaborator on this study because students are not permitted to serve as principal investigators. Please accept the Board's best wishes for success in your research.
Yours sincerely,

M.C

Michael King, PhD ABPP (CI, CN)
Acting Chair, Conjoint Health Research Ethics Board

MK/sg
c.c. Adult Research Committee Dr. J. Conly (information) Research Services Ali Rezaie (Student)
Office of Information & Privacy Commissioner

AUG-14-2006 09:09 *** TRANSMISSION ERROR REPORT *** U OF C FAC MED BIOETHICS 403 283 8524 P.01/01

AUG 10,2006 00:29 Model # Lexmark X6100 Series 00140324187 42

START TIME	SENT TO	PAGES	RESULT
00:27	01133251898918	1	REMOTE FAX ERROR

--> Retry job or try different remote fax machine.

Foothills Medical Centre
1403 29 Street NW
Calgary, Alberta, Canada T2N 2T9
website www.calgaryhealthregion.ca

calgary health region
Foothills Medical Centre

August 11, 2006

Dr. Robert Hilsden
Department of Medicine
Room 1703 - HSC
University of Calgary
Calgary, Alberta

Dear Dr. Hilsden:

RE: E-20322 - Development and Validation of an ICD Coding Definition for Inflammatory Bowel Disease

Thank you for submitting an application regarding the above project for review by the Adult Research Committee of the Calgary Health Region (CHR). This will confirm that the committee has granted institutional approval for this project, and that the CHR has granted approval under Sections 53 and 54 of the Health Information Act. ***This approval is contingent on approval by the Conjoint Health Research Ethics Board.***

It is understood from your submission that your study will be entirely funded through external sources and that the CHR will be reimbursed for all research costs associated with this project. **To facilitate a smooth startup of your project, please notify affected departments in the Region well in advance of your intent to initiate this study.**

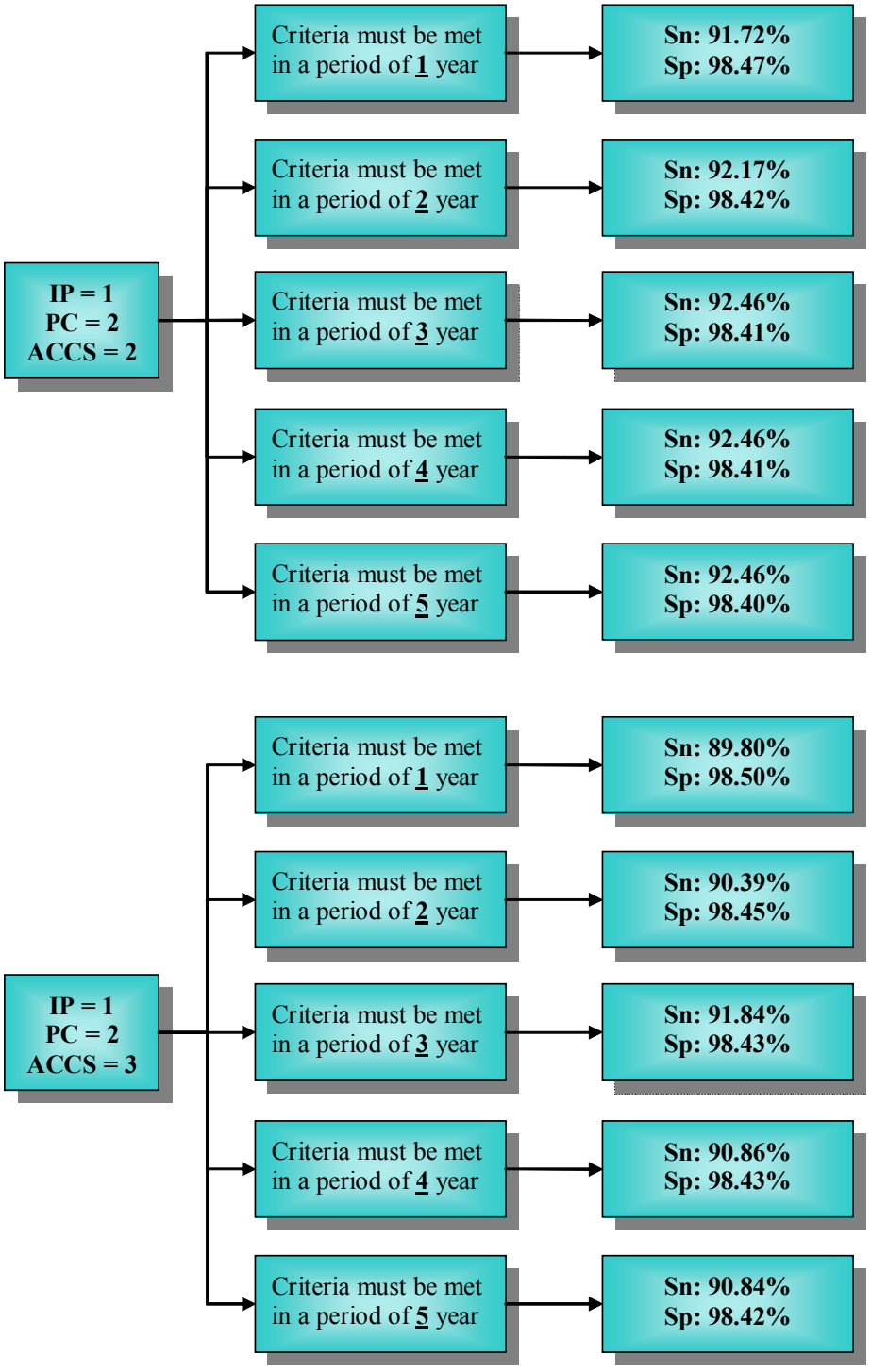
Please accept the committee's best wishes for success in your research.

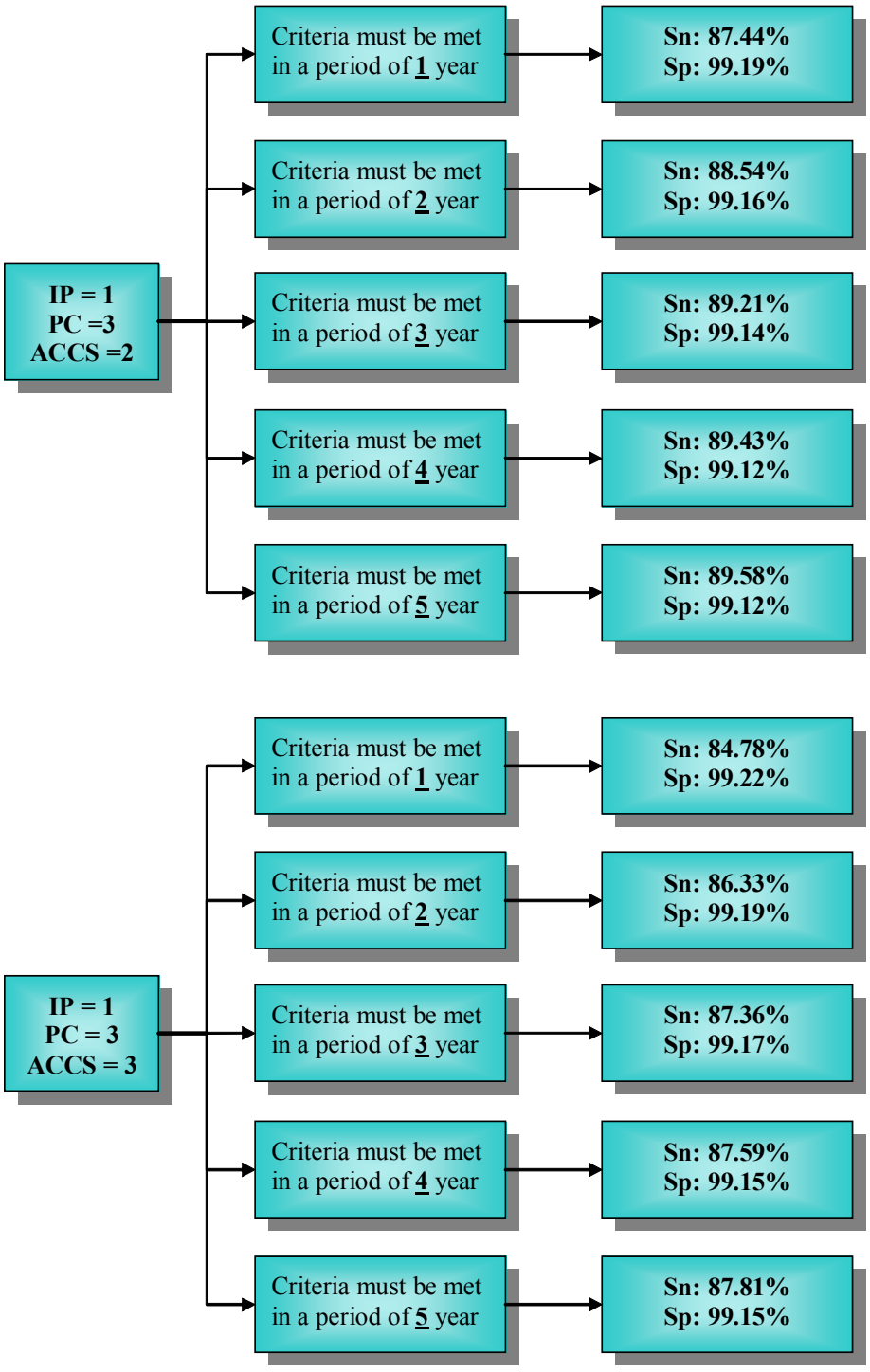
Yours sincerely,

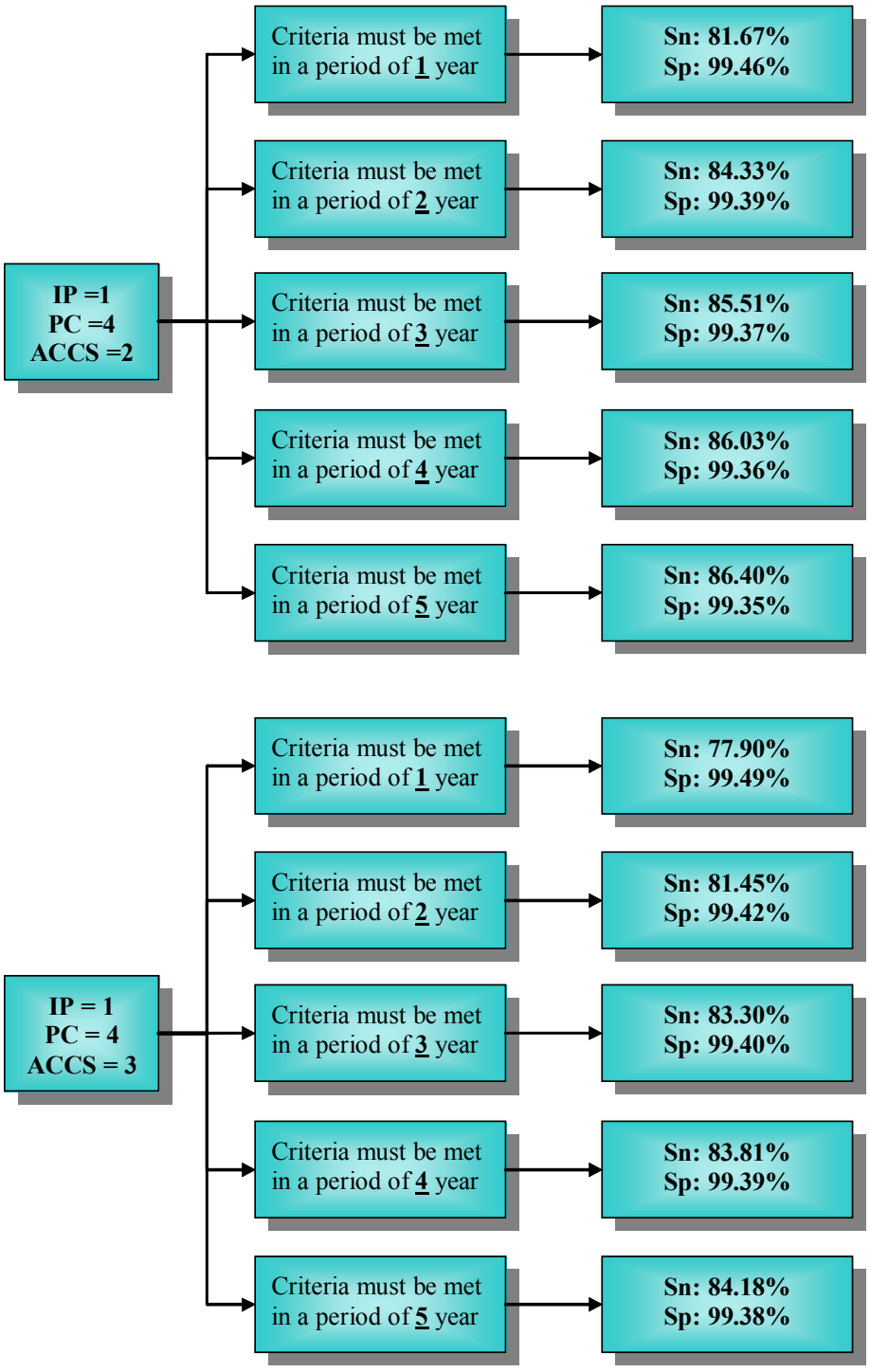
Elizabeth MacKay, MD
Acting Chair, Adult Research Committee

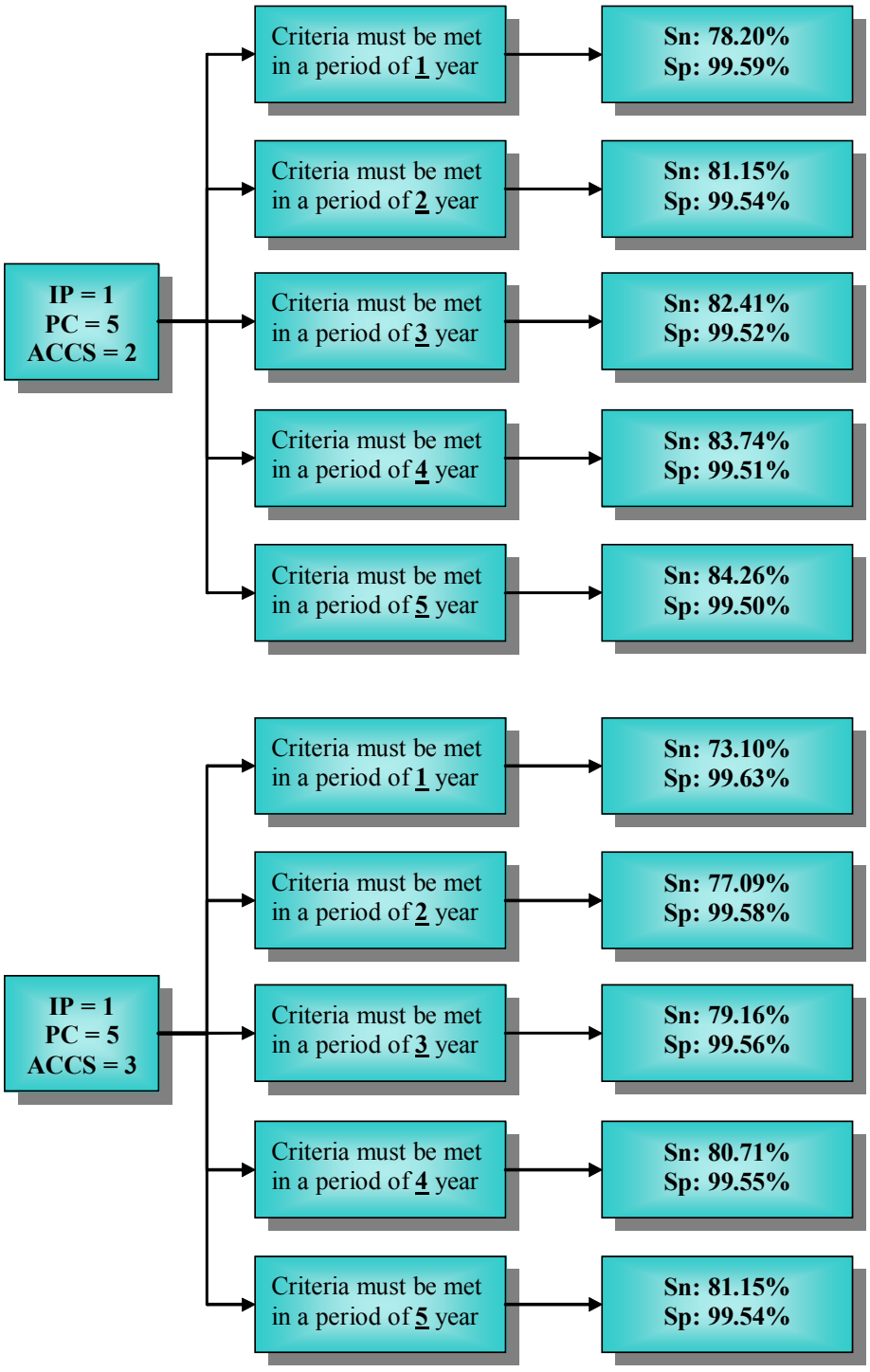
cc: Dr. J. Conly (information), Conjoint Health Research Ethics Board, Ali Rezaie (Student), Health Records

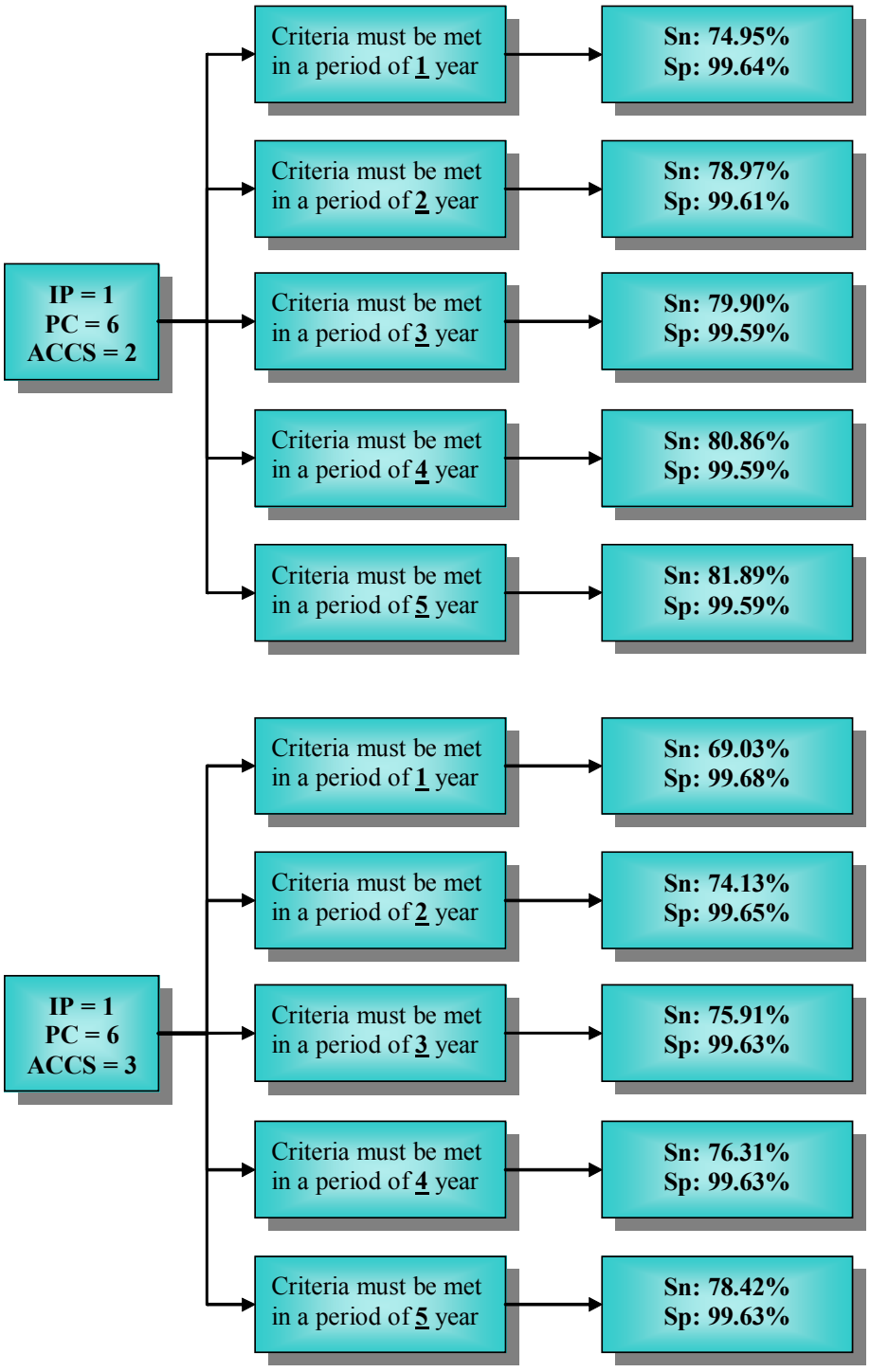
APPENDIX 5. SENSITIVITIES AND SPECIFICITIES FOR 100 PRELIMINARY DEFINITIONS

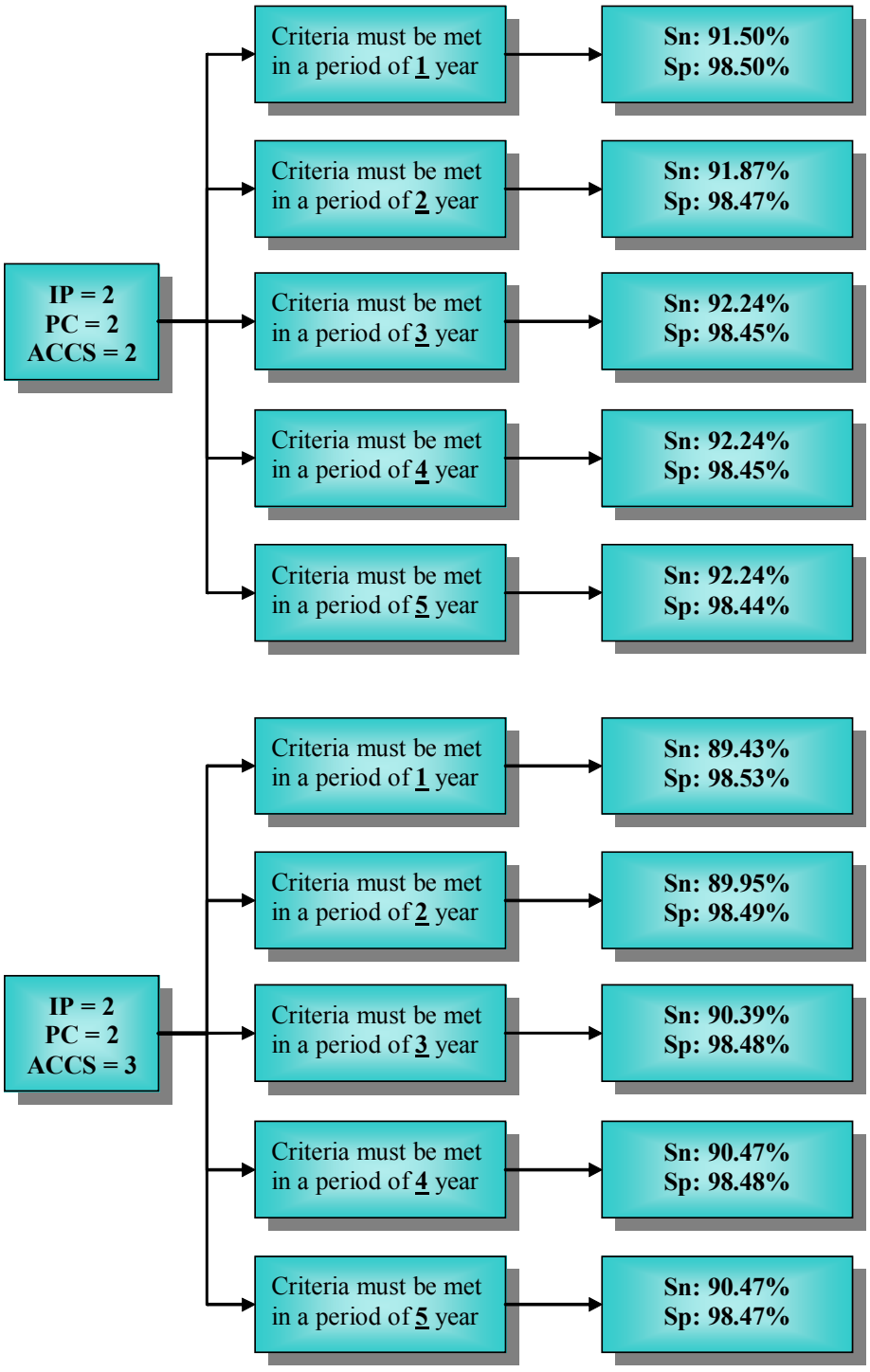


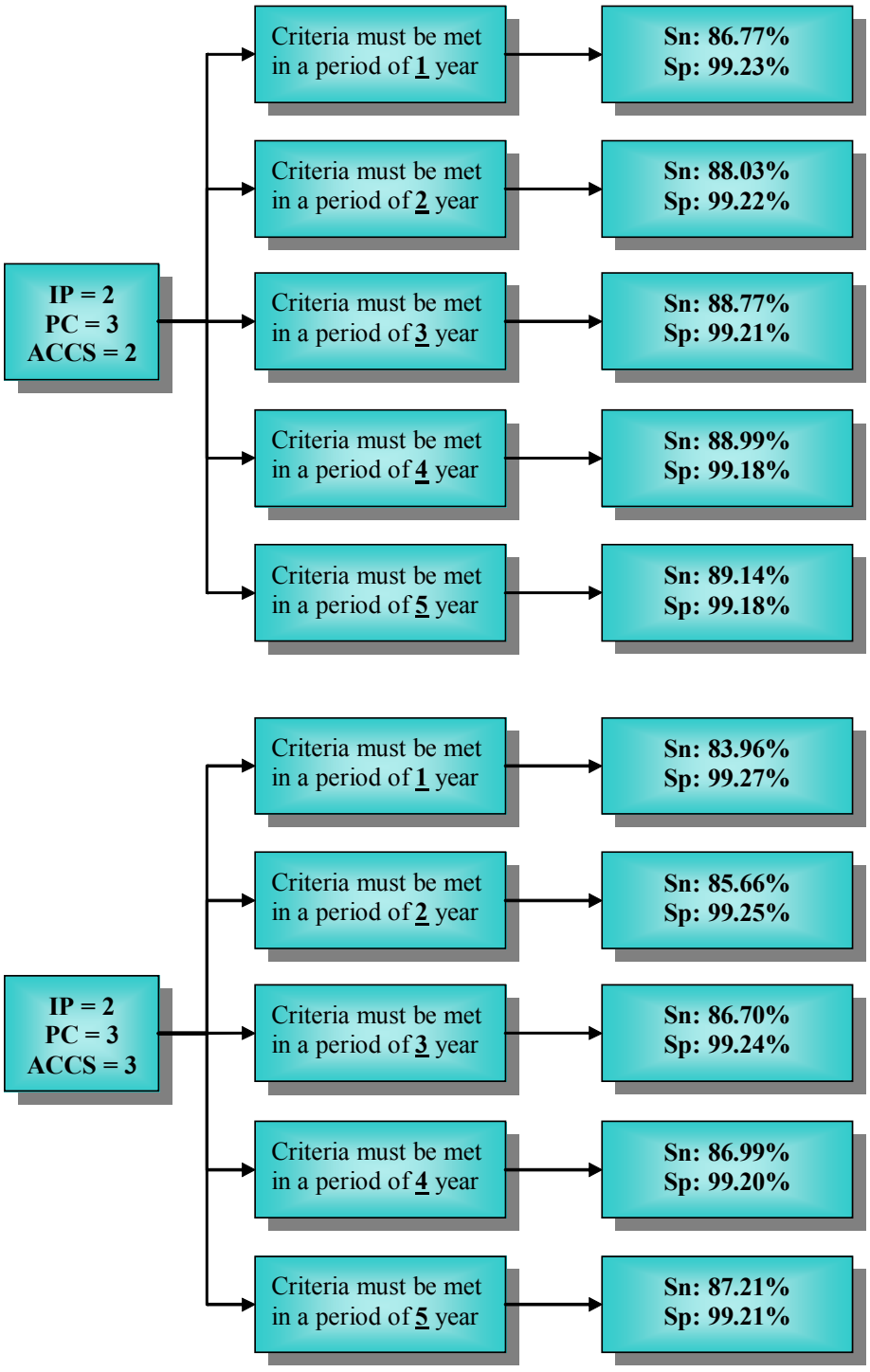

IP = 1
PC =3
ACCS =2
Criteria must be met in a period of 1 year
Sn: 87.44%
Sp: 99.19%
Criteria must be met in a period of 2 year
Sn: 88.54%
Sp: 99.16%
Criteria must be met in a period of 3 year
Sn: 89.21%
Sp: 99.14%
Criteria must be met in a period of 4 year
Sn: 89.43%
Sp: 99.12%
Criteria must be met in a period of 5 year
Sn: 89.58%
Sp: 99.12%
IP = 1
PC = 3
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 84.78%
Sp: 99.22%
Criteria must be met in a period of 2 year
Sn: 86.33%
Sp: 99.19%
Criteria must be met in a period of 3 year
Sn: 87.36%
Sp: 99.17%
Criteria must be met in a period of 4 year
Sn: 87.59%
Sp: 99.15%
Criteria must be met in a period of 5 year
Sn: 87.81%
Sp: 99.15%

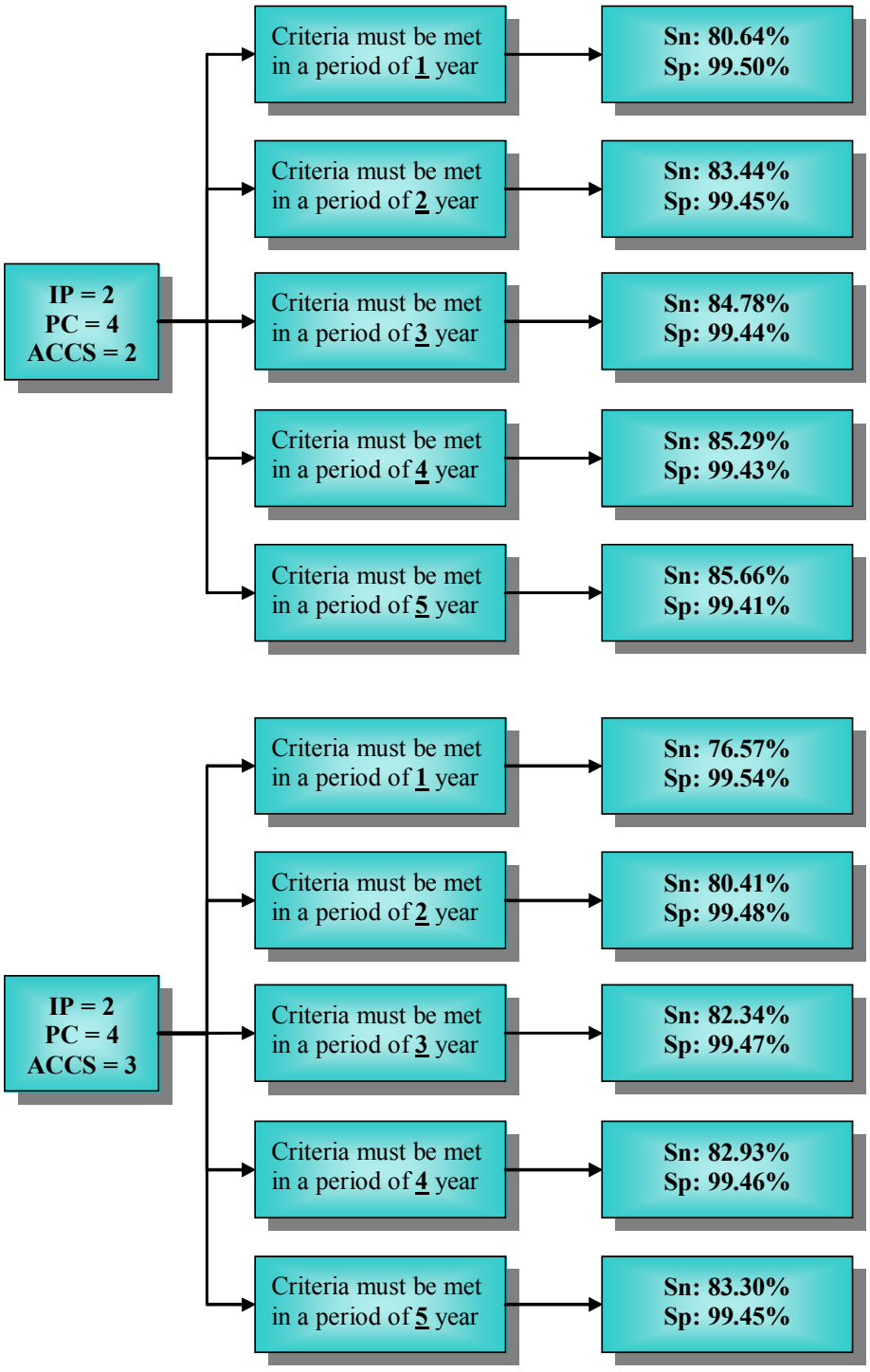

IP =1
PC =4
ACCS =2
Criteria must be met in a period of 1 year
Sn: 81.67%
Sp: 99.46%
Criteria must be met in a period of 2 year
Sn: 84.33%
Sp: 99.39%
Criteria must be met in a period of 3 year
Sn: 85.51%
Sp: 99.37%
Criteria must be met in a period of 4 year
Sn: 86.03%
Sp: 99.36%
Criteria must be met in a period of 5 year
Sn: 86.40%
Sp: 99.35%
IP = 1
PC = 4
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 77.90%
Sp: 99.49%
Criteria must be met in a period of 2 year
Sn: 81.45%
Sp: 99.42%
Criteria must be met in a period of 3 year
Sn: 83.30%
Sp: 99.40%
Criteria must be met in a period of 4 year
Sn: 83.81%
Sp: 99.39%
Criteria must be met in a period of 5 year
Sn: 84.18%
Sp: 99.38%

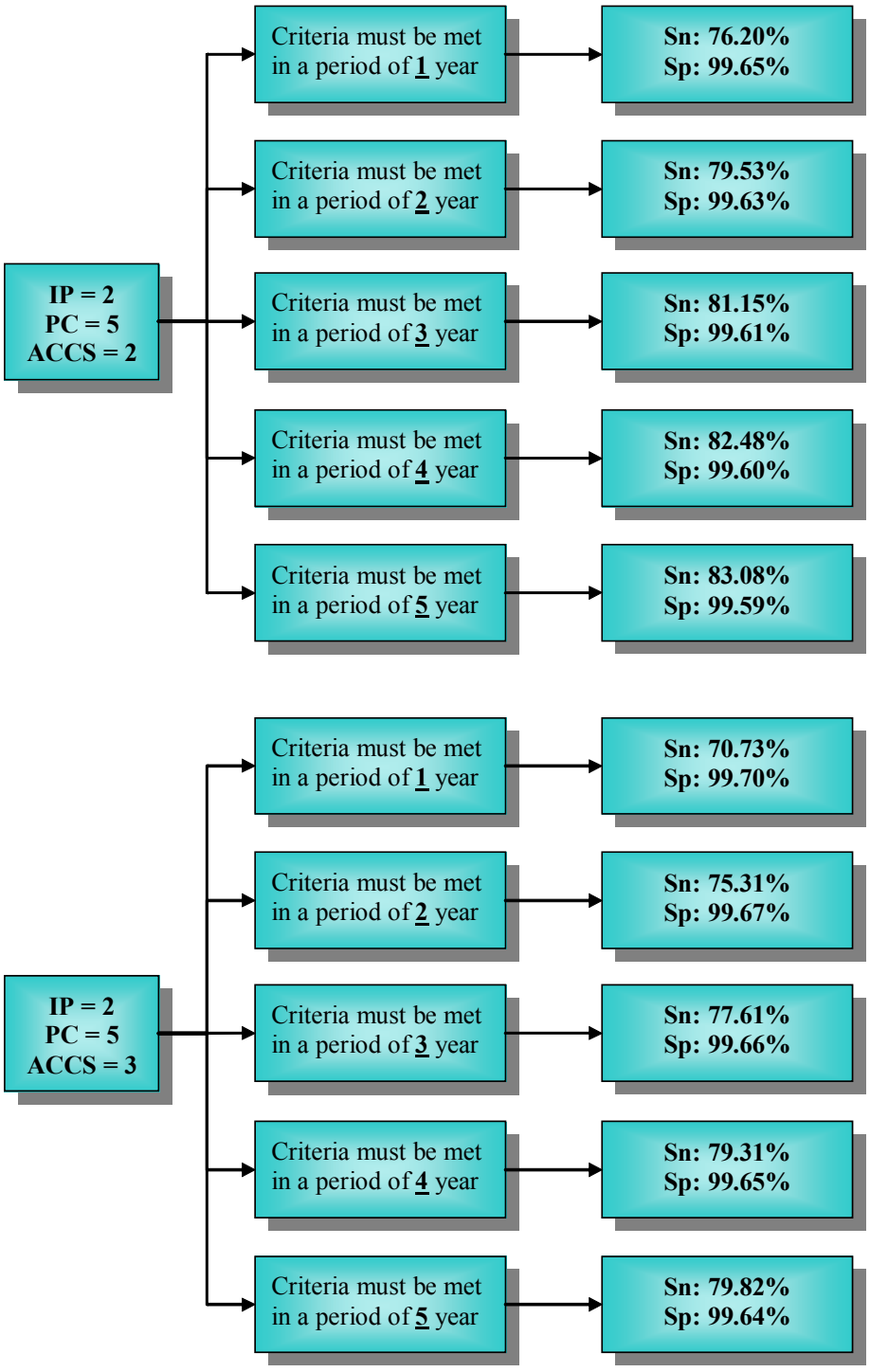

IP = 1
PC = 5
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 78.20%
Sp: 99.59%
Criteria must be met in a period of 2 year
Sn: 81.15%
Sp: 99.54%
Criteria must be met in a period of 3 year
Sn: 82.41%
Sp: 99.52%
Criteria must be met in a period of 4 year
Sn: 83.74%
Sp: 99.51%
Criteria must be met in a period of 5 year
Sn: 84.26%
Sp: 99.50%
IP = 1
PC = 5
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 73.10%
Sp: 99.63%
Criteria must be met in a period of 2 year
Sn: 77.09%
Sp: 99.58%
Criteria must be met in a period of 3 year
Sn: 79.16%
Sp: 99.56%
Criteria must be met in a period of 4 year
Sn: 80.71%
Sp: 99.55%
Criteria must be met in a period of 5 year
Sn: 81.15%
Sp: 99.54%

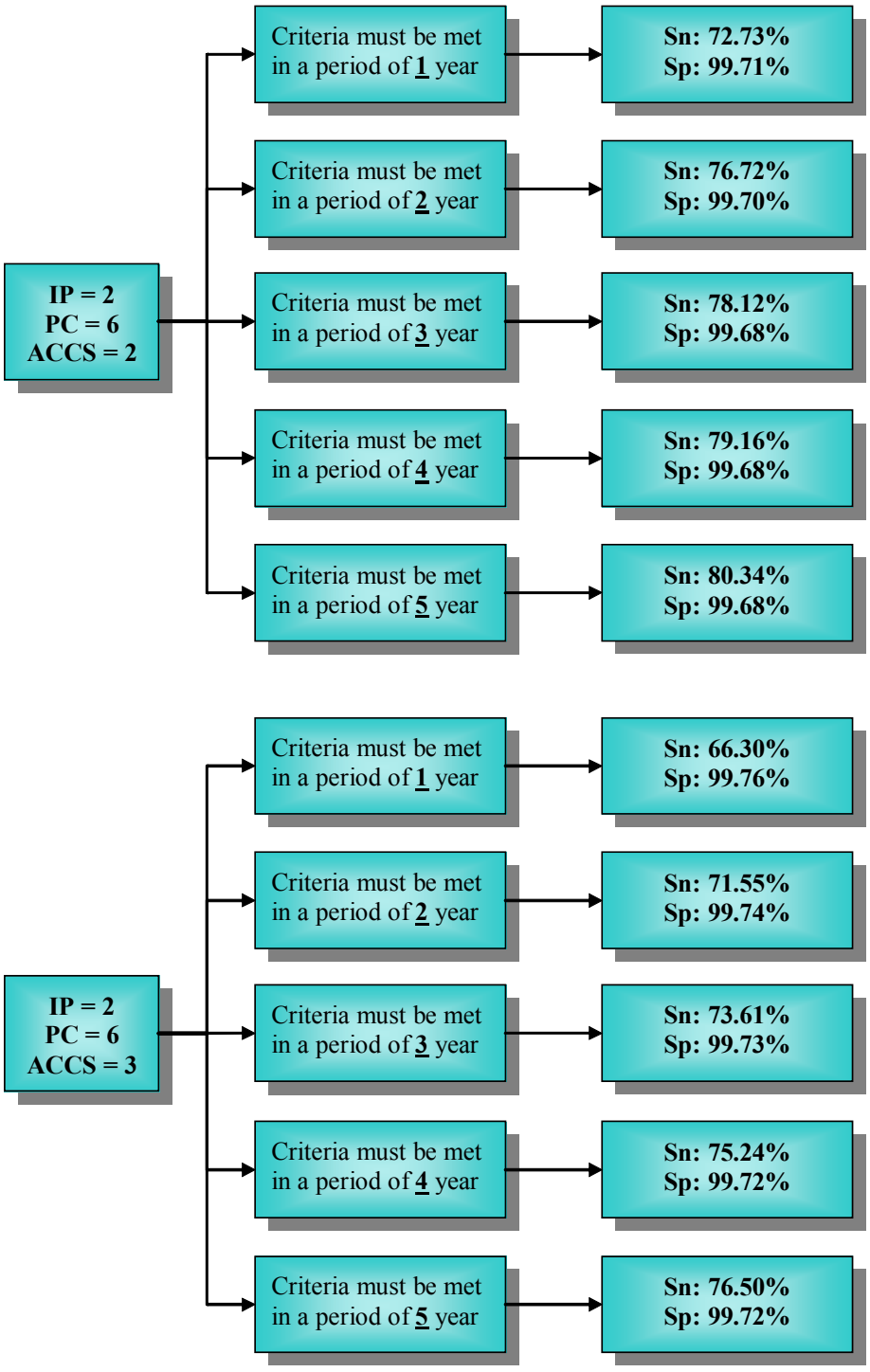

IP = 1
PC = 6
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 74.95%
Sp: 99.64%
Criteria must be met in a period of 2 year
Sn: 78.97%
Sp: 99.61%
Criteria must be met in a period of 3 year
Sn: 79.90%
Sp: 99.59%
Criteria must be met in a period of 4 year
Sn: 80.86%
Sp: 99.59%
Criteria must be met in a period of 5 year
Sn: 81.89%
Sp: 99.59%
IP = 1
PC = 6
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 69.03%
Sp: 99.68%
Criteria must be met in a period of 2 year
Sn: 74.13%
Sp: 99.65%
Criteria must be met in a period of 3 year
Sn: 75.91%
Sp: 99.63%
Criteria must be met in a period of 4 year
Sn: 76.31%
Sp: 99.63%
Criteria must be met in a period of 5 year
Sn: 78.42%
Sp: 99.63%


IP = 2
PC = 2
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 91.50%
Sp: 98.50%
Criteria must be met in a period of 2 year
Sn: 91.87%
Sp: 98.47%
Criteria must be met in a period of 3 year
Sn: 92.24%
Sp: 98.45%
Criteria must be met in a period of 4 year
Sn: 92.24%
Sp: 98.45%
Criteria must be met in a period of 5 year
Sn: 92.24%
Sp: 98.44%
IP = 2
PC = 2
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 89.43%
Sp: 98.53%
Criteria must be met in a period of 2 year
Sn: 89.95%
Sp: 98.49%
Criteria must be met in a period of 3 year
Sn: 90.39%
Sp: 98.48%
Criteria must be met in a period of 4 year
Sn: 90.47%
Sp: 98.48%
Criteria must be met in a period of 5 year
Sn: 90.47%
Sp: 98.47%

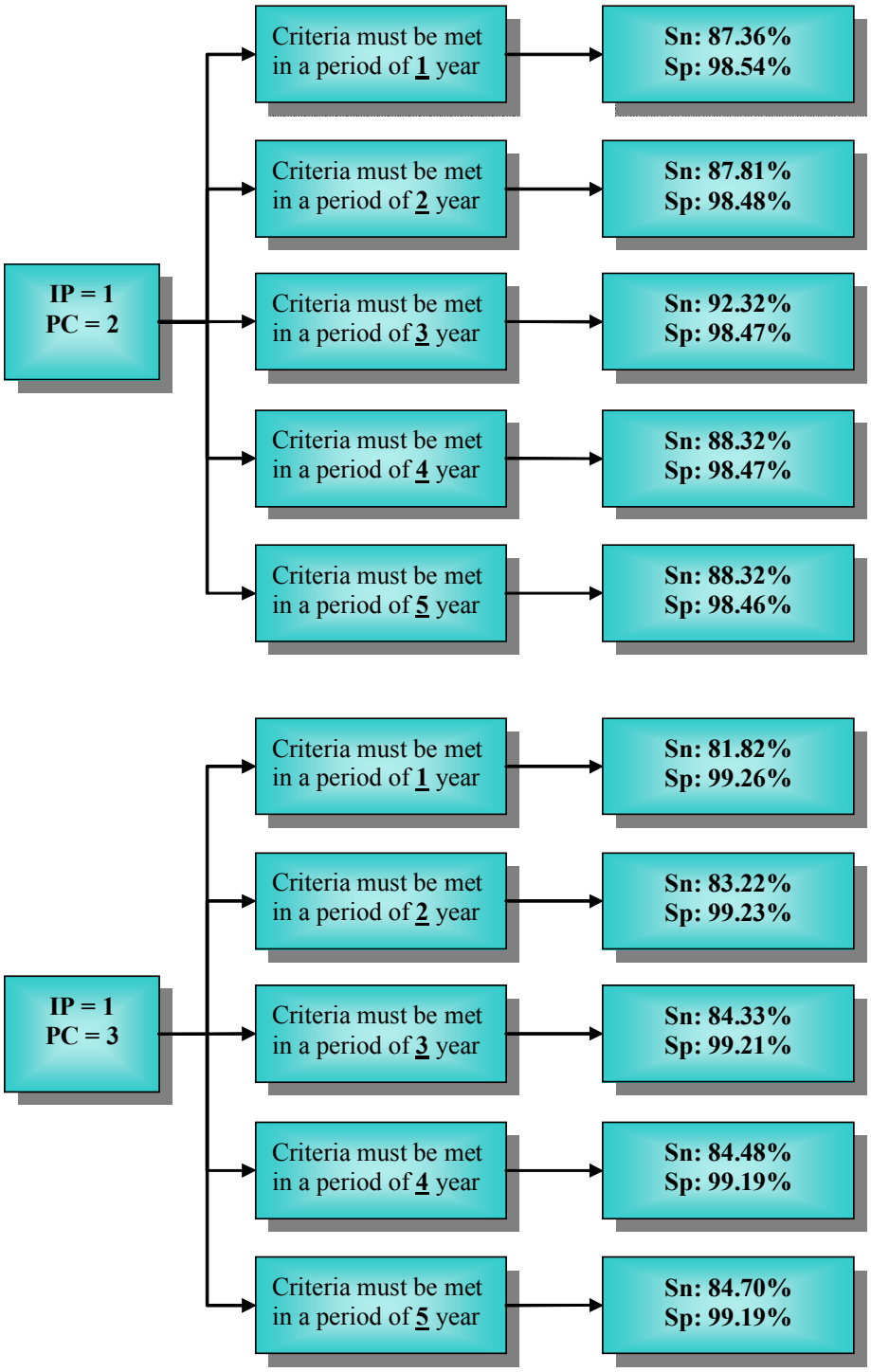

IP = 2
PC = 3
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 86.77%
Sp: 99.23%
Criteria must be met in a period of 2 year
Sn: 88.03%
Sp: 99.22%
Criteria must be met in a period of 3 year
Sn: 88.77%
Sp: 99.21%
Criteria must be met in a period of 4 year
Sn: 88.99%
Sp: 99.18%
Criteria must be met in a period of 5 year
Sn: 89.14%
Sp: 99.18%
IP = 2
PC = 3
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 83.96%
Sp: 99.27%
Criteria must be met in a period of 2 year
Sn: 85.66%
Sp: 99.25%
Criteria must be met in a period of 3 year
Sn: 86.70%
Sp: 99.24%
Criteria must be met in a period of 4 year
Sn: 86.99%
Sp: 99.20%
Criteria must be met in a period of 5 year
Sn: 87.21%
Sp: 99.21%

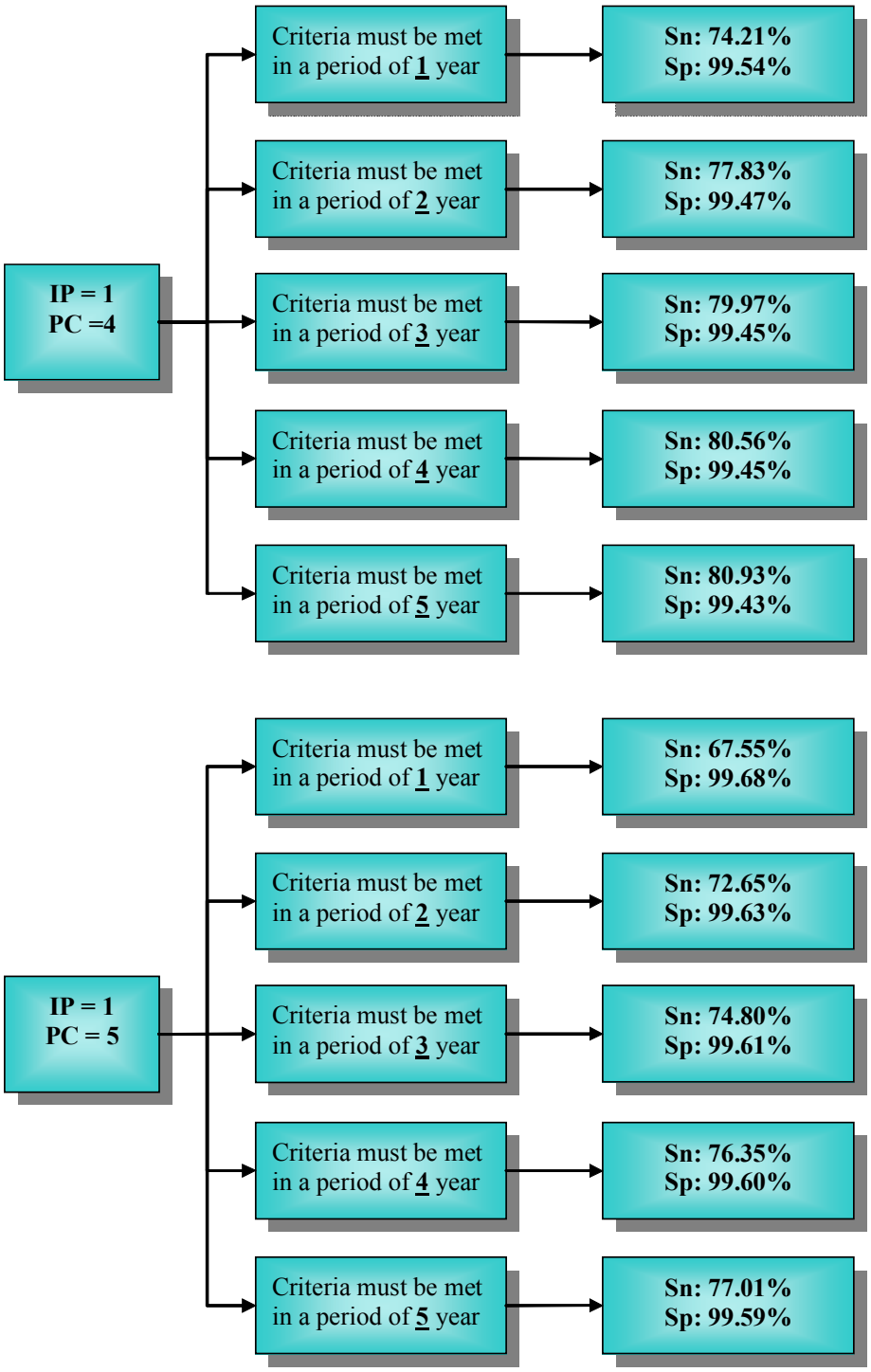

IP = 2
PC = 4
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 80.64%
Sp: 99.50%
Criteria must be met in a period of 2 year
Sn: 83.44%
Sp: 99.45%
Criteria must be met in a period of 3 year
Sn: 84.78%
Sp: 99.44%
Criteria must be met in a period of 4 year
Sn: 85.29%
Sp: 99.43%
Criteria must be met in a period of 5 year
Sn: 85.66%
Sp: 99.41%
IP = 2
PC = 4
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 76.57%
Sp: 99.54%
Criteria must be met in a period of 2 year
Sn: 80.41%
Sp: 99.48%
Criteria must be met in a period of 3 year
Sn: 82.34%
Sp: 99.47%
Criteria must be met in a period of 4 year
Sn: 82.93%
Sp: 99.46%
Criteria must be met in a period of 5 year
Sn: 83.30%
Sp: 99.45%

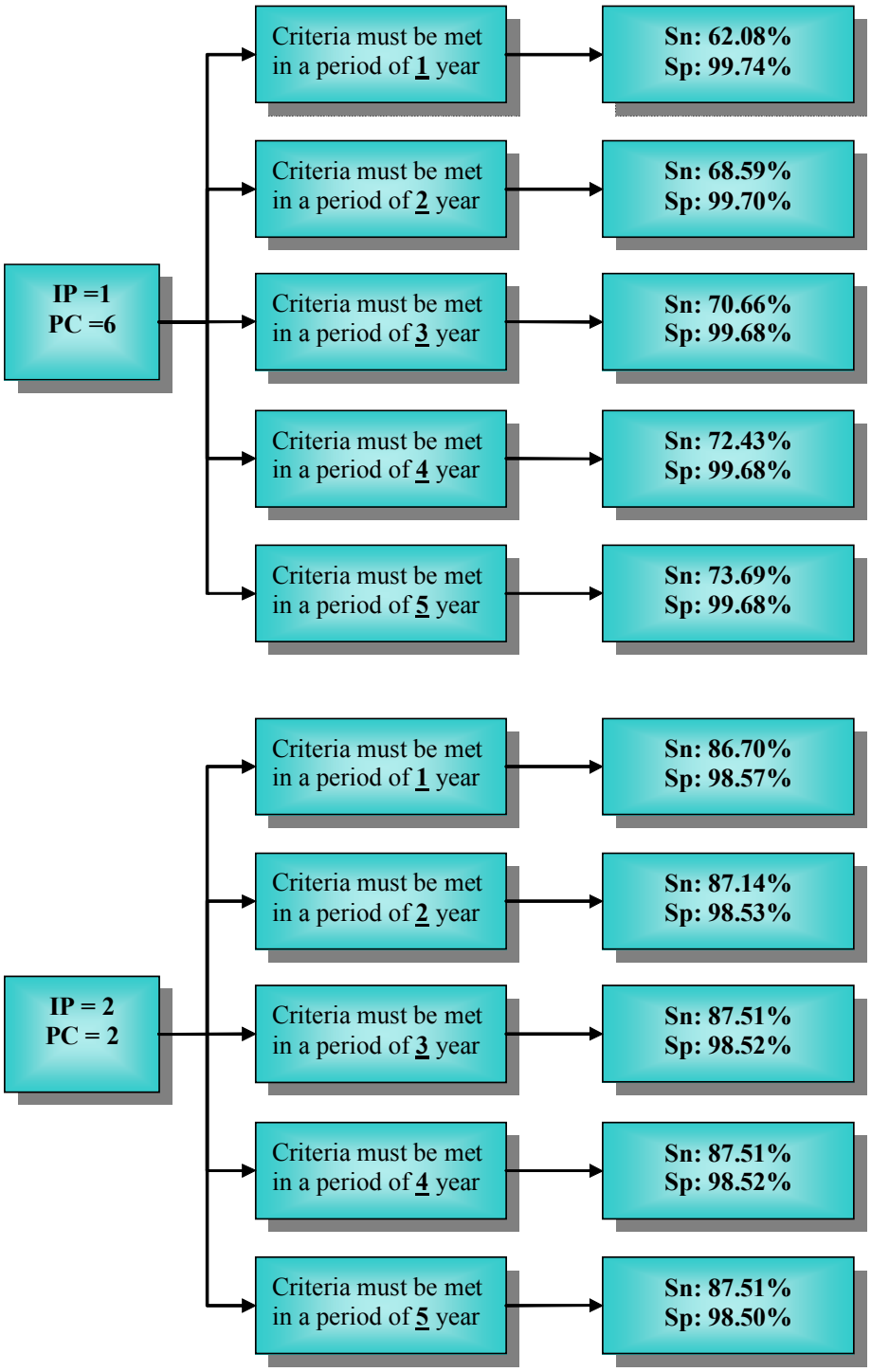

IP = 2
PC = 5
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 76.20%
Sp: 99.65%
Criteria must be met in a period of 2 year
Sn: 79.53%
Sp: 99.63%
Criteria must be met in a period of 3 year
Sn: 81.15%
Sp: 99.61%
Criteria must be met in a period of 4 year
Sn: 82.48%
Sp: 99.60%
Criteria must be met in a period of 5 year
Sn: 83.08%
Sp: 99.59%
IP = 2
PC = 5
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 70.73%
Sp: 99.70%
Criteria must be met in a period of 2 year
Sn: 75.31%
Sp: 99.67%
Criteria must be met in a period of 3 year
Sn: 77.61%
Sp: 99.66%
Criteria must be met in a period of 4 year
Sn: 79.31%
Sp: 99.65%
Criteria must be met in a period of 5 year
Sn: 79.82%
Sp: 99.64%

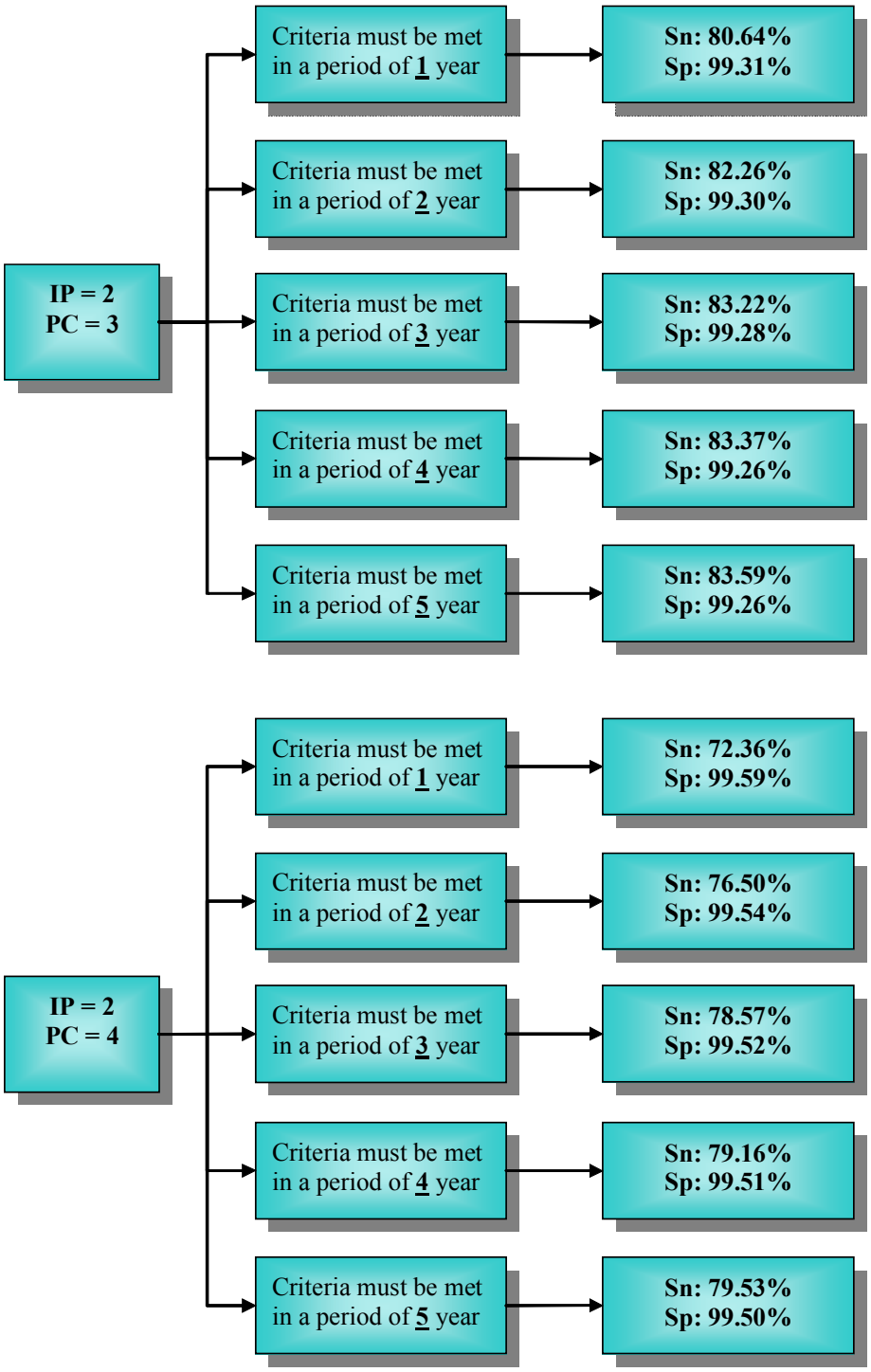

IP = 2
PC = 6
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 72.73%
Sp: 99.71%
Criteria must be met in a period of 2 year
Sn: 76.72%
Sp: 99.70%
Criteria must be met in a period of 3 year
Sn: 78.12%
Sp: 99.68%
Criteria must be met in a period of 4 year
Sn: 79.16%
Sp: 99.68%
Criteria must be met in a period of 5 year
Sn: 80.34%
Sp: 99.68%
IP = 2
PC = 6
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 66.30%
Sp: 99.76%
Criteria must be met in a period of 2 year
Sn: 71.55%
Sp: 99.74%
Criteria must be met in a period of 3 year
Sn: 73.61%
Sp: 99.73%
Criteria must be met in a period of 4 year
Sn: 75.24%
Sp: 99.72%
Criteria must be met in a period of 5 year
Sn: 76.50%
Sp: 99.72%

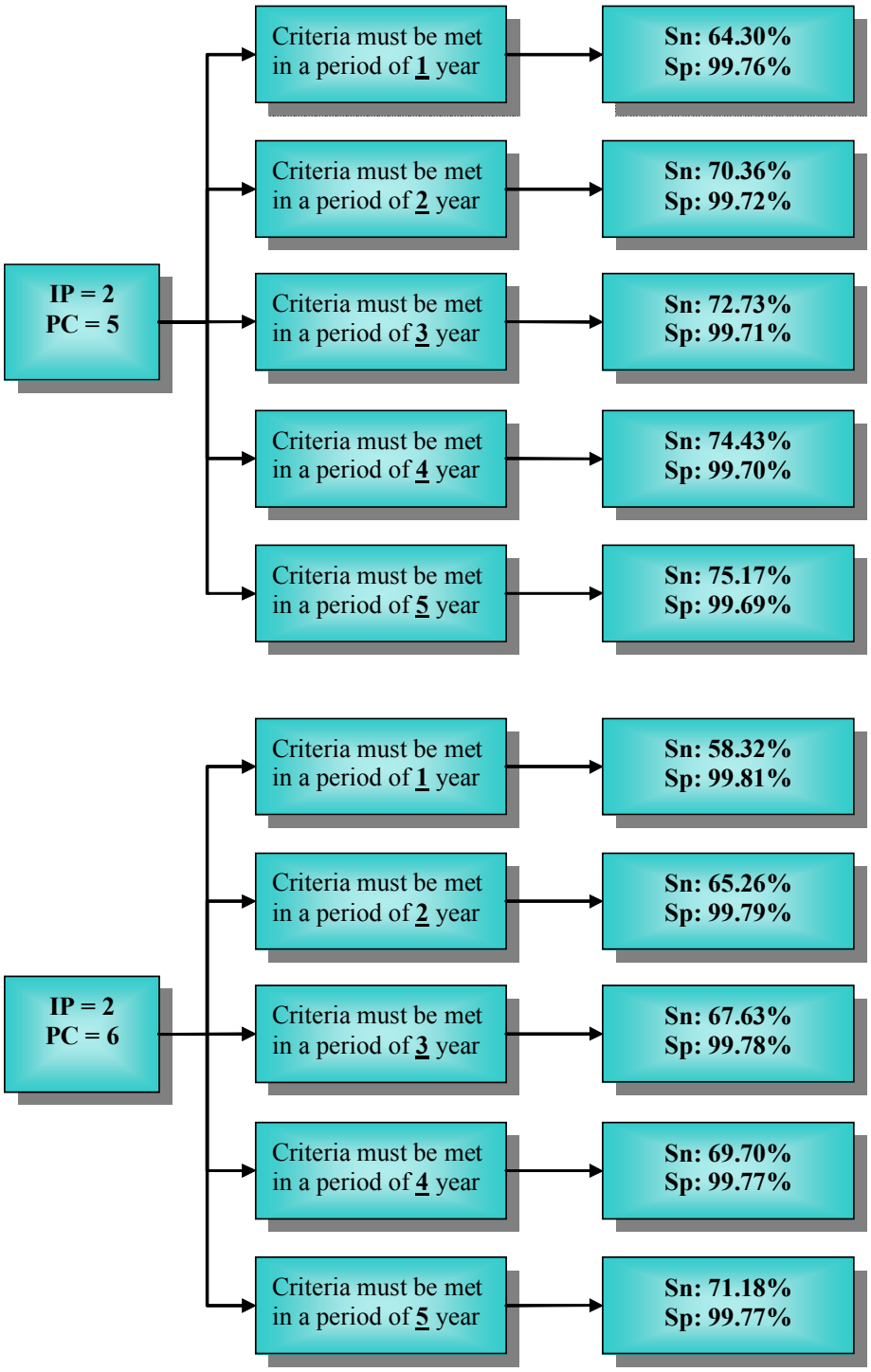
APPENDIX 6. SENSITIVITIES AND SPECIFICITIES FOR 50 PRELIMINARY DEFINITIONS THAT DID NOT CONSIDER ACCS DATABASE




IP = 1
PC =4
Criteria must be met in a period of 1 year
Sn: 74.21%
Sp: 99.54%
Criteria must be met in a period of 2 year
Sn: 77.83%
Sp: 99.47%
Criteria must be met in a period of 3 year
Sn: 79.97%
Sp: 99.45%
Criteria must be met in a period of 4 year
Sn: 80.56%
Sp: 99.45%
Criteria must be met in a period of 5 year
Sn: 80.93%
Sp: 99.43%
IP = 1
PC = 5
Criteria must be met in a period of 1 year
Sn: 67.55%
Sp: 99.68%
Criteria must be met in a period of 2 year
Sn: 72.65%
Sp: 99.63%
Criteria must be met in a period of 3 year
Sn: 74.80%
Sp: 99.61%
Criteria must be met in a period of 4 year
Sn: 76.35%
Sp: 99.60%
Criteria must be met in a period of 5 year
Sn: 77.01%
Sp: 99.59%


IP =1
PC =6
Criteria must be met in a period of 1 year
Sn: 62.08%
Sp: 99.74%
Criteria must be met in a period of 2 year
Sn: 68.59%
Sp: 99.70%
Criteria must be met in a period of 3 year
Sn: 70.66%
Sp: 99.68%
Criteria must be met in a period of 4 year
Sn: 72.43%
Sp: 99.68%
Criteria must be met in a period of 5 year
Sn: 73.69%
Sp: 99.68%
IP = 2
PC = 2
Criteria must be met in a period of 1 year
Sn: 86.70%
Sp: 98.57%
Criteria must be met in a period of 2 year
Sn: 87.14%
Sp: 98.53%
Criteria must be met in a period of 3 year
Sn: 87.51%
Sp: 98.52%
Criteria must be met in a period of 4 year
Sn: 87.51%
Sp: 98.52%
Criteria must be met in a period of 5 year
Sn: 87.51%
Sp: 98.50%


IP = 2
PC = 3
Criteria must be met in a period of 1 year
Sn: 80.64%
Sp: 99.31%
Criteria must be met in a period of 2 year
Sn: 82.26%
Sp: 99.30%
Criteria must be met in a period of 3 year
Sn: 83.22%
Sp: 99.28%
Criteria must be met in a period of 4 year
Sn: 83.37%
Sp: 99.26%
Criteria must be met in a period of 5 year
Sn: 83.59%
Sp: 99.26%
IP = 2
PC = 4
Criteria must be met in a period of 1 year
Sn: 72.36%
Sp: 99.59%
Criteria must be met in a period of 2 year
Sn: 76.50%
Sp: 99.54%
Criteria must be met in a period of 3 year
Sn: 78.57%
Sp: 99.52%
Criteria must be met in a period of 4 year
Sn: 79.16%
Sp: 99.51%
Criteria must be met in a period of 5 year
Sn: 79.53%
Sp: 99.50%


IP = 2
PC = 5
Criteria must be met in a period of 1 year
Sn: 64.30%
Sp: 99.76%
Criteria must be met in a period of 2 year
Sn: 70.36%
Sp: 99.72%
Criteria must be met in a period of 3 year
Sn: 72.73%
Sp: 99.71%
Criteria must be met in a period of 4 year
Sn: 74.43%
Sp: 99.70%
Criteria must be met in a period of 5 year
Sn: 75.17%
Sp: 99.69%
IP = 2
PC = 6
Criteria must be met in a period of 1 year
Sn: 58.32%
Sp: 99.81%
Criteria must be met in a period of 2 year
Sn: 65.26%
Sp: 99.79%
Criteria must be met in a period of 3 year
Sn: 67.63%
Sp: 99.78%
Criteria must be met in a period of 4 year
Sn: 69.70%
Sp: 99.77%
Criteria must be met in a period of 5 year
Sn: 71.18%
Sp: 99.77%

APPENDIX 7. 2×2 TABLES AND TEST CHARACTERISTICS PREFERABLE PRELIMINARY ADMINISTRATIVE DEFINITIONS OUT OF 150 PRELIMINARY DEFINITIONS.

A. Definitions considering ACCS database

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,198	155
	Non-IBD	131	15,375

Sn: 88.54% (95%CI=86.73,90.19)

Sp: 99.16% (95% CI=99.00,99.29)

PPV: 90.14%(95%CI=88.41,91.69)

NPV:99.00%(95%CI=98.83,99.15)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,207	146
	Non-IBD	134	15,375

Sn: 89.21% (95%CI=87.43,90.81)

Sp: 99.14% (95% CI=98.98,99.28)

PPV: 90.01%(95%CI=88.28,91.56)

NPV:99.06%(95%CI=98.89,99.21)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,210	143
	Non-IBD	137	15,369

Sn: 89.43% (95%CI=87.67,91.02)

Sp: 99.12% (95% CI=98.96,99.26)

PPV: 89.83%(95%CI=88.09,91.39)

NPV:99.08%(95%CI=98.92,99.22)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,212	141
	Non-IBD	137	15,369

Sn: 89.58% (95%CI=87.83,91.16)

Sp: 99.12% (95% CI=98.96,99.26)

PPV: 89.84%(95%CI=88.11,91.41)

NPV:99.09%(95%CI=98.93,99.23)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,141	212
	Non-IBD	95	15,411

Sn: 84.33% (95%CI=82.28,86.23)

Sp: 99.39% (95% CI=99.25,99.50)

PPV: 92.31%(95%CI=90.69,93.74)

NPV:98.64%(95%CI=98.45,98.82)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,157	186
	Non-IBD	98	15,408

Sn: 85.51% (95%CI=83.52,87.35)

Sp: 99.37% (95% CI=99.23,99.49)

PPV: 92.19%(95%CI=90.57,93.62)

NPV:98.74%(95%CI=98.56,98.91)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,164	189
	Non-IBD	99	15,407

Sn: 86.03% (95%CI=84.07,87.86)

Sp: 99.36% (95% CI=99.22,99.48)

PPV: 92.16%(95%CI=90.54,93.58)

NPV:98.79%(95%CI=98.60,98.95)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,169	184
	Non-IBD	101	15,405

Sn: 86.40% (95%CI=84.46,88.18)

Sp: 99.35% (95% CI=99.21,99.47)

PPV: 92.05%(95%CI=90.42,93.48)

NPV:98.82%(95%CI=98.64,98.98)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,098	255
	Non-IBD	71	15,435

Sn: 81.15% (95%CI=78.97,83.20)

Sp: 99.54% (95% CI=99.42,99.64)

PPV: 93.93%(95%CI=92.40,95.23)

NPV:98.38%(95%CI=98.17,98.57)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,115	238
	Non-IBD	74	15,432

Sn: 82.41% (95%CI=80.27,84.40)

Sp: 99.52% (95% CI=99.40,99.63)

PPV: 93.78%(95%CI=92.25,95.08)

NPV:98.48%(95%CI=98.28,98.66)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,133	220
	Non-IBD	76	15,430

Sn: 83.74% (95%CI=81.66,85.67)

Sp: 99.51% (95% CI=99.39,99.61)

PPV: 93.71%(95%CI=92.20,95.02)

NPV:98.59%(95%CI=98.40,98.77)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,140	213
	Non-IBD	77	15,429

Sn: 84.26% (95%CI=82.21,86.16)

Sp: 99.50% (95% CI=99.38,99.61)

PPV: 93.67%(95%CI=92.16,94.98)

NPV:98.64%(95%CI=98.44,98.81)

B. Definitions exclusive of ACCS database

- Definition "IP≥1 *or* PC≥3 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,126	227
	Non-IBD	119	15,387

Sn: 83.22% (95%CI=81.12,85.18)

Sp: 99.23% (95% CI=99.08,99.36)

PPV: 90.44%(95%CI=88.67,92.02)

NPV:98.55%(95%CI=98.35,98.73)

- Definition "IP≥1 *or* PC>3 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,141	212
	Non-IBD	122	15,384

Sn: 84.33% (95%CI=82.28,86.23)

Sp: 99.21% (95% CI=99.06,99.35)

PPV: 90.34%(95%CI=88.58,91.91)

NPV:98.64%(95%CI=98.45,98.82)

- Definition "IP≥1 *or* PC≥3 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,143	210
	Non-IBD	125	15,381

Sn: 84.48% (95%CI=82.44,86.37)

Sp: 99.20% (95% CI=99.04,99.33)

PPV: 90.14%(95%CI=88.37,91.73)

NPV:98.65%(95%CI=98.46,98.83)

- Definition "IP≥1 *or* PC≥3 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,146	207
	Non-IBD	125	15,381

Sn: 84.70% (95%CI=82.67,86.58)

Sp: 99.19% (95% CI=99.04,99.33)

PPV: 90.17%(95%CI=88.40,91.75)

NPV:98.67%(95%CI=98.48,98.85)

- Definition "IP≥1 *or* PC≥4 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,053	300
	Non-IBD	82	15,424

Sn: 77.83% (95%CI=75.52,80.02)

Sp: 99.47% (95% CI=99.34,99.58)

PPV: 92.76%(95%CI=91.11,94.21)

NPV:98.09%(95%CI=97.87,98.30)

- Definition "IP≥1 *or* PC≥4 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,082	271
	Non-IBD	85	15,421

Sn: 79.97% (95%CI=77.74,82.07)

Sp: 99.45% (95% CI=99.32,99.56)

PPV:92.72%(95%CI=91.07,94.14)

NPV:98.27%(95%CI=98.06,98.47)

- Definition "IP≥1 *or* PC≥4 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,090	263
	Non-IBD	86	15,420

Sn: 80.56% (95%CI=78.35,82.64)

Sp: 99.45% (95% CI=99.32,99.56)

PPV: 92.69%(95%CI=91.05,94.11)

NPV:98.32%(95%CI=98.11,98.52)

- Definition "IP≥1 *or* PC≥4 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,095	258
	Non-IBD	88	15,418

Sn: 80.93% (95%CI=78.74,82.99)

Sp: 99.43% (95% CI=99.30,99.55)

PPV: 92.56%(95%CI=90.92,93.99)

NPV:98.35%(95%CI=98.14,98.59)

- Definition "IP≥1 *or* PC≥5 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	983	370
	Non-IBD	57	15,449

Sn: 72.65% (95%CI=70.19,75.01)

Sp: 99.63% (95% CI=99.52,99.72)

PPV: 94.52%(95%CI=92.96,95.82)

NPV:97.66%(95%CI=97.41,97.89)

- Definition "IP≥1 *or* PC≥5 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,012	341
	Non-IBD	60	15,446

Sn: 74.80% (95%CI=72.39,77.09)

Sp: 99.61% (95% CI=99.50,99.71)

PPV: 94.40%(95%CI=92.85,95.70)

NPV:97.84%(95%CI=97.60,98.06)

- Definition "IP≥1 *or* PC≥5 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,033	320
	Non-IBD	62	15,444

Sn: 76.35% (95%CI=73.99,78.59)

Sp: 99.60% (95% CI=99.49,99.69)

PPV: 94.34%(95%CI=92.80,95.63)

NPV:97.97%(95%CI=97.74,98.18)

- Definition "IP≥1 *or* PC≥5 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro search (Gold standard test)	IBD	1,042	311
	Non-IBD	63	15,443

Sn: 77.01% (95%CI=74.68,79.23)

Sp: 99.59% (95% CI=99.48,99.69)

PPV: 94.30%(95%CI=92.76,95.59)

NPV:98.03%(95%CI=97.80,98.24)

APPENDIX 8. HEALTH SERVICE RECORDS OF OUR PATIENTS (AFTER CHART REVIEWS AND FINAL CLASSIFICATION) IN IP, PC AND ACCS DATABASES.

- **IP database**

Number of subjects with at least one hospitalization and frequency of IBD diagnostic codes in hospital discharge abstracts among IBD and non-IBD patients are shown in Diagram 5.

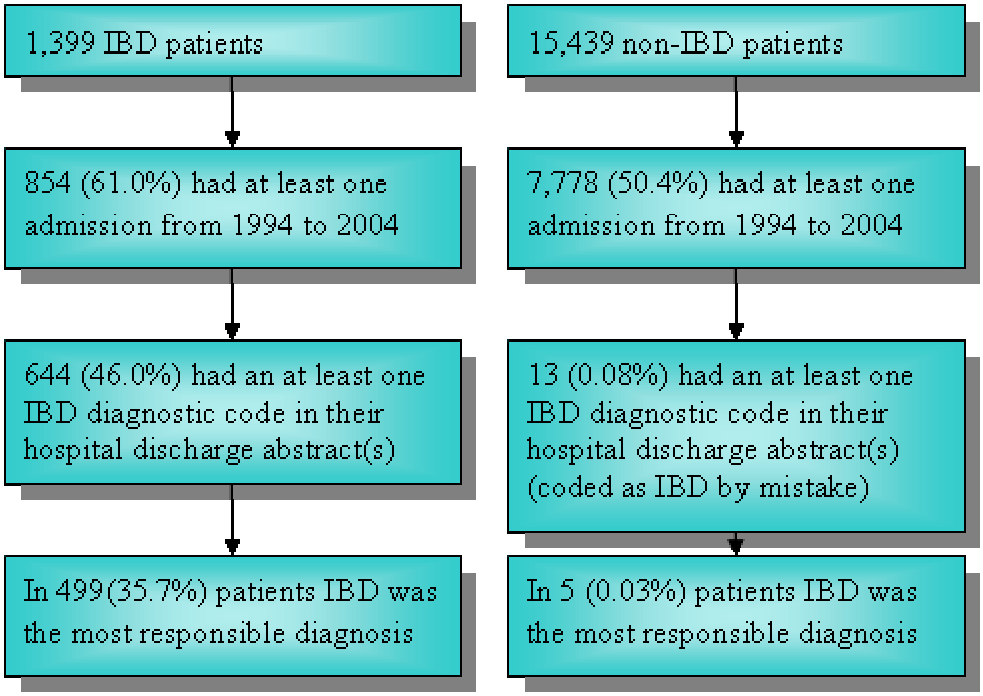


Diagram 5. Number of hospitalizations and IBD diagnostic codes in IBD and non-IBD patients after final classification.

IBD patients had a median number of 2 hospitalizations with a diagnostic code of IBD within the period of 1994-2004 (Mean, 2.49±3.13; range, 0-33).Within the same period

IBD patients had a median number of 1 hospitalization with a most responsible diagnosis of IBD (Mean, 1.44±2.05; range, 0-20).

- **PC database**

In total, 28,586 claims were identified with an ICD code of IBD (25,061claims in IBD patients, 2,669in possible IBD cases, and 856 claims in non-IBD patients).

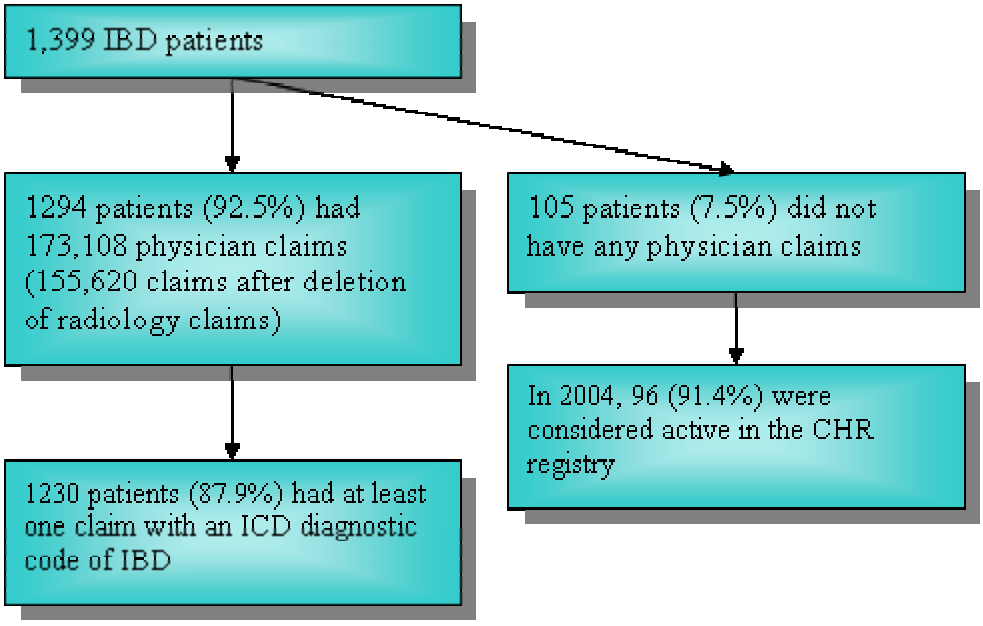


Diagram 6. Frequency of physician claims and IBD diagnostic codes among IBD patients after final classification.

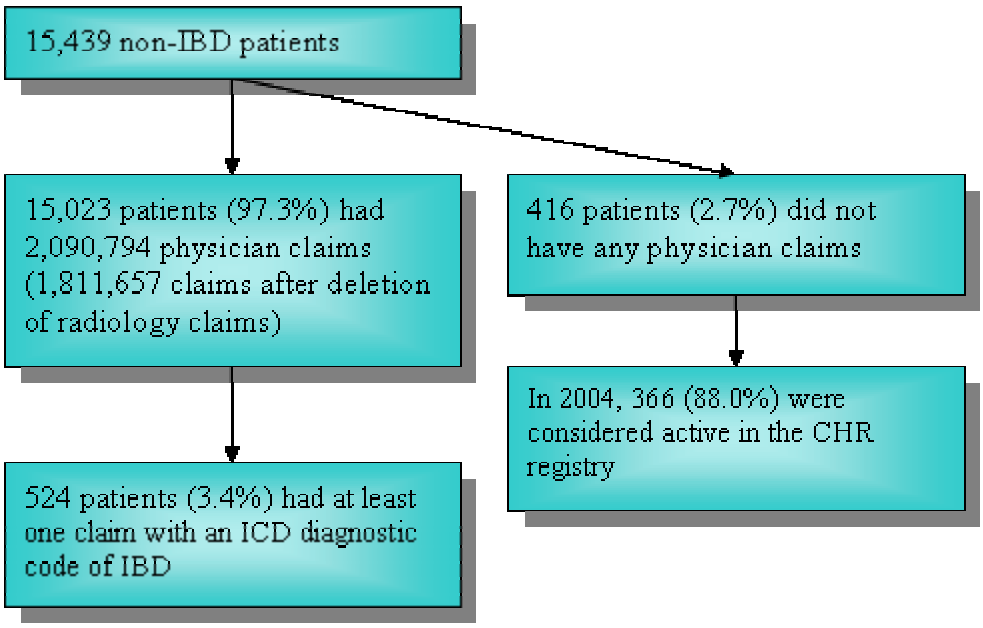


Diagram 7. Frequency of physician claims and IBD diagnostic codes among non-IBD patients after final classification.

IBD patients had a median number of 13 claims with a diagnostic code of IBD within the period of 1995-2004 (Mean, 19.59±23.45; range, 0-240). Non-IBD patients had a median number of 0 claims (95^{th} percentile = 0) with a diagnostic code of IBD within the period of 1995-2004 (Mean, 0.06±0.47; range, 0-29 (second largest = 11)).

For IBD patients on average IBD claims were 84.7±191.9 days apart. A year (365.25 days) lied on the 94.6^{th} percentile for the time difference between the IBD claims.

- **ACCS database**

Distribution of ACCS contacts is shown in Diagram 8.

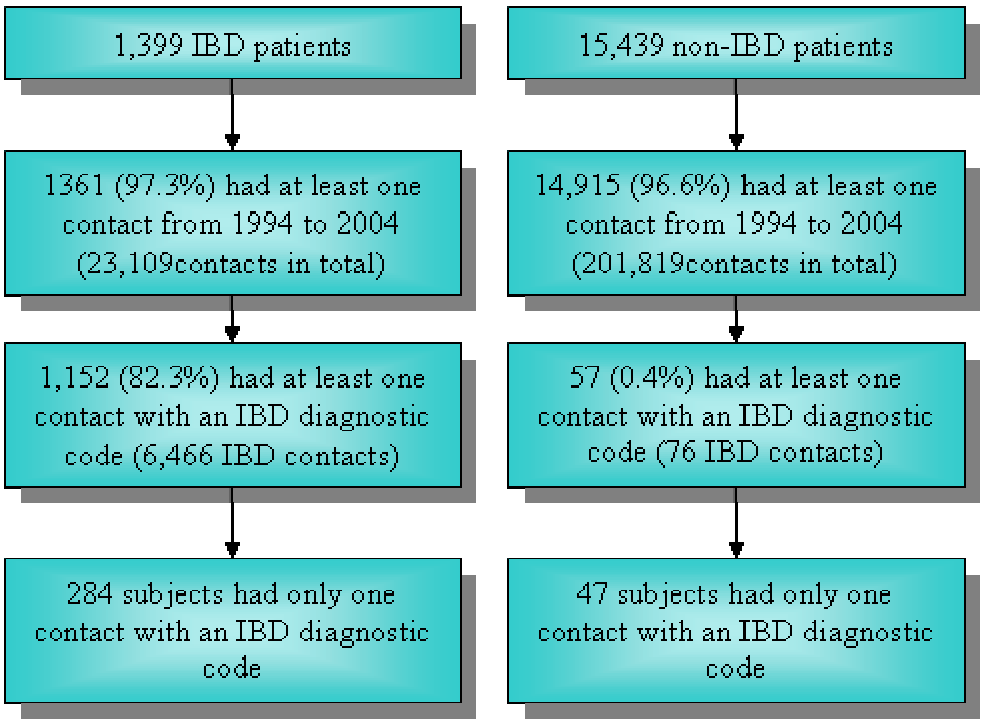
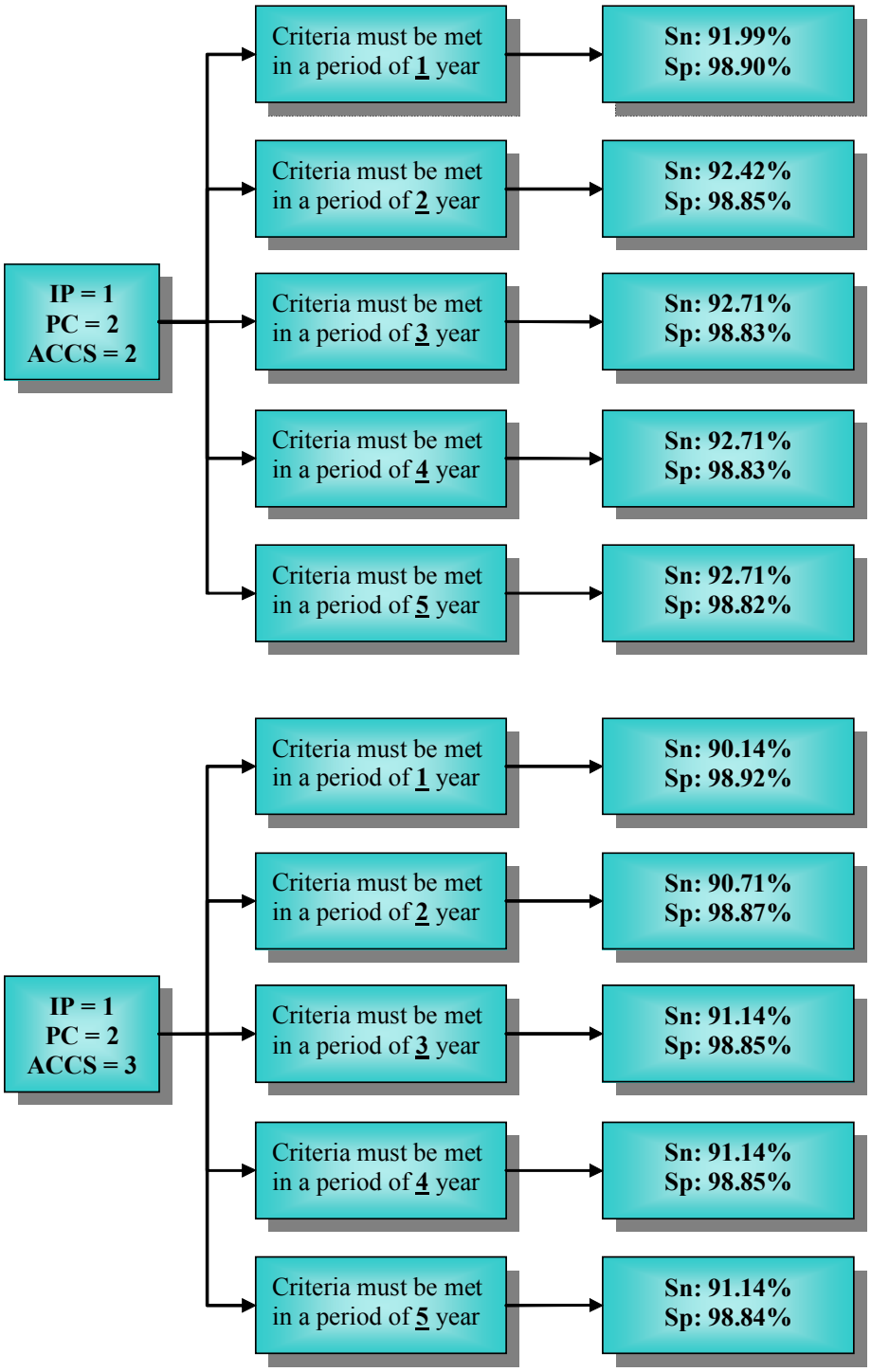
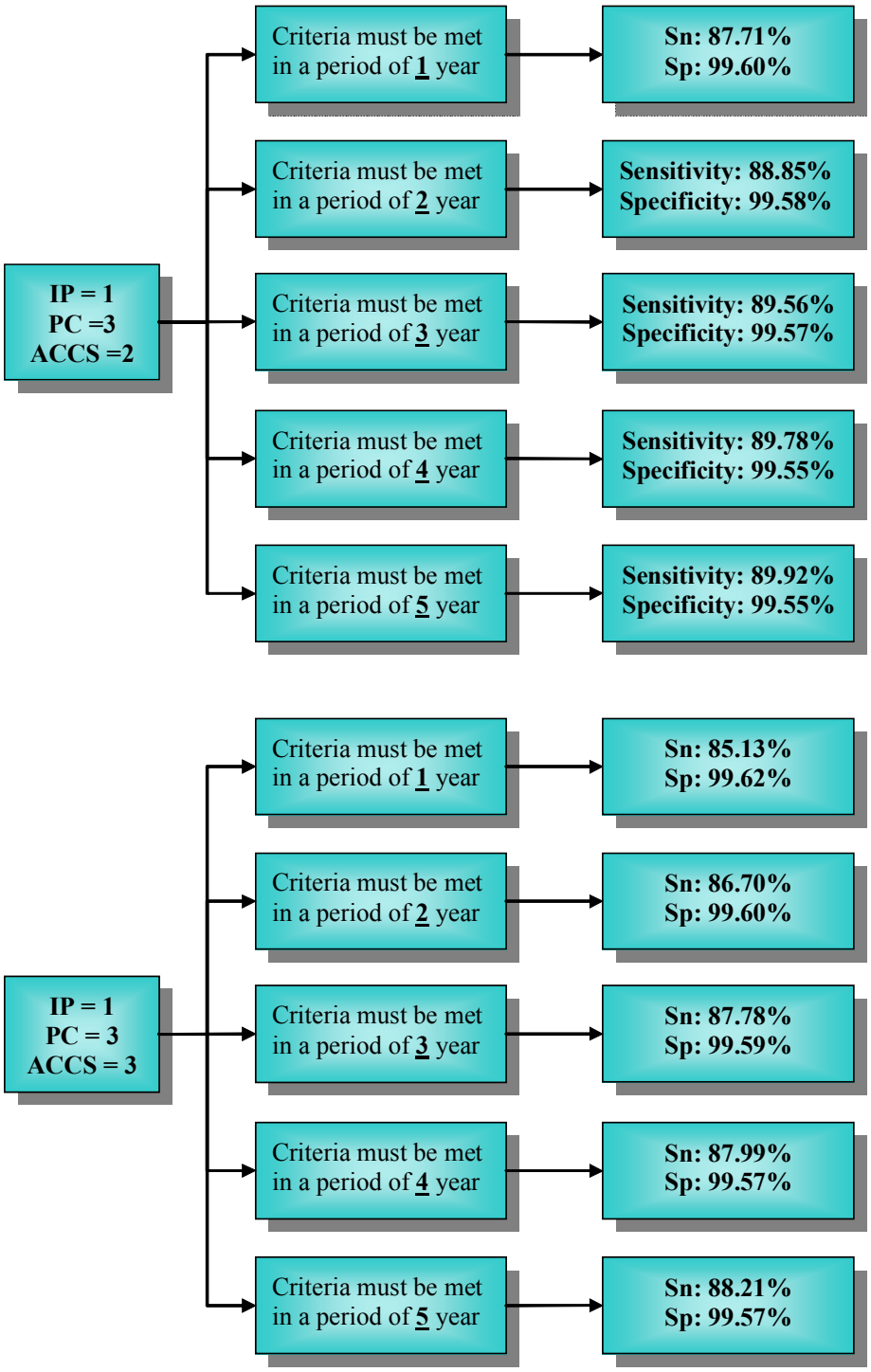


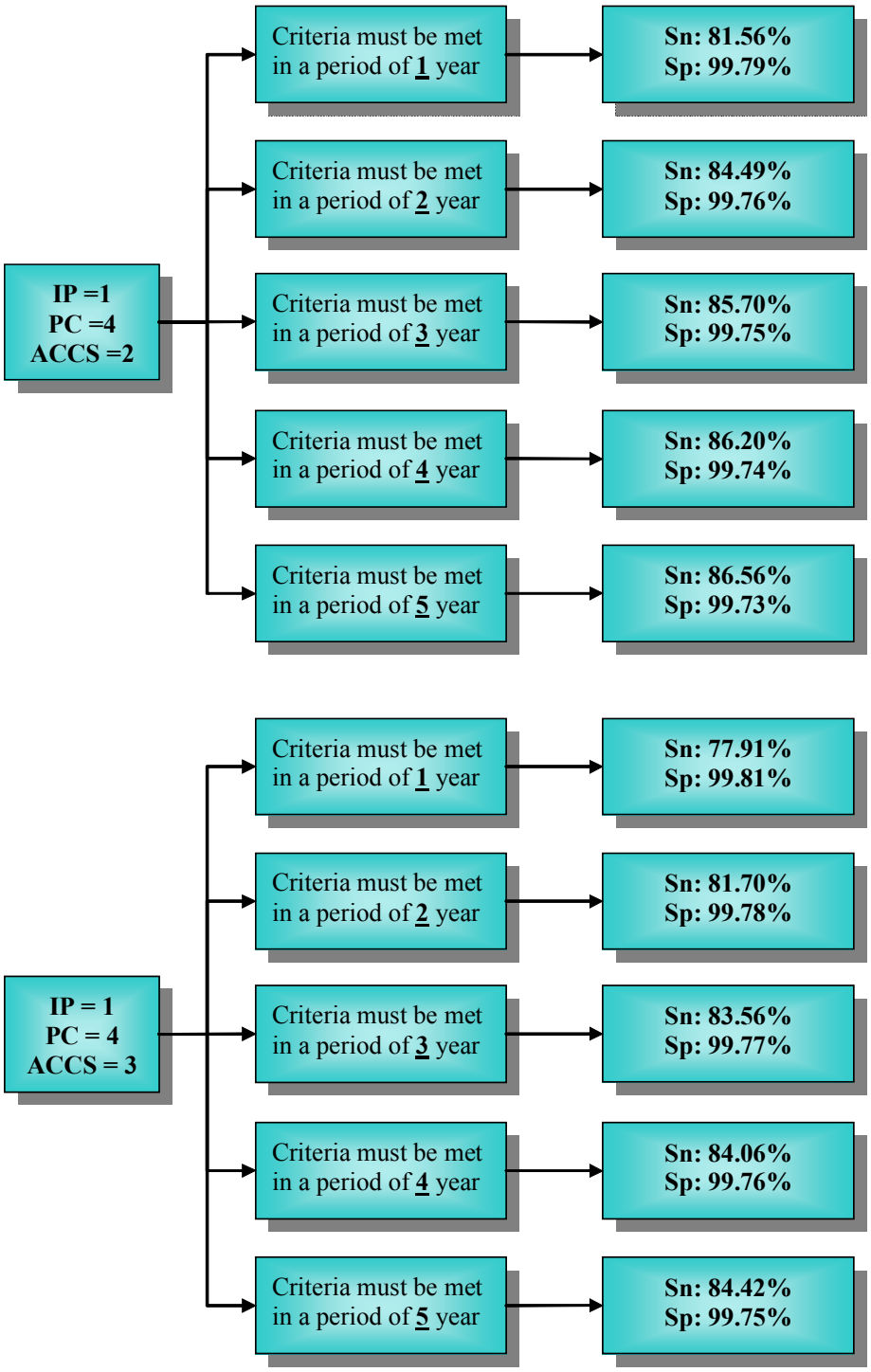
Diagram 8. Frequency of ACCS contacts and IBD diagnostic codes among IBD and non-IBD patients after final classification.

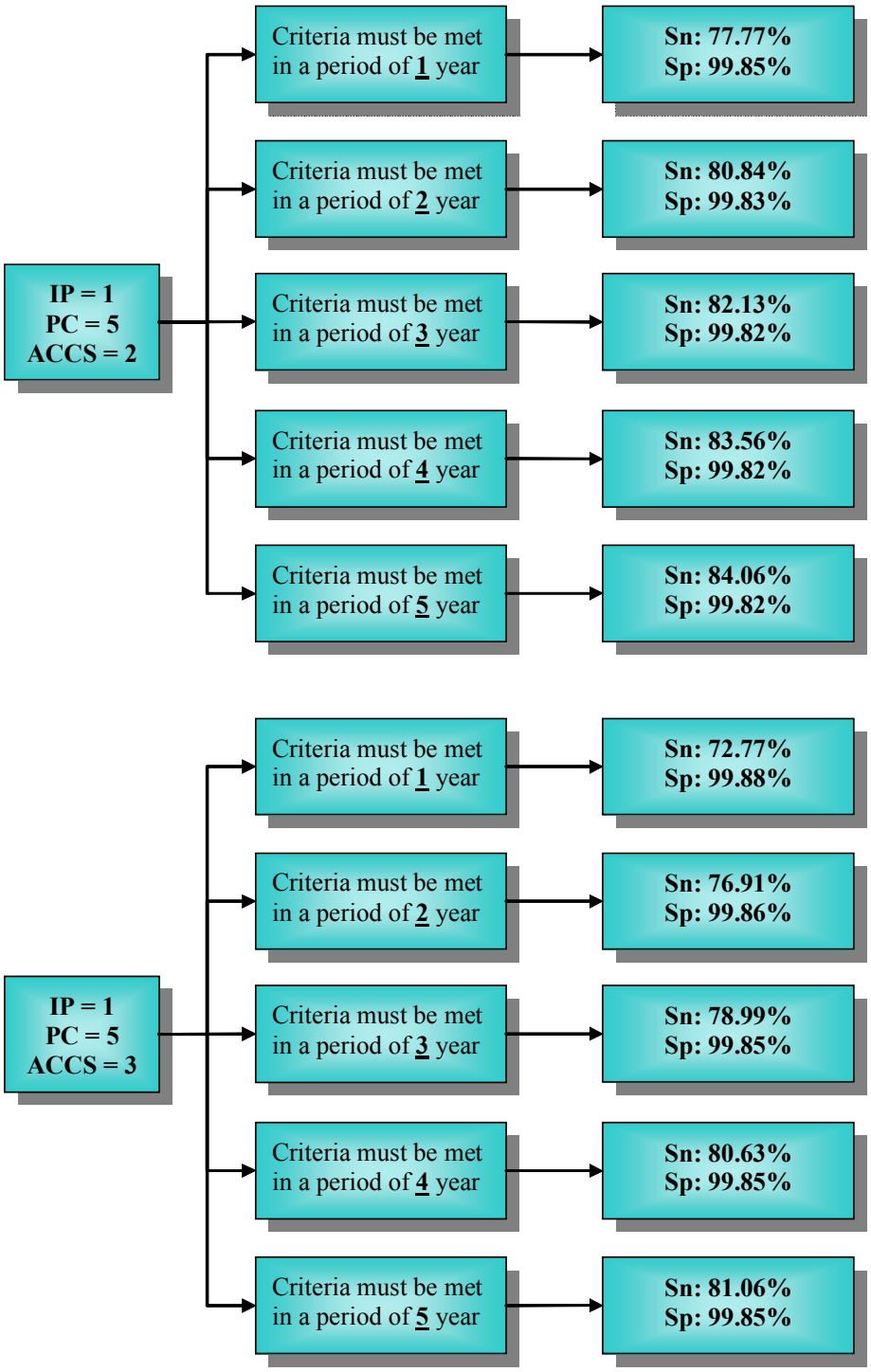
IBD patients had a median number of 2 ACCS contacts with a diagnostic code of IBD within the period of 1996-2004 (Mean, 4.78±7.79; range, 0-87). Non-IBD patients had a median number of 0 claims (99^{th} percentile = 0) with a diagnostic code of IBD within the period of 1996-2004 (Mean, 0.02±0.6; range, 0-62).

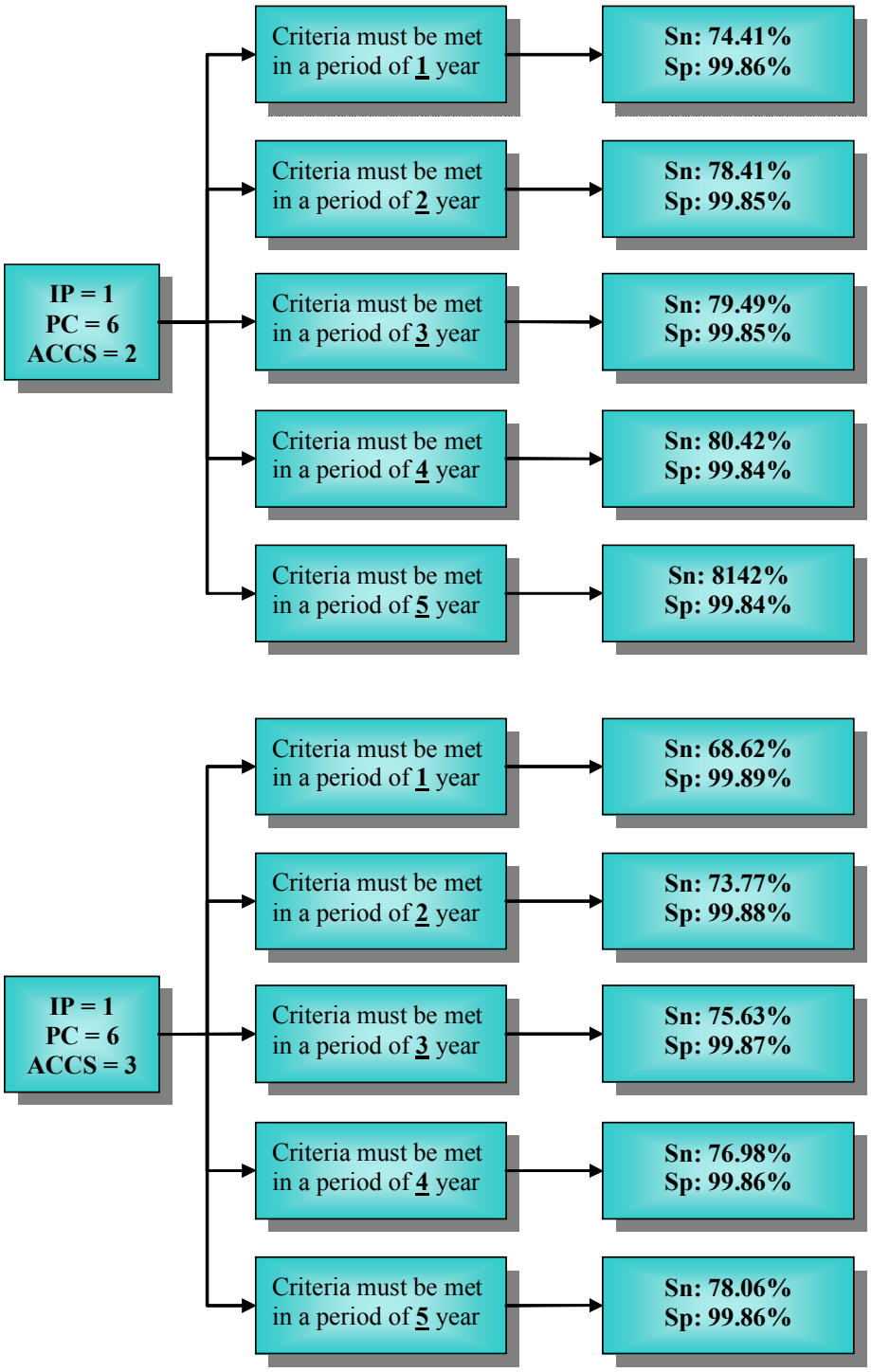
APPENDIX 9. SENSITIVITIES AND SPECIFICITIES FOR 100 DEFINITIONS INCLUSIVE OF ACCS DATABASE (FINAL CLASSIFICATION)

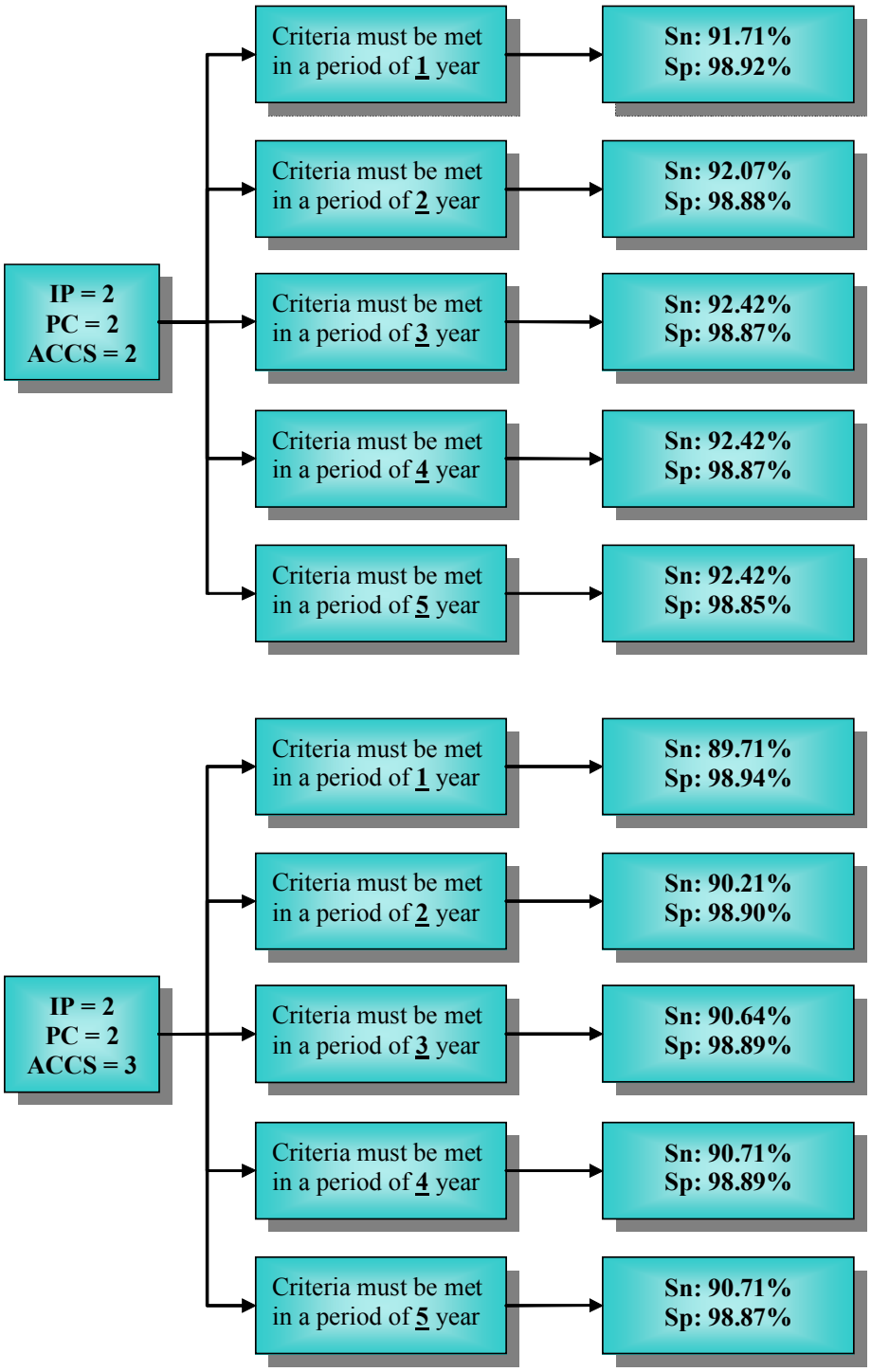


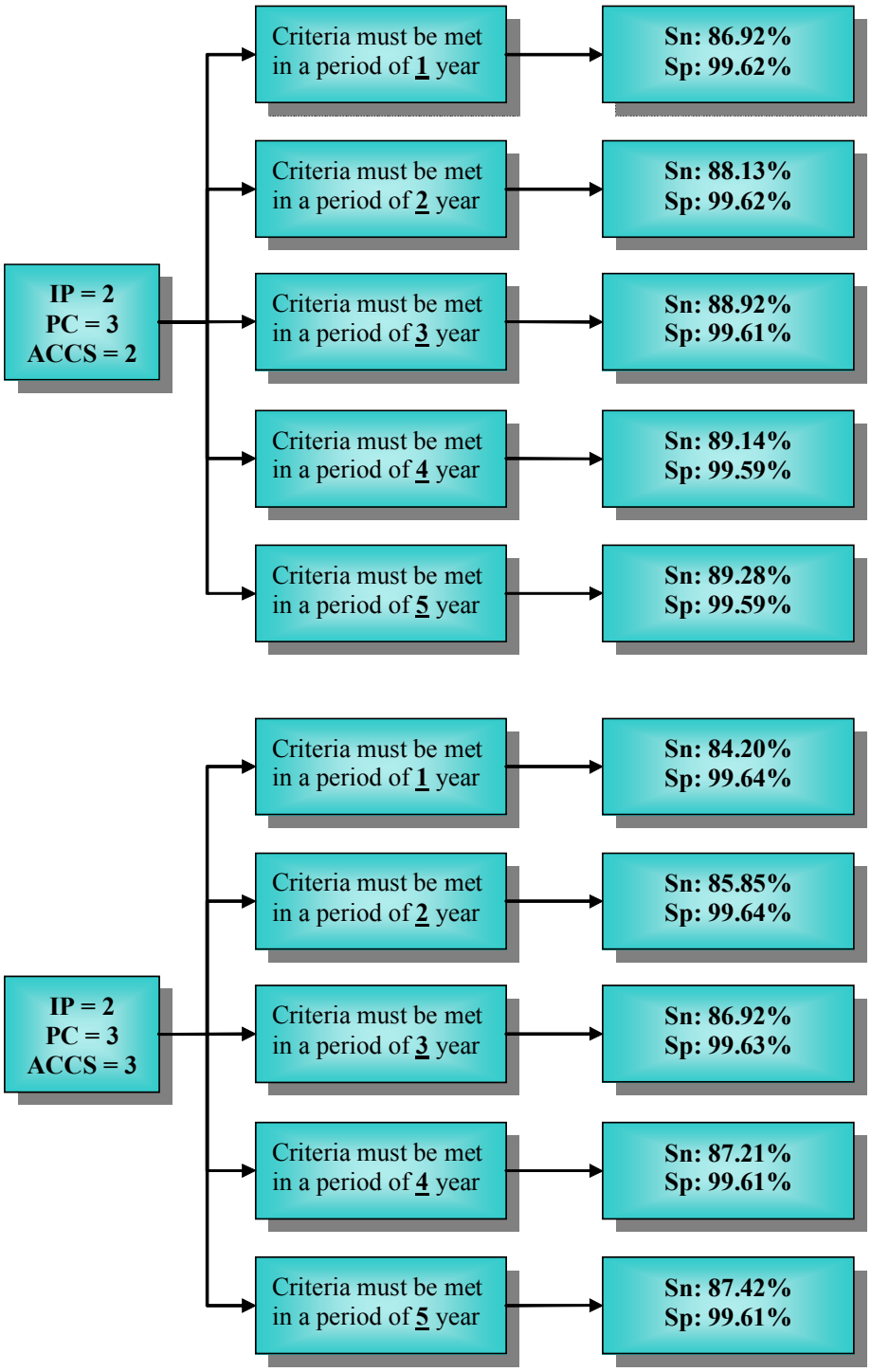

IP = 1
PC =3
ACCS =2
Criteria must be met in a period of 1 year
Sn: 87.71%
Sp: 99.60%
Criteria must be met in a period of 2 year
Sensitivity: 88.85%
Specificity: 99.58%
Criteria must be met in a period of 3 year
Sensitivity: 89.56%
Specificity: 99.57%
Criteria must be met in a period of 4 year
Sensitivity: 89.78%
Specificity: 99.55%
Criteria must be met in a period of 5 year
Sensitivity: 89.92%
Specificity: 99.55%
IP = 1
PC = 3
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 85.13%
Sp: 99.62%
Criteria must be met in a period of 2 year
Sn: 86.70%
Sp: 99.60%
Criteria must be met in a period of 3 year
Sn: 87.78%
Sp: 99.59%
Criteria must be met in a period of 4 year
Sn: 87.99%
Sp: 99.57%
Criteria must be met in a period of 5 year
Sn: 88.21%
Sp: 99.57%

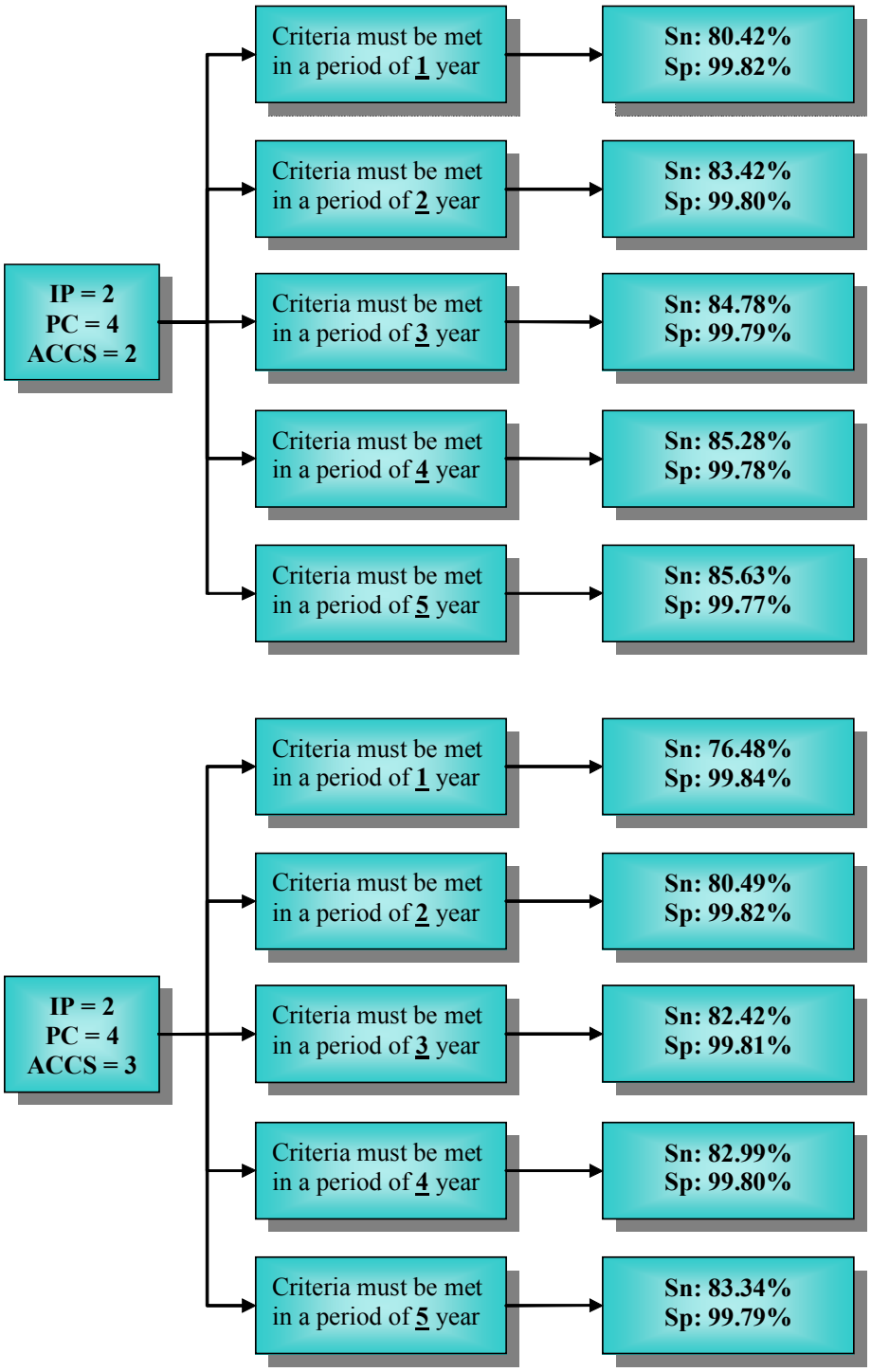

IP =1
PC =4
ACCS =2
Criteria must be met in a period of 1 year
Sn: 81.56%
Sp: 99.79%
Criteria must be met in a period of 2 year
Sn: 84.49%
Sp: 99.76%
Criteria must be met in a period of 3 year
Sn: 85.70%
Sp: 99.75%
Criteria must be met in a period of 4 year
Sn: 86.20%
Sp: 99.74%
Criteria must be met in a period of 5 year
Sn: 86.56%
Sp: 99.73%
IP = 1
PC = 4
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 77.91%
Sp: 99.81%
Criteria must be met in a period of 2 year
Sn: 81.70%
Sp: 99.78%
Criteria must be met in a period of 3 year
Sn: 83.56%
Sp: 99.77%
Criteria must be met in a period of 4 year
Sn: 84.06%
Sp: 99.76%
Criteria must be met in a period of 5 year
Sn: 84.42%
Sp: 99.75%

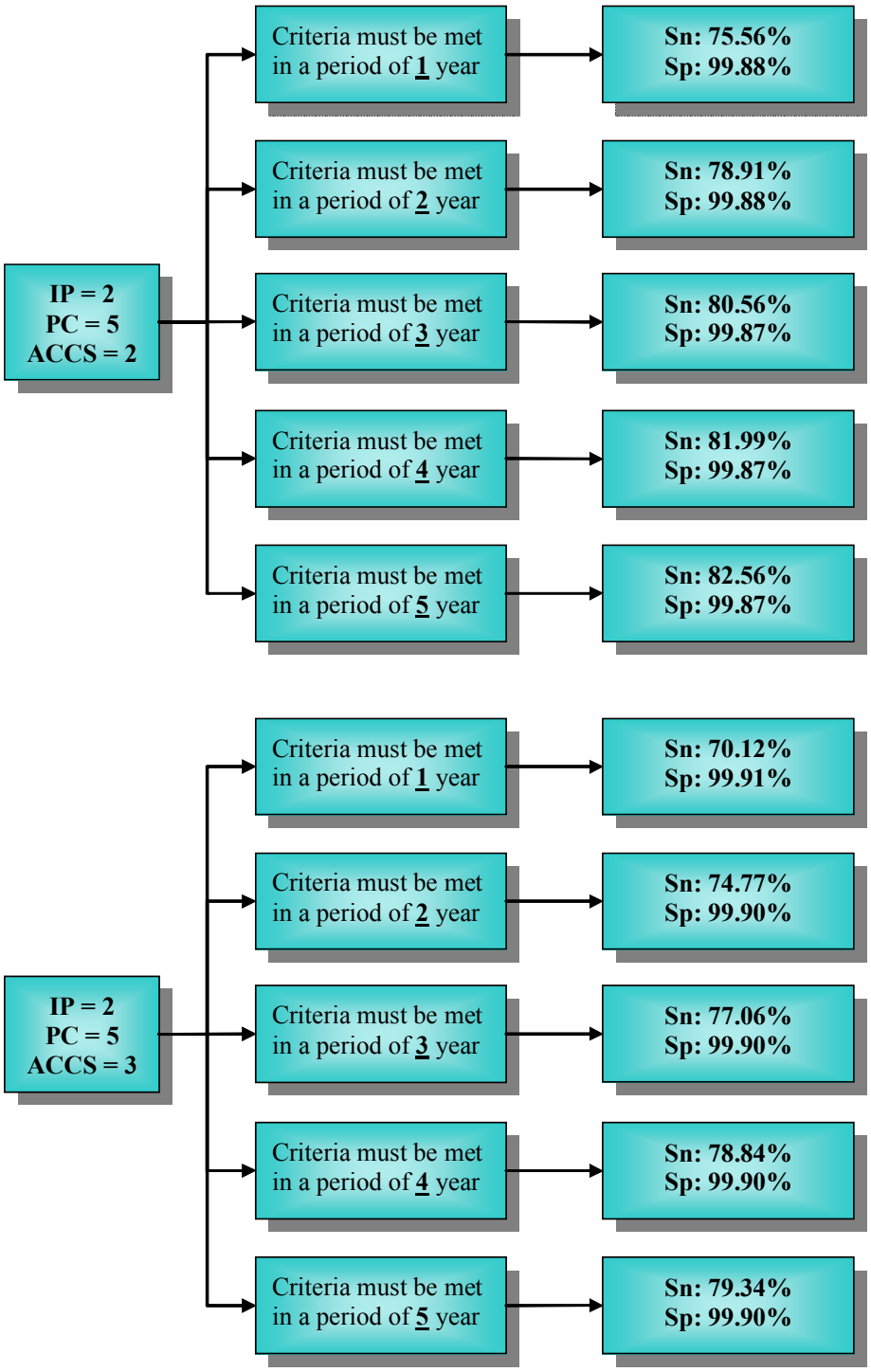

IP = 1
PC = 5
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 77.77%
Sp: 99.85%
Criteria must be met in a period of 2 year
Sn: 80.84%
Sp: 99.83%
Criteria must be met in a period of 3 year
Sn: 82.13%
Sp: 99.82%
Criteria must be met in a period of 4 year
Sn: 83.56%
Sp: 99.82%
Criteria must be met in a period of 5 year
Sn: 84.06%
Sp: 99.82%
IP = 1
PC = 5
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 72.77%
Sp: 99.88%
Criteria must be met in a period of 2 year
Sn: 76.91%
Sp: 99.86%
Criteria must be met in a period of 3 year
Sn: 78.99%
Sp: 99.85%
Criteria must be met in a period of 4 year
Sn: 80.63%
Sp: 99.85%
Criteria must be met in a period of 5 year
Sn: 81.06%
Sp: 99.85%

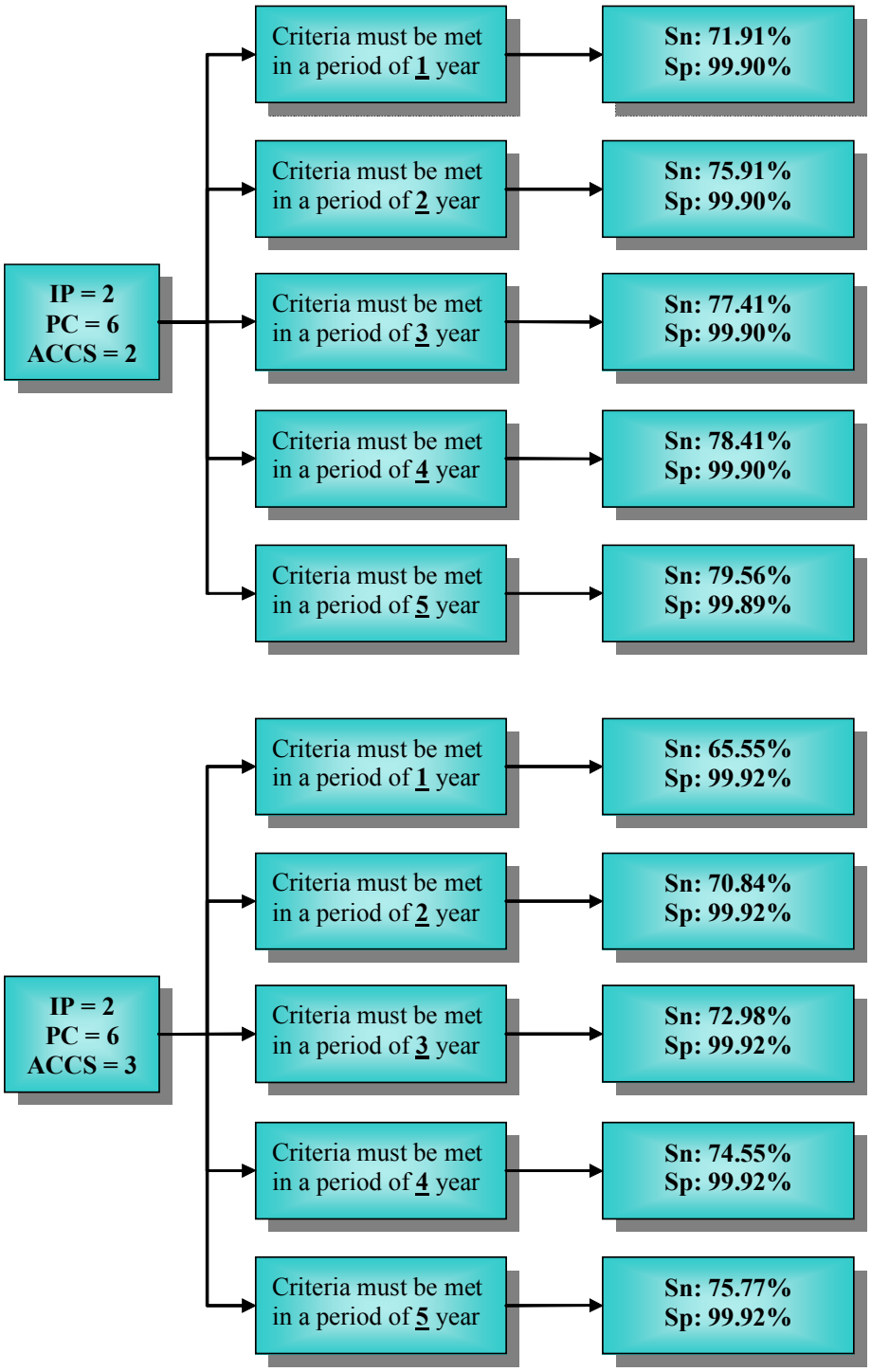

IP = 1
PC = 6
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 74.41%
Sp: 99.86%
Criteria must be met in a period of 2 year
Sn: 78.41%
Sp: 99.85%
Criteria must be met in a period of 3 year
Sn: 79.49%
Sp: 99.85%
Criteria must be met in a period of 4 year
Sn: 80.42%
Sp: 99.84%
Criteria must be met in a period of 5 year
Sn: 8142%
Sp: 99.84%
IP = 1
PC = 6
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 68.62%
Sp: 99.89%
Criteria must be met in a period of 2 year
Sn: 73.77%
Sp: 99.88%
Criteria must be met in a period of 3 year
Sn: 75.63%
Sp: 99.87%
Criteria must be met in a period of 4 year
Sn: 76.98%
Sp: 99.86%
Criteria must be met in a period of 5 year
Sn: 78.06%
Sp: 99.86%


IP = 2
PC = 2
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 91.71%
Sp: 98.92%
Criteria must be met in a period of 2 year
Sn: 92.07%
Sp: 98.88%
Criteria must be met in a period of 3 year
Sn: 92.42%
Sp: 98.87%
Criteria must be met in a period of 4 year
Sn: 92.42%
Sp: 98.87%
Criteria must be met in a period of 5 year
Sn: 92.42%
Sp: 98.85%
IP = 2
PC = 2
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 89.71%
Sp: 98.94%
Criteria must be met in a period of 2 year
Sn: 90.21%
Sp: 98.90%
Criteria must be met in a period of 3 year
Sn: 90.64%
Sp: 98.89%
Criteria must be met in a period of 4 year
Sn: 90.71%
Sp: 98.89%
Criteria must be met in a period of 5 year
Sn: 90.71%
Sp: 98.87%

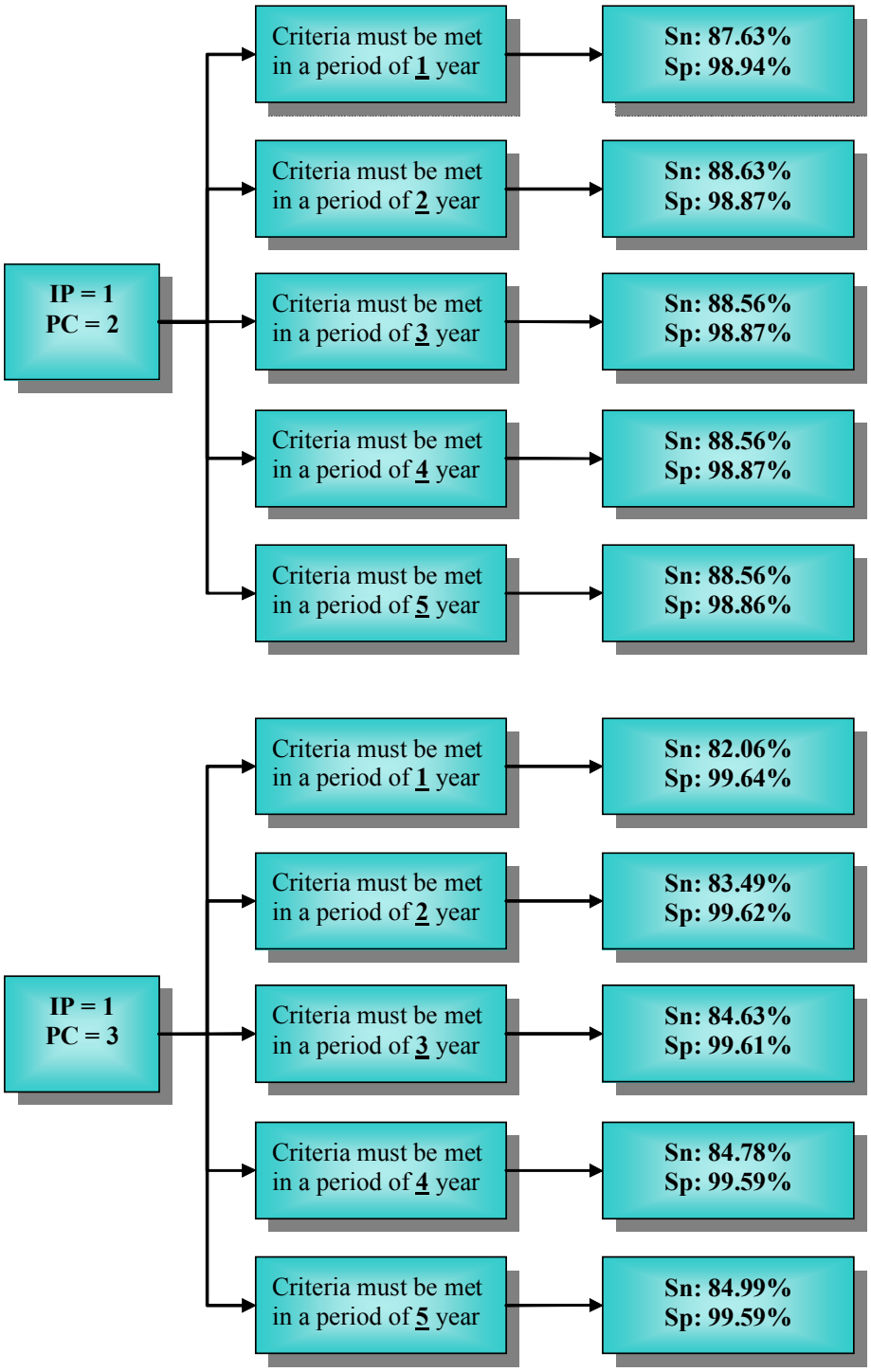

IP = 2
PC = 3
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 86.92%
Sp: 99.62%
Criteria must be met in a period of 2 year
Sn: 88.13%
Sp: 99.62%
Criteria must be met in a period of 3 year
Sn: 88.92%
Sp: 99.61%
Criteria must be met in a period of 4 year
Sn: 89.14%
Sp: 99.59%
Criteria must be met in a period of 5 year
Sn: 89.28%
Sp: 99.59%
IP = 2
PC = 3
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 84.20%
Sp: 99.64%
Criteria must be met in a period of 2 year
Sn: 85.85%
Sp: 99.64%
Criteria must be met in a period of 3 year
Sn: 86.92%
Sp: 99.63%
Criteria must be met in a period of 4 year
Sn: 87.21%
Sp: 99.61%
Criteria must be met in a period of 5 year
Sn: 87.42%
Sp: 99.61%

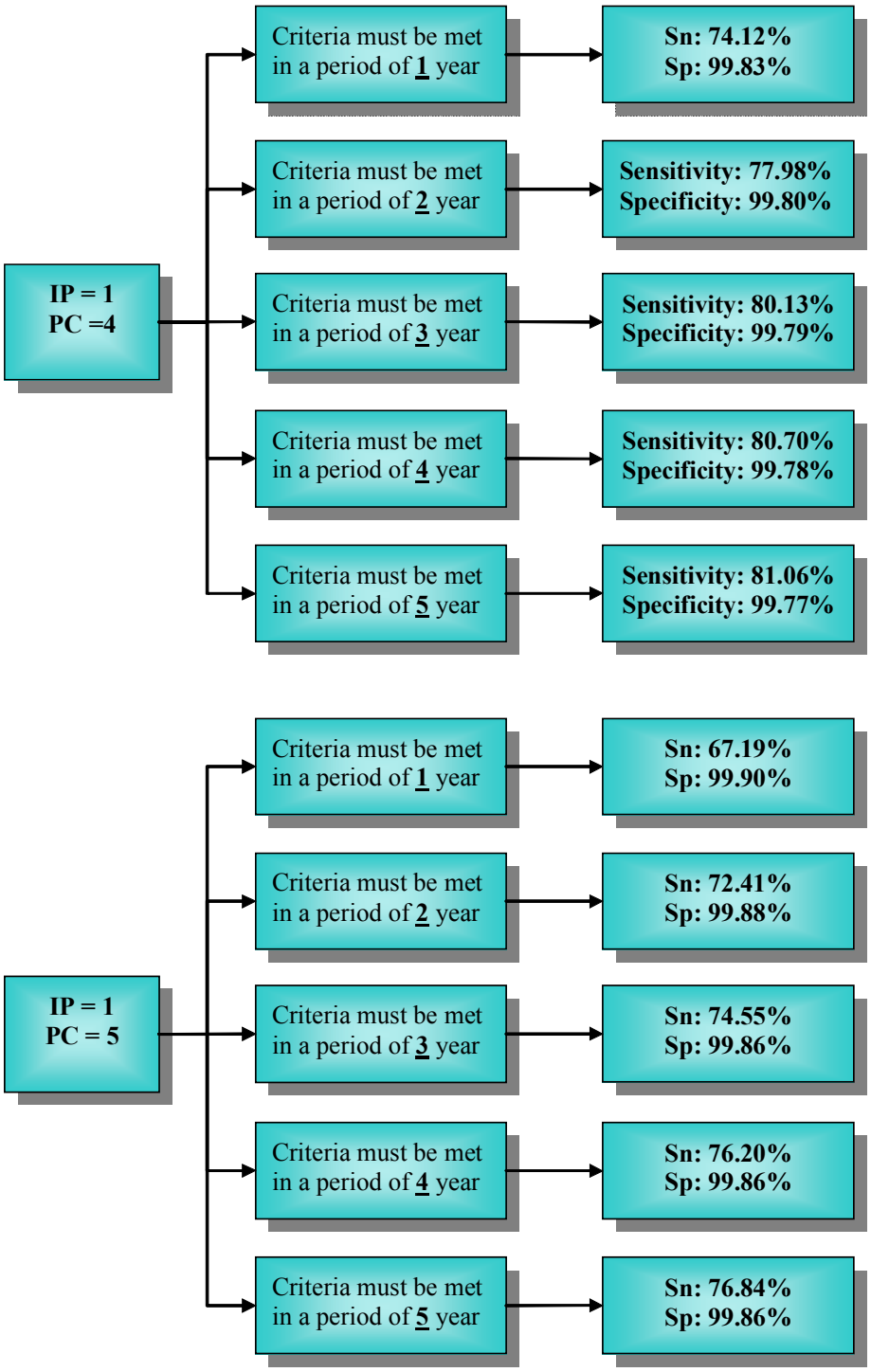

IP = 2
PC = 4
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 80.42%
Sp: 99.82%
Criteria must be met in a period of 2 year
Sn: 83.42%
Sp: 99.80%
Criteria must be met in a period of 3 year
Sn: 84.78%
Sp: 99.79%
Criteria must be met in a period of 4 year
Sn: 85.28%
Sp: 99.78%
Criteria must be met in a period of 5 year
Sn: 85.63%
Sp: 99.77%
IP = 2
PC = 4
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 76.48%
Sp: 99.84%
Criteria must be met in a period of 2 year
Sn: 80.49%
Sp: 99.82%
Criteria must be met in a period of 3 year
Sn: 82.42%
Sp: 99.81%
Criteria must be met in a period of 4 year
Sn: 82.99%
Sp: 99.80%
Criteria must be met in a period of 5 year
Sn: 83.34%
Sp: 99.79%

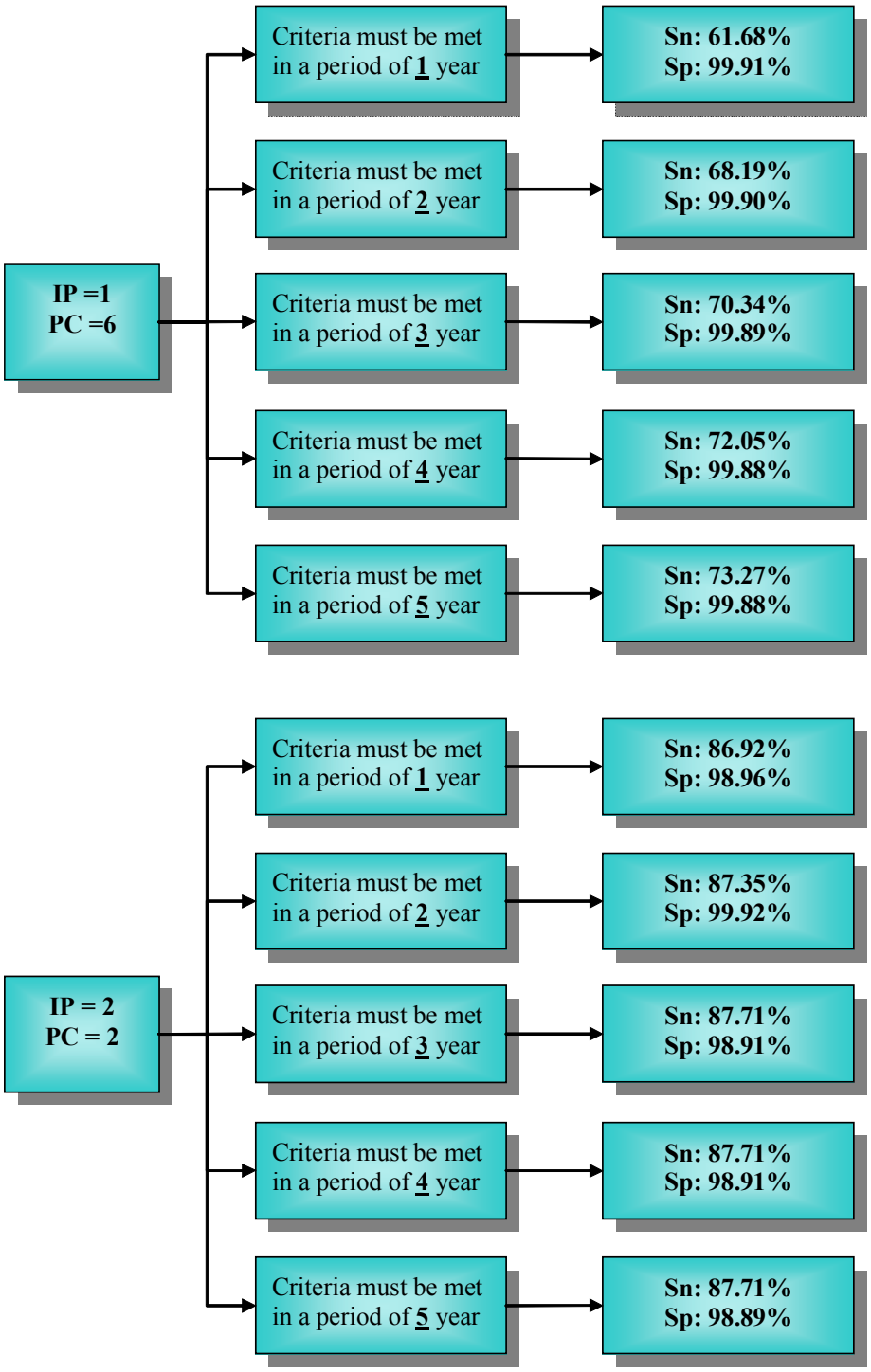

IP = 2
PC = 5
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 75.56%
Sp: 99.88%
Criteria must be met in a period of 2 year
Sn: 78.91%
Sp: 99.88%
Criteria must be met in a period of 3 year
Sn: 80.56%
Sp: 99.87%
Criteria must be met in a period of 4 year
Sn: 81.99%
Sp: 99.87%
Criteria must be met in a period of 5 year
Sn: 82.56%
Sp: 99.87%
IP = 2
PC = 5
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 70.12%
Sp: 99.91%
Criteria must be met in a period of 2 year
Sn: 74.77%
Sp: 99.90%
Criteria must be met in a period of 3 year
Sn: 77.06%
Sp: 99.90%
Criteria must be met in a period of 4 year
Sn: 78.84%
Sp: 99.90%
Criteria must be met in a period of 5 year
Sn: 79.34%
Sp: 99.90%

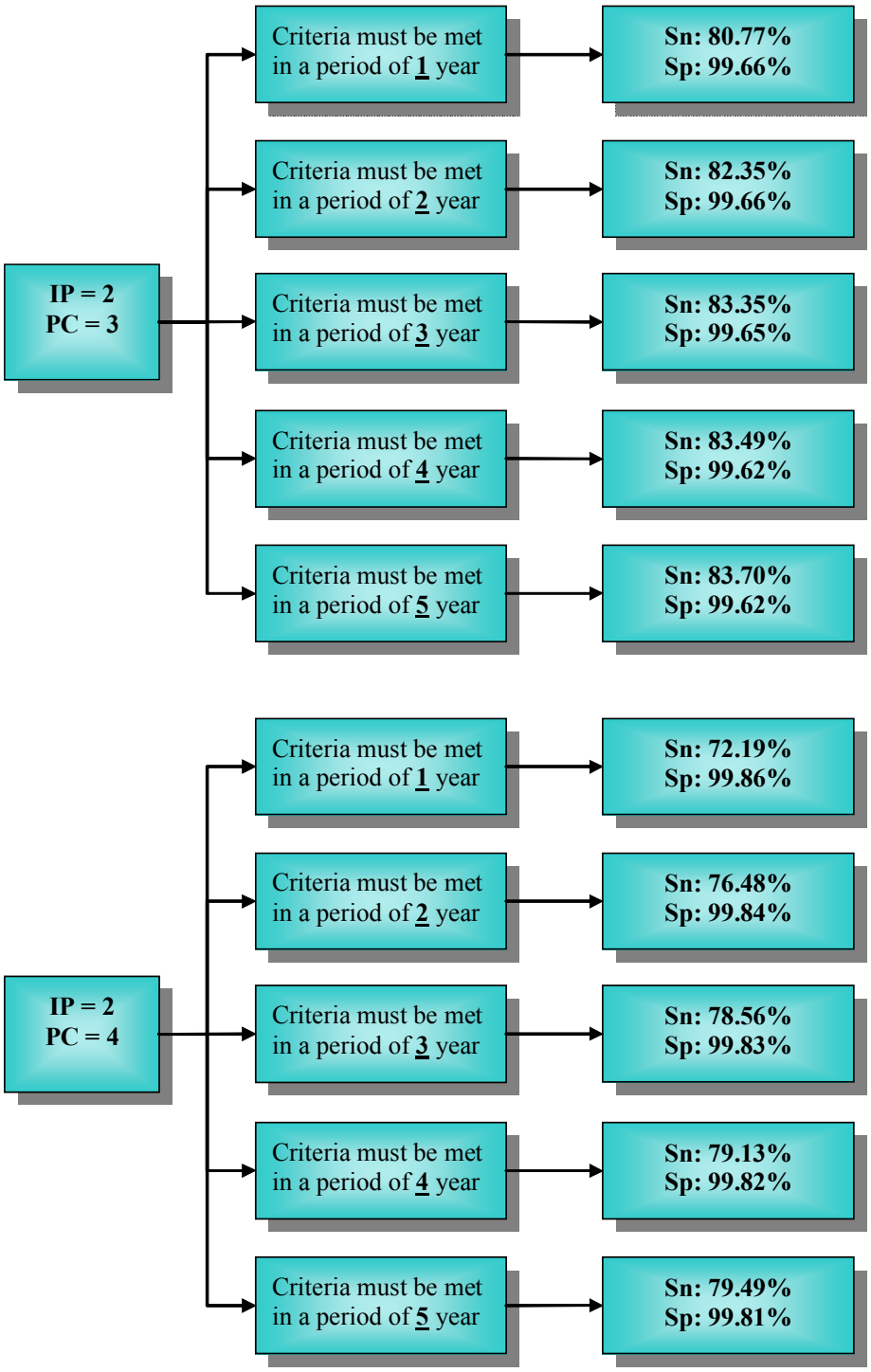

IP = 2
PC = 6
ACCS = 2
Criteria must be met in a period of 1 year
Sn: 71.91%
Sp: 99.90%
Criteria must be met in a period of 2 year
Sn: 75.91%
Sp: 99.90%
Criteria must be met in a period of 3 year
Sn: 77.41%
Sp: 99.90%
Criteria must be met in a period of 4 year
Sn: 78.41%
Sp: 99.90%
Criteria must be met in a period of 5 year
Sn: 79.56%
Sp: 99.89%
IP = 2
PC = 6
ACCS = 3
Criteria must be met in a period of 1 year
Sn: 65.55%
Sp: 99.92%
Criteria must be met in a period of 2 year
Sn: 70.84%
Sp: 99.92%
Criteria must be met in a period of 3 year
Sn: 72.98%
Sp: 99.92%
Criteria must be met in a period of 4 year
Sn: 74.55%
Sp: 99.92%
Criteria must be met in a period of 5 year
Sn: 75.77%
Sp: 99.92%

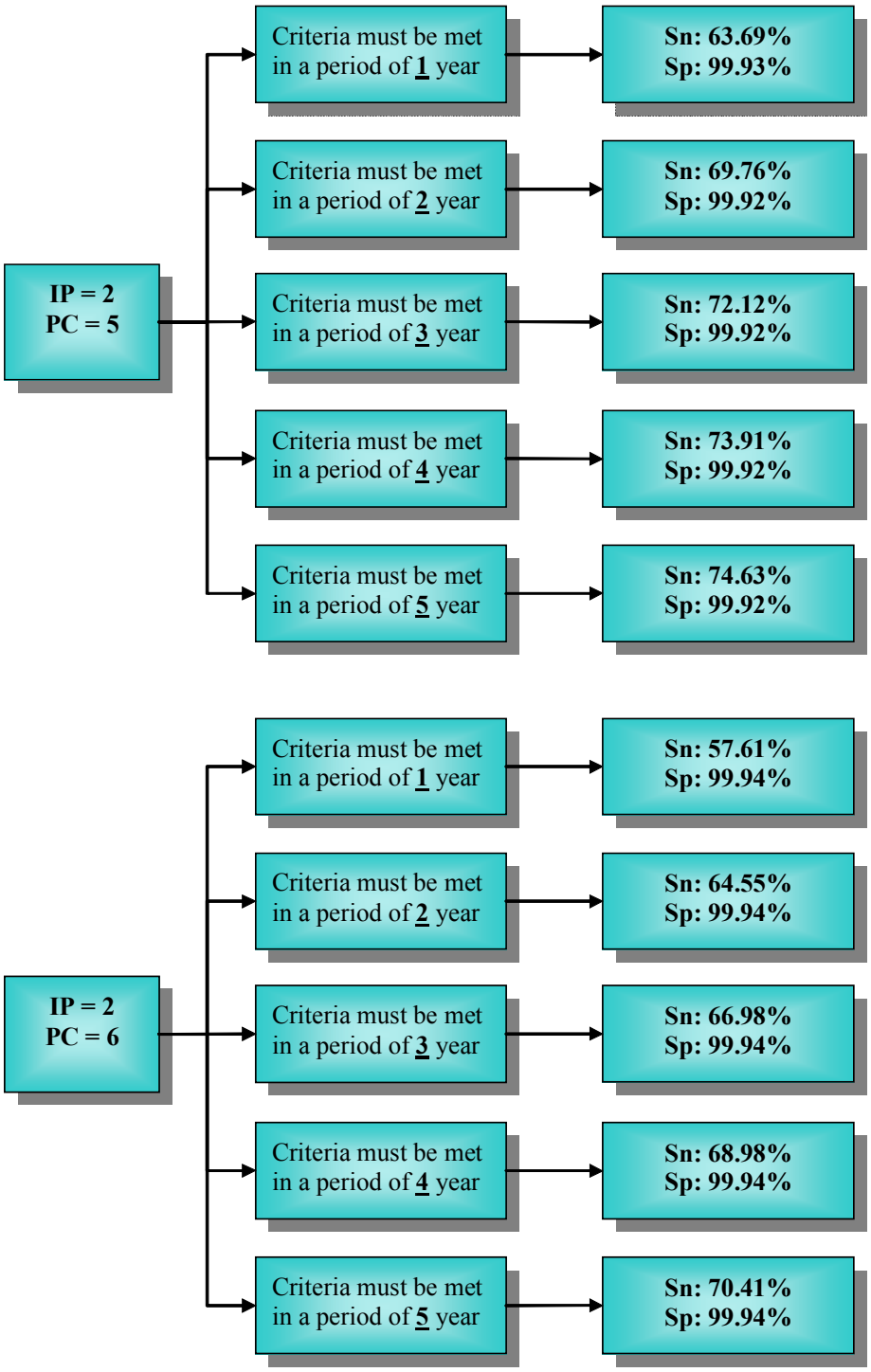
APPENDIX 10. SENSITIVITIES AND SPECIFICITIES FOR 50 DEFINITIONS THAT DID NOT INCLUDE ACCS DATABASE (FINAL CLASSIFICATION)




IP = 1
PC =4
Criteria must be met in a period of 1 year
Sn: 74.12%
Sp: 99.83%
Criteria must be met in a period of 2 year
Sensitivity: 77.98%
Specificity: 99.80%
Criteria must be met in a period of 3 year
Sensitivity: 80.13%
Specificity: 99.79%
Criteria must be met in a period of 4 year
Sensitivity: 80.70%
Specificity: 99.78%
Criteria must be met in a period of 5 year
Sensitivity: 81.06%
Specificity: 99.77%
IP = 1
PC = 5
Criteria must be met in a period of 1 year
Sn: 67.19%
Sp: 99.90%
Criteria must be met in a period of 2 year
Sn: 72.41%
Sp: 99.88%
Criteria must be met in a period of 3 year
Sn: 74.55%
Sp: 99.86%
Criteria must be met in a period of 4 year
Sn: 76.20%
Sp: 99.86%
Criteria must be met in a period of 5 year
Sn: 76.84%
Sp: 99.86%


IP =1
PC =6
Criteria must be met in a period of 1 year
Sn: 61.68%
Sp: 99.91%
Criteria must be met in a period of 2 year
Sn: 68.19%
Sp: 99.90%
Criteria must be met in a period of 3 year
Sn: 70.34%
Sp: 99.89%
Criteria must be met in a period of 4 year
Sn: 72.05%
Sp: 99.88%
Criteria must be met in a period of 5 year
Sn: 73.27%
Sp: 99.88%
IP = 2
PC = 2
Criteria must be met in a period of 1 year
Sn: 86.92%
Sp: 98.96%
Criteria must be met in a period of 2 year
Sn: 87.35%
Sp: 99.92%
Criteria must be met in a period of 3 year
Sn: 87.71%
Sp: 98.91%
Criteria must be met in a period of 4 year
Sn: 87.71%
Sp: 98.91%
Criteria must be met in a period of 5 year
Sn: 87.71%
Sp: 98.89%


IP = 2
PC = 3
Criteria must be met in a period of 1 year
Sn: 80.77%
Sp: 99.66%
Criteria must be met in a period of 2 year
Sn: 82.35%
Sp: 99.66%
Criteria must be met in a period of 3 year
Sn: 83.35%
Sp: 99.65%
Criteria must be met in a period of 4 year
Sn: 83.49%
Sp: 99.62%
Criteria must be met in a period of 5 year
Sn: 83.70%
Sp: 99.62%
IP = 2
PC = 4
Criteria must be met in a period of 1 year
Sn: 72.19%
Sp: 99.86%
Criteria must be met in a period of 2 year
Sn: 76.48%
Sp: 99.84%
Criteria must be met in a period of 3 year
Sn: 78.56%
Sp: 99.83%
Criteria must be met in a period of 4 year
Sn: 79.13%
Sp: 99.82%
Criteria must be met in a period of 5 year
Sn: 79.49%
Sp: 99.81%


IP = 2
PC = 5
Criteria must be met in a period of 1 year
Sn: 63.69%
Sp: 99.93%
Criteria must be met in a period of 2 year
Sn: 69.76%
Sp: 99.92%
Criteria must be met in a period of 3 year
Sn: 72.12%
Sp: 99.92%
Criteria must be met in a period of 4 year
Sn: 73.91%
Sp: 99.92%
Criteria must be met in a period of 5 year
Sn: 74.63%
Sp: 99.92%
IP = 2
PC = 6
Criteria must be met in a period of 1 year
Sn: 57.61%
Sp: 99.94%
Criteria must be met in a period of 2 year
Sn: 64.55%
Sp: 99.94%
Criteria must be met in a period of 3 year
Sn: 66.98%
Sp: 99.94%
Criteria must be met in a period of 4 year
Sn: 68.98%
Sp: 99.94%
Criteria must be met in a period of 5 year
Sn: 70.41%
Sp: 99.94%

APPENDIX 11. 2×2 TABLES AND TEST CHARACTERISTICS FOR PREFERABLE ADMINISTRATIVE DEFINITIONS OUT OF 150 DEFINITIONS.

A. Definitions considering ACCS database

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,243	156
	Non-IBD	65	15,374

Sn: 88.85% (95%CI=87.08,90.45)

Sp: 99.58% (95% CI=99.46,99.68)

PPV: 95.03%(95%CI=93.71,96.14)

NPV:99.00%(95%CI=98.83,99.15)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test))	IBD	1,253	146
	Non-IBD	67	15,372

Sn: 89.56% (95%CI=87.84,91.12)

Sp: 99.57% (95% CI=99.45,99.66)

PPV: 94.92%(95%CI=93.60,96.05)

NPV:99.06%(95%CI=98.89,99.21)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,256	143
	Non-IBD	70	15,369

Sn: 89.78% (95%CI=88.07,91.32)

Sp: 99.55% (95% CI=99.43,99.65)

PPV: 94.72%(95%CI=93.38,95.86)

NPV:99.08%(95%CI=98.92,99.22)

- Definition "IP≥1 *or* PC≥3 *or* ACCS≥2 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,258	141
	Non-IBD	70	15,369

Sn: 89.92% (95%CI=88.22,91.45)

Sp: 99.55% (95% CI=99.43,99.65)

PPV: 94.73%(95%CI=93.39,95.87)

NPV:99.09%(95%CI=98.93,99.23)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,182	217
	Non-IBD	37	15,402

Sn: 84.49% (95%CI=82.48,86.35)

Sp: 99.76% (95% CI=99.67,99.83)

PPV: 96.97%(95%CI=95.84,97.85)

NPV:98.61%(95%CI=98.42,98.79)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,199	200
	Non-IBD	39	15,400

Sn: 85.70% (95%CI=83.76,87.50)

Sp: 99.75% (95% CI=99.66,99.82)

PPV: 96.85%(95%CI=95.72,97.75)

NPV:98.72%(95%CI=98.53,98.89)

Positive Likelihood Ratio: 339 (95%CI = 248, 465)

Negative Likelihood Ratio: 0.14 (95%CI = 0.13, 0.16)

Kappa statistics= 0.90

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,206	193
	Non-IBD	40	15,399

Sn: 86.20% (95%CI=84.29,87.97)

Sp: 99.74% (95% CI=99.65,99.82)

PPV: 96.79%(95%CI=95.65,97.70)

NPV:98.76%(95%CI=98.58,98.93)

- Definition "IP≥1 *or* PC≥4 *or* ACCS≥2 within 5 years"

		Administrative definition IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,211	188
	Non-IBD	41	15,398

Sn: 86.56% (95%CI=84.66,88.31)

Sp: 99.73% (95% CI=99.64,99.81)

PPV: 96.73%(95%CI=95.58,97.64)

NPV:98.79%(95%CI=98.61,98.96)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 2 years"

		Administrative definition IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,131	268
	Non-IBD	26	15,413

Sn: 80.84% (95%CI=78.68,82.88)

Sp: 99.83% (95% CI=99.75,99.89)

PPV: 97.75%(95%CI=96.73,98.53)

NPV:98.29%(95%CI=98.08,98.49)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 3 years"

		Administrative definition IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,149	250
	Non-IBD	28	15,411

Sn: 82.13% (95%CI=80.02,84.10)

Sp: 99.82% (95% CI=99.74,99.88)

PPV: 97.62%(95%CI=96.58,98.41)

NPV:98.40%(95%CI=98.20,98.59)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,169	230
	Non-IBD	28	15,411

Sn: 83.56% (95%CI=81.51,85.47)

Sp: 99.82% (95% CI=99.74,99.88)

PPV: 97.66%(95%CI=96.64,98.44)

NPV:98.53%(95%CI=98.33,98.71)

- Definition "IP≥1 *or* PC≥5 *or* ACCS≥2 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,176	223
	Non-IBD	28	15,411

Sn: 84.06% (95%CI=82.04,85.94)

Sp: 99.82% (95% CI=99.74,99.88)

PPV: 97.67%(95%CI=96.66,98.45)

NPV:98.57%(95%CI=98.38,98.75)

B. Definitions exclusive of ACCS database

- Definition "IP≥1 *or* PC≥3 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,168	231
	Non-IBD	59	15,380

Sn: 83.49% (95%CI=81.44,85.40)

Sp: 99.62% (95% CI=99.51,99.71)

PPV: 95.19%(95%CI=93.84,96.32)

NPV:98.52%(95%CI=98.32,98.70)

- Definition "IP≥1 *or* PC≥3 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,184	215
	Non-IBD	61	15,378

Sn: 84.63% (95%CI=82.63,86.48)

Sp: 99.61% (95% CI=99.49,99.70)

PPV: 95.10%(95%CI=93.75,96.23)

NPV:98.62%(95%CI=98.43,98.80)

- Definition "IP≥1 *or* PC≥3 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,186	213
	Non-IBD	64	15,375

Sn: 84.78% (95%CI=82.78,86.62)

Sp: 99.59% (95% CI=99.47,99.68)

PPV: 94.88%(95%CI=93.51,96.04)

NPV:98.63%(95%CI=98.44,98.81)

- Definition "IP≥1 *or* PC≥3 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,189	210
	Non-IBD	64	15,375

Sn: 84.99% (95%CI=83.01,86.82)

Sp: 99.59% (95% CI=99.47,99.68)

PPV: 94.89%(95%CI=93.52,96.05)

NPV:98.65%(95%CI=98.46,98.83)

- Definition "IP≥1 *or* PC≥4 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,091	308
	Non-IBD	31	15,408

Sn: 77.93% (95%CI=75.72,80.13)

Sp: 99.80% (95% CI=99.72,99.86)

PPV: 97.24%(95%CI=96.10,98.12)

NPV:98.04%(95%CI=97.81,98.25)

- Definition "IP≥1 *or* PC≥4 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,121	278
	Non-IBD	33	15,406

Sn: 80.13% (95%CI=77.94,82.19)

Sp: 99.79% (95% CI=99.70,99.85)

PPV:97.14%(95%CI=96.01,98.02)

NPV:98.23%(95%CI=98.01,98.43)

Positive Likelihood Ratio: 375 (95%CI = 266, 528)

Negative Likelihood Ratio: 0.20 (95%CI = 0.18, 0.22)

- Definition "IP≥1 *or* PC≥4 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,129	270
	Non-IBD	34	15,405

Sn: 80.70% (95%CI=78.53,82.74)

Sp: 99.78% (95% CI=99.69,99.85)

PPV: 97.08%(95%CI=95.94,97.97)

NPV:98.28%(95%CI=98.06,98.48)

- Definition "IP≥1 *or* PC≥4 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,134	265
	Non-IBD	35	15,404

Sn: 81.06% (95%CI=78.90,83.08)

Sp: 99.77% (95% CI=99.69,99.84)

PPV: 97.01%(95%CI=95.86,97.91)

NPV:98.31%(95%CI=98.09,98.51)

- Definition "IP≥1 *or* PC≥5 within 2 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,013	386
	Non-IBD	19	15,420

Sn: 67.19% (95%CI=64.66,69.65)

Sp: 99.90% (95% CI=99.83,99.94)

PPV: 98.33%(95%CI=97.40,99.04)

NPV:97.11%(95%CI=96.84,97.36)

- Definition "IP≥1 *or* PC≥5 within 3 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,043	356
	Non-IBD	21	15,418

Sn: 74.55% (95%CI=72.19,76.82)

Sp: 99.86% (95% CI=99.79,99.92)

PPV: 98.03%(95%CI=97.00,98.77)

NPV:97.74%(95%CI=97.50,97.97)

- Definition "IP≥1 *or* PC≥5 within 4 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,066	333
	Non-IBD	21	15,418

Sn: 76.20% (95%CI=73.88,78.41)

Sp: 99.86% (95% CI=99.79,99.92)

PPV: 98.07%(95%CI=97.06,98.80)

NPV:97.89%(95%CI=97.65,98.11)

- Definition "IP≥1 *or* PC≥5 within 5 years"

		Administrative definition	
		IBD	Non-IBD
EndoPro + chart reviews (Gold standard test)	IBD	1,075	324
	Non-IBD	21	15,418

Sn: 76.84% (95%CI=74.54,79.03)

Sp: 99.86% (95% CI=99.79,99.92)

PPV: 98.08%(95%CI=97.09,98.81)

NPV:97.94%(95%CI=97.71,98.16)