HHS Public Access

Challenges in Designing a National Surveillance Program for Inflammatory Bowel Disease in the United States

Millie D. Long, MD, MPH[1], **Susan Hutfless, PhD**[2], **Michael D. Kappelman, MD, MPH**[1], **Hamed Khalili, MD, MPH**[3], **Gil Kaplan, MD, MPH**[4], **Charles N. Bernstein, MD**[5], **Jean Frederic Colombel, MD**[6], **Lisa Herrinton, PhD**[7], **Fernando Velayos, MD, MPH**[8], **Edward V. Loftus Jr., MD**[9], **Geoffrey C. Nguyen, MD, PhD**[10], **Ashwin N. Ananthakrishnan, MD, MPH**[3], **Amnon Sonnenberg, MD, MSc**[11], **Andrew Chan, MD, MPH**[3], **Robert S. Sandler, MD, MPH**[1], **Ashish Atreja, MD, MPH**[6], **Samir A. Shah, MD**[12], **Kenneth Rothman, DMD, PhD**[13,14], **Neal S. Leleiko, MD, PhD**[12], **Renee Bright, MS**[6], **Paolo Boffetta, MD**[6], **Kelly D. Myers**[15], and **Bruce E. Sands, MD, MS**[6]

[1] University of North Carolina at Chapel Hill, Departments of Medicine and Pediatrics, Division of Gastroenterology and Hepatology

[2] Johns Hopkins University School of Medicine, Baltimore, MD

[3] Massachussetts General Hospital, Center for Crohn's and Colitis, Harvard Medical School, Boston, MA

[4] Faculty of Medicine, Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada

[5] Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada

[6] Mount Sinai School of Medicine, New York, New York

[7] Kaiser Permanente Northern California, Oakland, CA

[8] University of California, San Francisco, Center for Crohn's and Colitis, San Francisco, CA

[9] Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota

[10] Mount Sinai Hospital Centre for Inflammatory Bowel Disease, University of Toronto, Canada

[11] Portland VA Medical Center, Portland, OR

[12] Alpert Medical School of Brown University, Providence, RI

[13] Department of Epidemiology, Boston University School of Public Health

[14] Department of Epidemiology Research, Research Triangle Institute Health Solution

[15] QForma, Inc.

Abstract

Corresponding Author: Millie D. Long MD, MPH, Campus Box 7080, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7080, millie_long@med.unc.edu, Phone: 919-843-5795, Fax: 919-966-6842.

This review describes the history of US government funding for surveillance programs in IBD, provides current estimates of the incidence and prevalence of inflammatory bowel diseases (IBD) in the United States (US), and enumerates a number of challenges faced by current and future IBD surveillance programs. A rationale for expanding the focus of IBD surveillance beyond counts of incidence and prevalence, in order to provide a greater understanding of the burden of IBD, disease etiology and pathogenesis, is provided. Lessons learned from other countries are summarized, as well as potential resources that may be used to optimize a new form of IBD surveillance in the US. A consensus recommendation on the goals and available resources for a new model for disease surveillance are provided. This new model should focus upon "surveillance of the burden of disease," including 1) natural history of disease and 2) outcomes and complications of the disease and/or treatments.

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory diseases of the gastrointestinal tract. CD can occur anywhere in the gastrointestinal tract whereas UC is localized to the colon. Collectively, these diseases are known as inflammatory bowel diseases (IBD). The etiology of IBD is unknown, although it is thought to arise from a combination of factors. These etiologic factors include genetic influences, alterations in the gut microbiota, alterations in the innate and adaptive immune system and environmental exposures. Unfortunately, without further understanding of the etiology of IBD, a prevention or cure of IBD is not possible.

IBD can cause severe ongoing gastrointestinal symptoms, such as diarrhea, bleeding and abdominal pain. These symptoms can dramatically affect quality of life. Disease can be refractory to medical treatments and surgery is often needed. IBD is therefore a costly, morbid condition for which there is currently no cure. In 2008, direct treatment costs alone for patients with IBD were estimated to be greater than 6.8 billion dollars.[1] When considering indirect costs, such as work related opportunity loss, an additional estimated 5.5 billion in 2009 US dollars needs to be added to this estimate.[2] Due to this high burden of disease, legislation has been enacted within the past decade to improve research funding for these diseases and to target further understanding of IBD epidemiology and pathophysiology.

This review will describe current estimates of the incidence and prevalence of IBD in the US, discuss potentially under-counted populations, and describe the history of government funding for surveillance programs in IBD. Lessons learned from other countries on IBD surveillance will be summarized, as will potential resources that may be used to optimize IBD surveillance in the US. Finally, a consensus recommendation on the best means of optimizing public health surveillance in IBD will be offered.

Epidemiology of IBD

In the US, it is currently estimated that over 1.4 million people suffer from IBD. Estimates of disease prevalence among adults in the US are 201 cases per 100,000 persons for CD and 238 cases per 100,000 persons for UC.[3] The incidence rates in the US are approximately 8.8

cases per 100,000 person-years for CD and 7.9 cases per 100,000 person-years for UC, as estimated in the Olmsted County, Minnesota, population.[4] When compared to international rates, estimates of CD incidence are highest in North America (20.2 per 100,000 person-years); whereas the annual incidence of UC is highest in Europe (24.3 per 100,000 person-years). Europe also has the greatest prevalence of both UC and CD (505 per 100,000 and 322 per 100,000, respectively).[5] Interestingly, other areas of the world have significantly lower rates of IBD;[5] however, these rates appear to be increasing in parts of Asia and northern Africa.[6] IBD incidence is also increasing in other areas such as Australia[7] and New Zealand.[8] In these emerging areas, rising rates of UC appear before those of CD.[9] Data are not robust on IBD incidence and prevalence in under-developed countries. More accurate means of surveillance in these areas are needed.

Certain populations may also be undercounted in the surveillance of IBD in the US (Table 1). A better understanding of disease rates in subgroups of interest, such as minorities, immigrants, the elderly and children, is warranted. CD and UC incidence and prevalence are difficult to determine by race and ethnicity status. Those studies that have investigated race and ethnicity in the epidemiology of IBD have compared rates of hospitalization for disease by race, rather than incidence or prevalence as identified in inpatient and outpatient resources.[10-15] The estimates range from little difference in the rate of CD and UC between whites and African Americans[10, 13] to decreased rates of CD and UC for African Americans, Hispanics, Asians, and Native Americans/Pacific Islanders compared with whites in the same population.[11, 12, 14, 15] Rates of IBD for migrants to the US have not been reported, although there is evidence from Southeast Asian migrants to the United Kingdom that migrant populations rapidly assume the incidence rate of the underlying population or assume an even higher rate.[16] Evidence for variation in incidence rates of IBD by race among U.S. children is even more rare[17] with a similar trend in the ratio of cases of IBD lower among blacks and Hispanics compared with whites.

There are 3 studies reporting the incidence of pediatric IBD in the US, each using a different data source, but with similar overall estimates.[4, 18, 19] In general, data are from a single geographic region and must be extrapolated to other areas of the US. Data for African Americans are sparse, and none of the reports are from the southern US. In the Olmsted County population, Loftus et al estimated the incidence of pediatric CD (4.8/100,000 person-years) and UC (3.2/100,000 person-years) from 1990-2000.[4] Adamiak used data from Wisconsin, where IBD patients generally receive their care within the state, to calculate the incidence of IBD from 2000-2007, with similar estimates found for both CD and UC (6.6/100,000 person-years and 2.4/100,000 person-years, respectively).[18] Both Olmsted County and Wisconsin are predominantly Caucasian; therefore, it is unclear whether these estimates are applicable to other races/ethnicities. In the Kaiser Permanente integrated health plan, data were somewhat different, perhaps related to a different mix of race/ethnicity than other parts of the US (2.7/100,000 person-years for CD and 3.2/100,000 person-years for UC).[19] The prevalence of pediatric IBD in the US has been studied in these populations, as well as in nationwide health claims data. Prevalence estimates of pediatric IBD are relatively consistent, approximately 40/100,000 for CD and 28/100,000 for UC.[3, 19, 20] Again, data on racial minorities and certain geographic regions of the US are

limited. These limitations demonstrate a need to include data from all geographic regions and ethnicities in the US.

History of IBD Surveillance

The Crohn's& Colitis Foundation of America (CCFA) began a collaboration with the Centers for Disease Control and Prevention (CDC) in 2002 to increase the awareness of and funding for IBD epidemiologic research in the US. In 2003, the Inflammatory Bowel Disease Act was introduced. This act called on the National Institutes of Health (NIH) to expand research in the arena of IBD. Within this act was a provision for the CDC to establish a national program of prevention and epidemiology to determine the prevalence and incidence of IBD in the United States as well as to explore how practice variations among community physicians affected patient outcomes. Specific funding of 5 million dollars per cycle was provided to accomplish these goals. The components of this act were combined with other legislation in the Research Review Act of 2004. The bill was unanimously passed by the Senate on November 16. President Bush signed the Research Review Act of 2004 into law on November 30.

In its first cycle of IBD surveillance funding, CDC epidemiologists worked with CCFA and Kaiser-Permanente to better understand the epidemiology, incidence, prevalence, and outcomes of IBD.[19-23] Researchers also worked to determine factors that predicted the course of disease. In the second cycle, the funding went to a unique state-based inception cohort of IBD, the Ocean State Crohn's and Colitis Area Registry (OSCCAR).[24] How best to study IBD prevalence moving forward is of utmost importance.

Conference on IBD Surveillance

There is currently no unified surveillance system for IBD in the US. Thus far, estimates of incidence and prevalence in the US have come from regional cohorts and administrative data. In order to determine the best use of future funding for IBD surveillance, the CDC and CCFA convened a meeting of international experts in IBD epidemiology in February, 2013, in Philadelphia, PA. The group was challenged to develop and pilot the prototype for a national surveillance program for IBD. The charge was to gather national prevalence and incidence data; perhaps by utilizing an existing surveillance system.

Prior to developing a surveillance system for IBD, the group determined that the implications and goals of surveillance in IBD must be considered. Public health surveillance involves the ongoing and systematic collection, analysis and interpretation of health data and the timely dissemination of these data to those responsible for preventing and controlling the condition.[9] The specific purposes of a public health surveillance system are to: 1) formulate public health policies; 2) identify high risk populations; 3) target interventions; 4) evaluate progress in disease prevention; and 5) identify environmental risk factors contributing to disease.

In a disease group such as IBD, not all of these components of surveillance apply. For example, with an as of yet unknown specific etiology, targeting interventions and prevention is not yet possible. However, an accurate estimate of the prevalence of IBD remains

important to inform public policies, particularly in populations for whom little data are available (children, racial and ethnic minorities and immigrants). Surveillance of IBD and other disease processes can help to track health care resource needs moving forward. Clinical data beyond estimates of incidence and prevalence are needed, in order to understand the economic, population and individual-level burdens of IBD.

Challenges of IBD Surveillance in the United States

There are numerous challenges within the US that have an impact upon the creation of an IBD surveillance program. In general, unlike in other countries, there is no unique identifier with which to follow individuals through the fragmented health care system. This fragmentation of health care, with multiple separate systems of commercial and government-based health insurance, allows for the possibility of underrepresentation of certain groups in the data that are available and an inability to unify longitudinal tracking of disease course for any individual. In particular, minorities and children may not be adequately represented. Problems that are more specific to IBD include the lack of hard outcomes such as mortality, the fact that IBD is not a "reportable" condition, and the important problem that a diagnosis of IBD itself is not straightforward without an available gold standard. Laboratory testing for IBD is not considered reliable or accurate. As IBD is a collection of relatively rare entities, and there is a wide range of severity of disease, there is the potential for low levels of disease remaining unrecognized below the diagnostic threshold.

There are several potential solutions to these challenges. Importantly, established sources of data could be used to establish a surveillance infrastructure for IBD. These potential data sources include data from established population-based surveys or cohort studies, provider surveys, facility-based surveys, state-based registries, administrative health care data, electronic medical records and reporting systems, and internet-based surveys. During the course of the CCFA/CDC meeting, experts considered each of these sources as a possible solution to the problems of creating a national IBD surveillance program in the US, with emphasis upon feasibility, cost, value added and quality of the data. Experts also reviewed international surveillance systems already in place, which could provide important lessons for the development of such a system in the US.

Potential Methods of Surveillance

Lessons from elsewhere

There are a number of established registries that have been effectively used to provide insight into the epidemiology of IBD. The paragraphs that follow describe these important resources and how they been used to advance our understanding of IBD.

Scandinavian Health System—The Swedish National Board of Health and Welfare has collected individual-level data since 1964. The registry includes data on personal identification number, date of birth, dates of hospital admissions, hospital departments, and discharge diagnoses. The registry was expanded in 1997 to include data on ambulatory surgery and again in 2001 to include data on hospital-based outpatient clinic visits. In 2005,

the Swedish Prescribed Drug Register was established. The register contained data on dispensed prescriptions including such factors as the date of redemption, the amount of drug, and the dosage. Identifying data is missing in less than 0.3% of all items. Other notable Scandinavian registries include the Medical Birth Register Database, the Cause of Death Registry, and the Swedish Cancer Registry. The Medical Birth Register contains detailed information on prenatal care, delivery, and neonatal care. The Swedish Cancer Registry maintains data on sex, age, site of cancer, histology, stage of tumor, and date of death from cancer.

These data have been used in numerous studies investigating various aspects of IBD, beyond incidence and prevalence. In particular, several studies on malignancy risks over time in those with IBD have come from these data sources.[25-27] The primary advantages of these registries include the large sample size, the longitudinal data of individuals followed over a long time period, and availability of prescription, cancer, and cause of death information. The ideal types of studies to perform using these data include natural history studies and outcomes research in IBD. The primary disadvantage to these data is the lack of detailed and updated information on many lifestyle characteristics.

Scandinavian registries are ideal to evaluate the secular trends in incidence and prevalence of IBD. The presence of a unique identification number for every individual living in Sweden allows various registries to be merged and used for future research. Unfortunately, in the US it is impossible to implement such a system due to the lack of unique personal identifiers within our fragmented health care system.

Calgary, Canada—The Calgary Health Zone is a population-based health authority under a public, single-payer system, and provides all levels of medical and surgical care to the residents of the city of Calgary and over 20 nearby smaller cities, towns, villages, and hamlets. The estimated population of the Calgary Health Zone is over 2 million people. Administrative data are collected within the province. As these data are identifiable, they can be linked to other databases or to the electronic medical record to validate disease exposures and outcomes. The Data Integration, Measurement, and Reporting (DIMR) hospital discharge abstract administrative database captures all hospitalizations in the Calgary Health Zone of Alberta Health Services, Canada. The DIMR database contains 42 diagnostic and 25 procedural coding fields. The International Classification of Disease, Ninth Revision (ICD-9) was used up to 2001, whereas the ICD-10-CA (Canadian adaption of ICD-10) and the Canadian Classification of Health Intervention coding were used after 2001. Within the province, prevalence and inception cohorts were developed, with age- and sex-matched controls. Validation studies were performed (via chart review) to confirm cases and exposures. Validation studies are essential to evaluate biases in an administrative database. For example, in a recent validation study from within this cohort, administrative data identified the same risk factors as chart review, but overestimated the magnitude of risk.[28]

This data source has been used in studies of IBD, including outcomes, risk factor analysis, surveillance of comorbidities or complications, tracking of infections, health services utilization and health economics/costs.[29-32] For example, the incidence of IBD has been

reported to be 25/100,000 person-years from 2003-2011 in this region (equating to 800 new cases per year). Data from this cohort have also shown that risk factors for postoperative mortality in IBD patients include age and disease severity.[33] The advantages to the use of administrative data as an initial data source include the large sample size, the ability to completely capture any billed aspect of disease management (such as surgery) and detailed information on costs. The disadvantages include the need for validation of individual level data via chart review and the costs and time involved to do so. Detailed clinical information, such as phenotype, disease activity, and biological samples, are not available within administrative data, unless linked to other repositories. Unfortunately, due to the inability to gain access to identifiable data within conglomerate administrative data in the US, this type of prospective inception/prevalence cohort is not possible. Median time on a given commercial health plan in the US is only a few years. As we do not have the ability to track an individual across plans, long-term outcomes data would not be available.

Manitoba, Canada—The University of Manitoba IBD Epidemiology Database (UMIBDED) and the University of Manitoba IBD Research Registry (UMIBDRR) were established in 1995 in the central Canadian province of Manitoba (population of 1.27 million in 2012). Manitoba Health is the single health insurance provider in the region. Each resident of Manitoba has a personal health identification number (PHIN) dating back to 1984 through which every health system encounter is logged with a diagnostic code (ICD-9 for all outpatient encounters and inpatient encounters until 2003 and ICD-10 for inpatient encounters starting in 2004), a tariff that identifies what service was provided, and a provider ID number.

A population-based database of all Manitobans was established applying a validated (by self-report and chart review) administrative definition of IBD to the Manitoba Health database. A case is identified in the Manitoba Health database as being an IBD case if there are at least 5 distinct encounters with a code for IBD. Alternately, if the person is in the database for less than 2 years, then at least 3 distinct encounters are required. This database is updated regularly to provide contemporary data on all Manitobans with IBD. A matched cohort drawn from the Manitoba Health database was also established during the same timeframe. The matching was 10:1 on sex, age and geographic location (by postal code). In 1995 the Drug Program Information Network was established which identifies all outpatient prescriptions for all residents and this can be linked to the UMIBDED by the PHIN. The UMIBDED is anonymous and scrubbed of any subject identifiers. Factors that can be identified include gender, age, disease duration, disease diagnosis (CD versus UC), comorbidities, health care utilization, hospitalizations, surgeries, endoscopies, imaging procedures and area of residence. Factors that cannot be identified include disease phenotype and disease activity.

This database has been used to identify epidemiology outcomes, disease outcomes, comorbidities, health care utilization, and direct health care costs. At the time the database was developed, the UMIBDRR was also formed by enrolling persons who responded to an initial mailing from their doctors' offices. This Research Registry has been intermittently updated since 1995 through mailing to recently diagnosed subjects identified as such by the administrative definition applied to the Manitoba Health database. This Registry has been

used to enroll patients in several studies including the Manitoba IBD Risk Factor Study and the Manitoba IBD Cohort Study among others. The UMIBDRR is a population-based cohort that can be accessed for studies where personal data and disease characteristics can be more fully explored. Further interventions can be undertaken on these subjects.

Numerous publications have arisen from this population-based cohort, including studies of risk factors for disease, malignancy, disease complications and disease outcomes.[34-43] The UMIBDED is an administrative database and subject access or intervention is not possible. The UMIBDED is ideal for surveillance of incidence and prevalence data, burden of comorbidities and health care utilization, as well as pharmacoepidemiology studies. Its main limitation is the lack of patient-level data such as phenotype, disease activity and response to medications. The personal health identification number to link individuals through various aspects of the health care system and other registries makes this type of cohort possible, and is currently not available in the US. Additionally, the region has one insurer, allowing for greater confidence in complete capture of incident cases, which is not the case in the US, where there are many insurance providers and individuals without any form of insurance coverage.

Northern France—Unlike some other European countries, there is no national disease registration system in France. In 1988, the Epidémiologie des maladies inflammatoires chroniques de l'Intestin en France (EPIMAD) registry was established as a means of tracking incident cases of IBD, risk factors and outcomes. The registry covers an area in Northern France with over 6 million inhabitants (9.3% of the French population). This cohort study uses interviewer practitioners in the offices of 262 adult and pediatric gastroenterologists who work in private practice ($n = 168$), general practice ($n = 78$) or university hospitals ($n = 16$) within the catchment area. Only patients who are residents in the defined study area at the time of diagnosis are included. Each gastroenterologist reports all patients undergoing first consultations for symptoms compatible with IBD. For each new case, the interviewer practitioners collect all available information from the chart in a standardized format. These data include such factors as age, gender, year of diagnosis, time between onset of symptoms and diagnosis, as well as clinical, radiological, endoscopic, and histological findings at the time of diagnosis. Final diagnosis of IBD is then made by two expert gastroenterologists and recorded as definite, probable, or possible, according to the validated criteria.

This database has been used in many studies of incidence, risk factors, disease complications, disease phenotype and progression, natural history of disease, and genetic studies.[44-51] Important findings have arisen from this cohort, such as the description of a generally mild disease course in elderly-onset IBD.[45] The strengths of this cohort include the population-based nature of the cohort, the prospective data collection, the sample size (now with over 15,900 cases of confirmed IBD), and the detailed level of clinical information (such as phenotype and medication use). This cohort can also be used for recruitment into interventional studies as well. The disadvantages of this registry include the limited time frame for assessment (registry initiated in 1988), no prospective data collection following the diagnosis, the fact that detection of incident cases depends on a good collaboration with private gastroenterologists, labor intensive case detection and diagnosis confirmation, and the costs are quite high (annual costs of approximately 300,000 euros).

This type of incident disease cohort has been established in the US (OSCCAR) in Rhode Island over the past 5 years. The limitations to using this type of cohort for US epidemiologic surveillance include the small sample size, the limited geographic coverage (therefore limiting inclusion of underrepresented racial and ethnic minorities) and the costs involved in development and maintenance of the cohort.

US Experience

Administrative Data: Commercial Health Plans—Administrative claims data from within the US can be leveraged for robust epidemiological research. These data consist of the computerized encounter data submitted by providers (physicians, hospitals, nursing homes, home health agencies and pharmacies) to the payer. Data users can then use these de-identified claims to perform research. There are numerous commercially available data sources within the US, such as data from a single payer (such as United Healthcare (Ingenix)), or pooled health plan data (such as IMS Health Lifelink, MarketScan or Healthcore). There are also data from government insurers, namely Medicare and Medicaid.

There are several important advantages to these data. First, these data have great power and precision related to the extremely large sizes of the plans. The data are longitudinal, allowing individual to be followed over time, as long as they are insured by the same entity. The data are relatively complete and encompass care provided to individuals across all settings, provided that claims were submitted for payment. These data therefore represent "real world settings."

There are also many potential pitfalls, or disadvantages, to using these data. Most importantly, these data are not collected for the purpose of research (or for clinical care). Therefore, administrative data lack clinical details not captured by procedure or diagnosis codes. In the case of IBD, administrative data do not capture phenotype, extent, symptoms, or disease severity. Other important clinical characteristics are also missing, such as body mass index, race/ethnicity or smoking status. Validation of exposures and outcomes is not always possible. While individuals can be followed longitudinally, a limitation to the use of these data is that the median time period for health plan enrollment in the US is relatively short. For example, the median duration of enrollment in IMS Health data is 2.5 years per individual. Therefore, studies of long-term disease/treatment outcomes (such as malignancy) can be limited. Within these data, it is also not always clear whether an individual died, or left the health plan, it is only apparent that they are no longer within the dataset (which limits studies of mortality).

Health plan data are excellent for estimating the prevalence of disease. Various definitions of IBD within administrative data have been used across these studies. In a northern California population, Liu et al found a definition consisting of any 2 separate visits associated with an ICD-9 code for IBD to have a positive predictive value (PPV) of 95% when compared to chart review.[52] Herrinton et al investigated a random sample of 1.8 million members of 9 integrated healthcare organizations using several different combinations of international classification of diseases 9th edition (ICD-9) codes and medications to identify patients with IBD.[22] The final algorithm included at least one ICD-9 for CD or UC and at least one dispensing of a 5-aminosalicylate (5-ASA) medication. The

prevalence per 100,000 in 1999-2001 was 129 for CD, 191 for UC and 388 for IBD overall. There was a threefold variation in prevalence by plan.[22] Prevalence will vary based on characteristics used to define IBD, including the run-in period. There are differences in the Kaiser insured population (all-inclusive closed plan) as compared to other US insured populations, therefore limiting the generalizability of these definitions across all US commercial health plan data.

An updated prevalence study by Kappelman et al used data from Pharmetrics Choice Database (now IMS Health), which covered over 44 million lives. A somewhat more specific algorithm was used, requiring at least 3 health care contacts with a diagnosis code of IBD or 1 ICD-9 code for IBD in combination with 1 IBD-specific medication. Prevalence was standardized to the age, gender and regional distribution of the 2009 US census. Kappelman et al found a prevalence of pediatric and adult CD to be 58 and 241 per 100,000, respectively. Pediatric and adult UC were relatively similar, with somewhat lower pediatric rates and higher adult rates, as expected (34 and 263 per 100,000, respectively). In the various studies of IBD prevalence in the US since 2000, overall reported prevalence rates have been in a similar range (96-190 per 100,000 for CD and 156-214 per 100,000 for UC). Determination of incidence within health plan data is more difficult, particularly related to the low overall median health plan duration of coverage (2.5 years). Incidence data cannot be generated with such short time spans. As an extended period with no prior claims for IBD is needed to establish truly incident disease (rather than prevalent disease with low utilization), a limitation to the power of these datasets is the need to select only those individuals with longer periods of coverage. Kappelman et al limited their analysis to those with longer coverage time and found an incidence of 20 per 100,000 person-years for CD and 34 per 100,000 person-years for UC (unpublished data). These rates are similar to those reported by Loftus et al in their population-based cohort in Olmsted County from 2001 (31-71 per 100,000 person-years for CD and 18-31 per 100,000 person-years for UC).[4] Therefore, it may be possible to use administrative claims data to establish incidence with the appropriate methodological considerations. In summary, de-identified health plan data are readily available, reasonably current (12 month delay), and are fairly inexpensive to access and to work with. With these data, misclassification will happen, and definitions need to take into account the likely direction of the bias. Algorithms for definition of incident IBD could potentially be informed and calibrated by data from other sources.

These types of data could be leveraged as a component of a surveillance system for IBD. A surveillance system could partner with selected health plans to achieve the desired population sampling. In essence, this could become "real-time" surveillance, beginning with the first claim. The infrastructure and costs to support such a system would be quite high, and it may not be plausible given the need for chart review to validate cases. Therefore, the use of administrative data would likely need to be combined with other approaches to reach the goals of surveillance in IBD.

Administrative Data: Medicare and Medicaid—Medicare and Medicaid are organized broadly under the Center for Medicaid and Medicare Services. Medicare is organized under the federal government and Medicaid is organized individually under the state governments

within federal guidelines and oversight. Therefore, these data do represent a central repository of national data. However, these data only cover a segment of the population.

In 2000 there were 39.7 million people enrolled in Medicare and this increased to 50.7 million in 2012. Criteria for coverage under Medicare include those 65 years of age, those under 65 with certain disabilities, and all those with end-stage renal disease. The selected breakdown within the 2012 data includes 8.8 million disabled under the age of 65, 17 million age 75, 5.1 million African American, and 3.4 million of other race/ethnicity. Medicare is divided into 4 components. Part A is hospital insurance, Part B is outpatient insurance, Part C is Medicare Advantage, with services through provider organization, and Part D is prescription coverage, which started in 2006.[53]

Medicaid is a highly utilized program, and is steadily growing. For example, in 2000, there were 34.5 million enrollees and in 2012 there were 56.6 million enrollees. After the implementation in 2014 of the Patient Protection and Affordable Care Act, an additional 17 million enrollees are expected. Overall, Medicaid enrollees represent 22% of the US population.[53] In regards to Medicaid eligibility, all groups include low income as a criterion, and other qualifications include disability, pregnancy, children, and elderly individuals. In regards to race and ethnicity, there were 13.6 million African Americans, 15.5 million Hispanic and 1.9 million Asian enrollees. Each state determines the type, duration and scope of services it will provide, within broad federal guidelines. Delivery of care occurs in 2 separate models, including fee for service and also via managed care organizations.

Medicare and Medicaid data are used for a broad range of projects. One important aspect of these data is the potential to link them to other registries to enrich the data. For example, these data can be linked to Surveillance and End Epidemiology Results (SEER) to perform studies of cancer outcomes. Shaukat et al linked a SEER catchment area to Medicare claims data Parts A and B. The authors reported a 93% success rate for linking. The authors also investigated IBD as a risk factor for colorectal cancer in older individuals[54] and risks of small bowel adenocarcinoma among older individuals with CD.[55] Other linkage studies did not focus upon IBD, but were able to link Medicare or Medicaid data to other US government registries or disease-based registries.[56-58]

It is also possible to combine Medicare and Medicaid data with other data sources to enrich data or augment the power. For example, one group combined 4 data sources, Medicare/ Medicaid databases, Tennessee Medicaid, pharmacy benefits for Medicare in certain states, and Kaiser Permanente data, to investigate the cancer risks associated with anti-tumor necrosis factor alpha (anti-TNF) agents in chronic immune-mediated diseases. The authors used ICD-9 codes for identification and a variety of methods in combining data regarding cancers and also regarding infectious complications.[59, 60] There are also state-based data that can be leveraged to perform research. For example, in California, there is a research institute (California Medicaid Research Institute, CAMRI) within the MediCal program that focuses upon research relevant to health policy and health care decision makers. As of yet, state programs such as this do not focus upon surveillance or disease specific states. IBD researchers have used state government Medicaid repositories to investigate the incidence and risk of intestinal and extraintestinal complications over a 5-year period in IBD.[61] ICD-9

codes were used for identification of patients, requiring 2 claims for IBD, similar to prior studies using administrative data.

These data currently have an undefined role in surveillance. However, these data could form an aspect of a broad approach to surveillance. CMS data are the closest thing we currently have to a "national" population-based database. As data are in place, the costs may be less than those associated with initiating a surveillance program from scratch. Additionally, these data include historically underrepresented populations, including vulnerable populations such as children. The limitations include those of administrative data, without the ability to validate data unless linked to another data source. Clinical characteristics such as disease phenotype are not available. The population is limited, and it is not representative of the US population at large. These data have a high prevalence of elderly individuals and low-income individuals. If these data are used in combination with other commercial administrative claims data that include Medicare or Medicaid managed care programs, it is possible that individuals would be double counted or miscounted. In all, these data could not stand alone as a surveillance program for IBD, but could play a role in such a program.

Closed Panel HMO data (Kaiser-Permanente)—Kaiser Permanente Northern California (KPNC) is a closed, staff model, integrated health care delivery system with capitated payment founded in 1945 serving Sacramento and the Bay Area. Today, it is one of the largest health care delivery organizations in the United States. KPNC provides health care services to approximately 3.3 million members, about 30% of the population in 14 Northern California counties. Based on data from the 2003 California Health Interview Survey, the KPNC adult membership is very similar to the non-KPNC population with regards to education and health, differing by having lower percentages of White non-Hispanics and lower percentages of members with very low and very high household incomes. The KPNC membership is similar to the general population with regards to race/ethnicity, but appears to be better educated, in better overall health (as perceived by survey participants), and have lower rates of smoking.

The KPNC Division of Research (DOR) was established in 1961 with a mission to understand the determinants of illness and well-being and to improve the quality and cost effectiveness of health care for Kaiser Permanente members and society at large. DOR conducts epidemiological and health services research, including technology assessment, and serves as the principal consulting resource for health services, health disparities and clinical studies for KPNC. The DOR has validated algorithms for IBD case finding within their computerized clinical data. These definitions were previously validated against chart review. Cases of IBD have been identified and followed, allowing for publication of studies on natural history, incidence and prevalence, disease complications and outcomes, medication safety, as well as mortality.[20, 22, 23, 62-65]

Large computerized data systems are essential for studying rare diseases such as IBD and they represent real-world experience. Kaiser Permanente's population is large, diverse, and generally representative of urban Northern California. Many of the variables used in the IBD research program have been validated. In addition, the automated healthcare data used for this study have relatively long follow-up, allowing greater capture of the patient's history.

The setting and resources are excellent for assessing incidence and prevalence, and for studying health care processes and outcomes. There are also limitations to this data source. These important limitations include incomplete information on the patient's clinical history; interval censoring of patients, with patients cycling into and out of enrollment; and the use of algorithms with levels of validity that are not adequate to meet a study's aims. In addition, information on the incidence date and disease severity and activity are not recorded as structured data in the computerized system, although these data can be obtained from interpretation of text in progress notes.

Kaiser Permanente's computerized data has an important role to play in understanding the incidence and prevalence of IBD, including patterns in relation to sex, age, race/ethnicity, and calendar change over time. However, in regards to a national IBD surveillance program, this model is not feasible as the majority of US health plans are not closed, capitated health plans as in the Kaiser Permanente model. Other forms of administrative data within the US are not identifiable, and therefore cannot be linked to medical records for validation. A proposed national IBD surveillance program may include a component of administrative data, such as that within Kaiser Permanente.

Regional Longitudinal Cohort (Olmsted County)—The Rochester Epidemiology Project (REP) is a medical records linkage system that has existed since 1966 and captures healthcare information on all residents of Olmsted County, Minnesota.[66-69] The project provides computer linkage for medical diagnoses and surgical procedures (via ICD-9 and other codes) for all healthcare encounters including hospitalizations, emergency room visits, outpatient visits and nursing home care. The project has access to all records, including ambulatory care, within the county. The REP exploits the fact that virtually all healthcare in the county is provided by one of two integrated healthcare systems. The project now includes over 500,000 individual patients (1966-2010), with over 6 million years of follow-up. These data identify the full spectrum of disease. There are over 1800 epidemiologic reports on acute and chronic disease. The schema for development of an individual study with the REP includes retrieval of data using codes and dates, medical record review, and then performance of the study within these validated, abstracted data. This methodology has been used to develop an inception cohort of IBD.[4, 70, 71]

Case definitions of IBD within this data source can include clinical characteristics, due to the highly reliable level of clinical data available. For example, the case definition for CD includes a finding from at least 2 of the following categories: specific symptoms, endoscopic findings, radiologic findings, surgical findings, or pathologic findings. For UC the definition includes 2 studies separated by 6 months including endoscopy findings of diffuse granular or friable mucosa with continuous involvement on endoscopy or barium study. Examples of IBD studies within this data source have included several incidence and prevalence studies. Loftus et al showed incidence rates of both CD and UC in Olmsted County from 1940-2000 to peak in the 20-29 year old age group.[4] Temporal trends showed an increase in incidence of IBD over time in this cohort, particularly for CD. Other factors addressed in studies from within this cohort include costs,[72] malignancy risks,[73] fistula risks,[74] bone fracture risks,[75, 76] responses to medical therapies (such as corticosteroids)[77], progression of intestinal complications,[78] requirement for surgeries,[79] and overall survival.[80] In a survival

study, Jess et al demonstrated that Olmsted County residents with UC had similar survival to expected rates, actually trending towards better than expected. There were significantly fewer cardiovascular deaths, perhaps related to smoking status.[80] In CD, Olmsted County residents also had similar survival to expected rates, with a trend towards worsening survival. Approximately a 1/3 of deaths are related to GI causes.

The data within the REP have been shown to be representative of Minnesota and the upper Midwestern US. However, these data are not representative of the US as a whole. There are fewer non-white and Hispanic individuals than in the US population. Individuals in the REP are also more highly educated and have higher incomes than the US population.[68] While the data are rich in clinical characteristics, they cover a limited geographic area and cannot be used alone to form the basis of a surveillance program. This type of data source is ideal for population-based studies that allow for description of the full spectrum of disease. There is the potential for recognition of "high-risk" patients early in disease due to the level of clinical data available. Due to the relatively small sample size, it can be difficult to study low-frequency complications. While these data could not stand alone as a surveillance program in IBD, they would play an integral role in such a program by providing such a high level of clinical data and providing the necessary information on prognosis of disease. These data could be combined with other surveillance methods to answer important questions and enhance personalized medicine in IBD.

State-Based Cohort (OSCCAR)—The Ocean State Crohn's and Colitis Area Registry (OSCCAR) is a prospective inception cohort of IBD. The broad spectrum of IBD has not been well represented at cohorts of convenience at tertiary care centers. Retrospective cohorts do not permit for unbiased assessment of exposures, risk factors outcomes, or biobanking of baseline specimens. Rhode Island was chosen as the site of this inception cohort, as it is a densely populated area with care generally delivered within this state. The population is 10 fold larger than that of Olmsted County in Minnesota, and there is also somewhat greater ethnic diversity. The goals of this cohort were to estimate incidence, create models of risk stratification, and enhance studies of disease causation and susceptibility.

There is a complex biopsychosocial model of factors that influence prognosis in IBD, and understanding these components can likely only be done in a prospective fashion. To understand prognosis, data collection needs to include detailed genetic, phenotypic, environmental and therapeutic factors (Figure 1). For example, in the literature, factors such as corticosteroid use at first flare, age at diagnosis, smoking, perianal fistulas, disease extent, disease location, severity on endoscopic measurements, and serologic reactivity to microbial antigens have been associated with prognosis.

The study design includes academic detailing of practices, home visits to patients, a combination of prospective and retrospective data collection, and central reading of pathology and imaging. Data collection comes from the patient, with confirmation in medical records via direct chart abstraction. Data collection from the patient allows for collection of patient reported outcomes, such as scales of health related quality of life, work

productivity, and specific food frequency questionnaires. Specimens are also collected as baseline and follow-up visits.

Thus far, data from this cohort show the incidence of IBD in Rhode Island by age from 2008-2010 to be similar to that reported in Olmsted County, with a peak at age 20-29 years. The advantages to this type of inception cohort are myriad, with detailed information on clinical characteristics, patient-reported outcomes, validated disease and exposures from clinical record review. The results will yield important results on prognosis of disease. However, there are numerous limitations as well. The cohort is small, there is the potential for incomplete capture due to clinical care sought out of state, the population is not necessarily generalizable to the remainder of the US, and the effort is costly. However, as long as the range of relevant variables is covered in a regional cohort, getting representativeness across the US is an unnecessary goal. This type of inception cohort answers a series of very important questions, but cannot stand alone as a surveillance system in IBD.

Nationwide Inpatient Sample—The Nationwide Inpatient Sample (NIS) is part of the healthcare utilization project (HCUP) sponsored by the Agency for Healthcare Research and Quality (AHRQ). It comprises all discharge records from a 20% stratified sampling of acute care hospitals in the United States since 1988. This sample is representative of approximately 5,000 hospitals included in the American Hospital Association Survey but does not include Veterans Affairs and federal hospitals. Hospitals are sampled from within 60 strata based on geographic region, urban vs. rural location, teaching status, hospital bed-size, and ownership. The number of states that contribute data to the NIS has gradually increased from 8 states in 1988 to 45 states in 2010. Data that are collected in the NIS include demographic information on age, sex, race and ethnicity, health insurance payer, and neighborhood income. Clinical fields in the database include type of admission (emergent vs. elective), primary and secondary diagnoses and procedures, length of stay, inpatient charges, and disposition upon discharge. The NIS also contains data on hospital characteristics geographic location, rural status, teaching status, and hospital bed-size.

The NIS has been analyzed for many IBD-related studies including descriptive epidemiology of hospitalizations and surgery.[1] This analysis has been performed on underserved populations such as the uninsured, who are included in few other databases.[2] The NIS has been used to describe variation in utilization of surgery among geographic regions and racial/ethnic groups.[3, 4] It has also been used to characterize the impact of admission and procedural volume on clinical outcomes in the IBD population.[5, 6] Additionally, NIS data have been used to describe the increased prevalence of *C. difficile*, methicillin-resistant *Staph. aureus* (MRSA) and venous thromboembolism (VTE) in hospitalized IBD patients compared to general medical inpatients. Associations between these factors and mortality in IBD have also been reported.[7-1011]

Among the key strengths of the NIS is its large national sample size spanning over 2 decades which enables time trend and group analysis even in a relatively less common condition such as IBD. The NIS also contains racially and ethnically diverse data from all insurance payer types, particularly the uninsured. However, it has several important

limitations revolving around the lack of personal identifiers. Consequently, the validity of IBD diagnosis codes in the NIS cannot be confirmed by medical records. Though inpatient IBD diagnosis codes have been shown to be quite good in Canadian and VA databases, this validation is not necessarily generalizable to the NIS.[12] Any potential misclassification, even non-differential, may compromise findings from descriptive epidemiology studies. An additional limitation is the high rate of missing racial and ethnic data, which approaches 25%. However, the missing racial/ethnic data are specific to certain states that do not collect them. Thus, exclusion of these states from analyses does not necessarily introduce bias.

Despite these limitations, the NIS may be useful for identifying time trends in IBD-related hospitalizations, surgeries, and other clinical outcomes. Systematic errors in IBD coding are not expected to change year-to-year, so relative changes in hospitalizations and surgeries are likely to be valid and can be used to identify sentinel increases in those outcomes. Increases in hospitalizations, however, do not necessarily reflect increases in IBD prevalence but may also be attributable to increased disease severity or decreased access to ambulatory IBD care. As these data are limited to hospitalizations, they cannot be used as a mechanism of estimating incidence and prevalence of disease. Instead, these data would play an important role in understanding the burden of disease.

National Health Interview Survey—The National Health Interview Survey (NHIS) is an annual cross-sectional household health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The target population for the NHIS is the civilian non-institutionalized population of the US. The survey uses a multi-stage clustered sample design, with oversampling of black, Hispanic, and Asian persons, and produces nationally representative data on health insurance coverage, health care access and utilization, health status, health behaviors, and other health-related topics. The in-person interviews are conducted by roughly 750 trained interviewers with the US Census Bureau using computer-assisted personal interviewing (CAPI). In this technique, telephone interviewing is allowed to complete missing sections of the interview when in-person interviewing is not possible.

The NHIS questionnaire consists of a core set of questions that remain relatively unchanged from year to year, and supplemental questions that vary from year to year to collect additional data pertaining to current health issues of national importance. The core survey instrument contains four main modules: Household Composition, Family, Sample Child, and Sample Adult. For the household composition module, a household respondent provides basic sociodemographic information on all members of the household. Within each family, the family module is completed by a family respondent who provides health information on each member of the family. All available adults in the family are invited to participate in the family interview. Additional health information is subsequently collected from one randomly selected adult (the "sample adult") aged 18 years or older and from the parent or guardian of one randomly selected child under age 18 (the "sample child").

The last time IBD was included in the NHIS was in 1999 as a supplement to the sample adult interview. Interviewees were asked if they were given a diagnosis of CD or UC by a physician or other health professional, age at diagnosis, and whether they had IBD-related

symptoms within the prior 12 months. Longobardi and colleagues used the 1999 NHIS data to demonstrate increased healthcare utilization among IBD patients compared to non-IBD controls.[13] Using their analysis, we can also estimate that the prevalence of self-reported IBD to be 876 per 100,000. This figure is nearly 2-3-fold higher than recent estimates of IBD prevalence in North America (UC, 249 per 100,000 persons; CD, 319 per 100,000 persons).[5] This discrepancy underscores the difference between prevalence estimates based on self-report compared with those ascertained by IBD-related health encounters. It should be noted that self-reported IBD in the NHIS is not confirmed with the medical record and may reflect over-reporting. However, there are patients with IBD in long-term remission who do not seek IBD-related healthcare, and these individuals may be undercounted using health administrative data. Another potential limitation of the NHIS is that the final sample adult response rate in 1999 was 70%, and IBD patients may have been more likely to agree to the survey leading to a response bias. However, age is a correlate of both response and many health outcomes and the sample adult weights are calibrated to Census-based population controls including age. This certainly helps to address nonresponse bias in sample adult estimates. Lastly, the exclusion of children may have partly explained the higher prevalence estimates in the NHIS.

The NHIS is potentially a very useful tool to assess IBD burden in the US because it represents a nationwide population-based sample and uses an existing infrastructure to collect data. Its oversampling of Hisapnics, non-Hispanic Blacks, and non-Hispanic Asians also enables descriptive epidemiology for important minority groups. The presence of personal identifiers also enables linkage to databases such as the National Death Index and the Medical Expenditure Panel Survey (MEPS). Sample redesign of this survey is done every 10 years, with a redesigned sample due to be released next in 2016. A content redesign is scheduled for implementation in 2017. Inclusion of an IBD component would represent an invaluable opportunity. This can be expensive, with a cost of approximately 1 million dollars to add a 3-4 minute series of core questions to the survey, such as those specific to IBD. An additional challenge is the need to validate self-reported IBD. This could be achievable by asking the interviewee permission to contact treating physicians. The infrastructure for such validation is already in place in that the current survey does validate self-reported immunizations in children with the medical record. Finally, the total IBD sample size was relatively small (N=271 in 1999) and may be insufficient for subgroup analyses. This limitation could be addressed by including IBD-related questions on the NHIS in multiple years.

Population-based surveys or cohort studies (Harvard Professional Health Studies)—The Harvard Professional Health Studies are a group of prospective cohorts of women (Nurses' Health Study I & II) and men (Health Professional Follow-up studies) that were initiated in 1976, 1989, and 1986, respectively, and are followed through biennial questionnaires ascertaining diet, lifestyle, and health history including new medical diagnoses. With a follow-up rate exceeding 90%, they have contributed significantly to understanding chronic disease epidemiology. Diagnosis is through initial self-report, followed by a supplemental questionnaire, and medical record review of every reported case.

The infrastructure needs for using the cohort for the study of IBD studies is difficult to isolate as questionnaire development and administration is encompassed within the larger efforts of cohort follow-up, which include collection of multiple exposures and surveillance for the incidence of several disease outcomes. Resources generally required for IBD-specific studies require funding to support research assistants to contact participants, obtain medical records, and create case files. Currently, natural history of disease including complications, need for surgery and various treatments are obtained through self-report on a supplemental questionnaire and then validated on medical record review. However, this assessment is obtained at the time of initial self-report. At present, follow-up questionnaires are not being administered. Biospecimens are available (before or after diagnosis) in a subset (approximately one-third of the cohort) consisting of buffy coat for DNA and stored plasma.

There are several examples of IBD studies from within this cohort, particularly focused upon risk factors for development of disease (vitamin D, stress/anxiety, medication use, early life factors, smoking and dietary factors).[81-85] The advantages of this cohort pertain to an accurate ascertainment of incident IBD through physician-confirmation of reported cases. The cohort is ideal for studying environmental exposures as they relate to disease risk and possible interactions between genetic and environmental factors. Validation ensures accuracy of self-reported cases in all cases. A key advantage is that by simultaneously assessing multiple exposures and outcomes, "disease-specific" costs are kept lower than what they would otherwise be. There are also some distinct limitations to this cohort. Women enrolled in the initial Nurses' Health Study now represent an older population and the cohorts are relatively gender-specific. However ongoing expansion of the target population through newer cohorts would potentially mitigate some of this skewed distribution. Furthermore, the results from this cohort have been consistent with other published epidemiologic studies, supporting continued utility of such cohort. The high degree of medical knowledge of the participants is also a substantial advantage of this cohort. Use of this and other established cohorts in IBD surveillance would require additional infrastructure to obtain periodic follow-up, including time of medication initiation, surgery, hospitalization, and other disease-specific factors.

This form of surveillance is ideal in many ways for a US surveillance program as structured assessments of exposures prior to disease onset allows for identifying environmental risks through a high-quality prospective cohort, and allows for tracking incident disease which is otherwise challenging. The fact that the cohort assesses multiple diseases makes it efficient and less expensive. Studies from within this cohort have taken a structured approach to validation, which would be key in any future surveillance program. Participants need to remain motivated to ensure high follow-up rates. In these cohorts, this has been accomplished through periodic feedback of results (e.g., through newsletters) from the study. The Harvard Professional Health Studies have central data storage and a structured protocol for proposing a study, executing it, and publishing results. These factors ensure data integrity and would be needed should similar methodology be used in a national IBD surveillance program.

Veteran's Administration Data—The Department of Veterans Affairs (VA) is the largest healthcare system in the United States. The VA has kept computerized records of all

its patients for more than 30 years. These records comprise the Patient Treatment File (PTF) with inpatient data since 1970. From 1970 until 1980, diagnoses in the PTF were recorded using the 8th revision of the International Classification of Diseases (ICD-8). From 1981 until 2006, ICD-9 has been used. Codes have been validated within the VA system, with a reasonable reliability for disease status. The positive predictive value (PPV) of CD codes was superior (88-100%) to that of UC (50-93%), but the accuracy of extra-intestinal manifestations and surgeries was poor (PPV 0-29%). [86]

There are numerous examples of studies on IBD etiology, geographic variation of disease within the VA system, complications, medication effects, and quality assessment of preventive measures using the VA population.[87-93] For example, in a study of colorectal cancer (CRC) in patients with IBD as compared to those without, CRC was found to affect IBD patients at a younger age and have a more proximal localization when compared with CRC of non-IBD patients.[87] In a study of etiologic risk factors for IBD in the veteran population, military duty in Vietnam and a status as prisoner of war both were protective against the development of CD, but not UC (OR 0.84; 0.75-0.96 and 0.60; 0.41-0.87, respectively).[88] Khan et al performed a recent study on adherence to mesalamine and its influence on UC flares. Long-term high-adherence level was lower than previously reported (40% range). Adherence was a significant factor in predicting disease flares.[91] Other studies from this group have investigated metabolic bone disease, prevalence of corticosteroid use and lymphoma risk in individuals with UC on thiopurines.[94-96] Studies directly related to quality of care within the VA have also been performed. For example, Hou et al reviewed monitoring for myelosuppression in veterans on thopurines for IBD. Monitoring after medication initiation was low with wide variability based on facility.[90] As these few examples suggest, data from the VA system can be used for a wide variety of research and quality improvement purposes.

One of the distinct advantages to this database is that individuals can be followed throughout a long time period of coverage, allowing for cohort studies and studies of outcomes that can take years to develop (such as malignancies). Because medications are recorded within the database, pharmacoepidemiologic studies are also possible. One example is the recent study by Khan et al demonstrating an increased risk of lymphoma in individuals with UC on thiopurines.[94] There are several disadvantages to this data source as well. For each patient, a record of the previous medical history needs to be re-assembled from multiple individual annual data files, including separate inpatient and outpatient files, medical and surgical procedural files, medication file, and a death file. The programming to use these data is quite labor intensive, as is the approval process. Because medical test results are unavailable within the database, the investigator merely knows that the test has been performed (such as laboratory, pathology, radiology and endoscopy results). Additionally, some veterans use the VA for prescription medication coverage but are hospitalized or obtain much of their care outside the VA system; in these instances the external records and outcomes are not available. There are ongoing changes in eligibility for VA benefits, resulting in an unknown population base (or denominator). The lack of a known denominator affects the ability to perform disease surveillance. Historically, the VA system has been predominantly caring for male patients, limiting the generalizability of its data to women. Although the VA system includes a large population, for some epidemiologic questions the sample size still remains

too small. Such questions pertain, for instance, to the seasonal variation of IBD, its distribution among ethnic minorities, or its underlying birth-cohort patterns. Data from the VA system could not stand alone as a surveillance system for IBD, but could serve as a valuable component of a disease surveillance system.

Surveillance Epidemiology and End Results program (SEER)—The Surveillance Epidemiology and End Results program (SEER) is sponsored by the National Cancer Institute (NCI). This program collects data on cancer incidence from population-based registries covering approximately 28% of the US population. This program is designed to over sample traditionally under-represented populations such as Hispanics, American Indians and Alaskan Natives, Asians, and Pacific Islanders. The sample is comparable with the US population in regards to income and education. The data are updated annually. SEER registries are located in particular states, located in various regions of the country.

The process for capturing data within SEER relies on mandatory reporting of cancer diagnoses. Within each SEER area, health care providers who diagnose or provide first courses of treatment are required to report cancer cases to the SEER registry. Some health care facilities have onsite cancer registers to aid in this process. After a cancer report is received, a cancer register is sent to the facility with 6 months to abstract data on the case. This form of data collection can be time consuming and expensive, although the data are of high quality.

The goals of the SEER program are to collect complete and accurate cancer data. The program periodically reports on the burden of cancer in the US. Researchers and public health policy makers also use these data to describe temporal changes in cancer diagnoses. The program can identify any unusual patterns in cancer incidence, including iatrogenic cancers. Data are provided as a resource to researchers for a variety of other study purposes. The data currently collected within SEER include demographics, primary tumor site, tumor morphology, stage at diagnosis, first course of treatment, survival, and cause of death. Census data are also included every decade, such as median household income, percent with high school education, percent speaking English, population density, and urban versus rural status based on zip code. There are also data on availability of health resources (for example, facilities for radiation, etc).

Importantly, SEER data can be linked to other sources for research and public health purposes. For example, data can be linked to Medicare data or to other national data repositories such as the Medicare Health Outcomes Survey (MHOS) or the National Longitudinal Mortality Study (NLMS). SEER-Medicare data include information on 100% of cancer patients in SEER areas. These data also include a 5% random sample of non-cancer patients in these areas. Longitudinal Medicare claim data are also included, with factors such as short stay hospitalization data, physician or lab services, hospital outpatient claims and part D prescription drug claims. There are several strengths to SEER-Medicare data, including a large number of patients within a diverse geographical base, longitudinal follow up, data on non-cancer controls for comparisons, and the population-based nature of the cohort. There are also several limitations to SEER data. These data only include individuals age 65 and older and a lack of inclusion of data on non-covered services, reasons

for tests or services, the results of testing, or confounders associated with treatment. SEER-Medicare data are well suited for health care utilization research, or for studies on late outcomes of cancer treatments. Types of research that are more challenging using these data include: treatment effectiveness studies, long term symptoms or studies on recurrence of cancer or disease progression. MHOS is a supplemental questionnaire administered to a random sampling of 1000 beneficiaries from each managed care organization that participates in Medicare. There are 95 questions included related to demographics, socioeconomic status, activities of daily living, and quality of life. A new MHOS cohort is selected every year, and follow-up data are obtained 2 years later. A total of 8 MHOS cohorts have been linked to SEER (approximately 55,000 individuals). The strengths of this linkage include the ability to evaluate health-related quality of life (HRQoL) prior to and after a cancer diagnosis as well after an initial treatment. The impact of cancer on activities of daily living (ADLs) can also be determined. The weaknesses to these data are that they are limited to a small sample and that the entire sample comes from Medicare managed care beneficiaries. These data also do not contain information on recurrence or specifics of the cancer treatment. SEER-NLMS data combine socioeconomic data collected by Current Population Surveys and cause of death. These data do include such characteristics as race/ethnicity, marital status, education, income, employment status, occupation, household size, immigrant status, limited smoking data, and health insurance status. This resource includes 3 million people, with cause of death in approximately 250,000. These data have been linked to SEER for approximately 35,000 individuals with cancer diagnoses through 2003.

SEER data have been used in prior publications studying IBD-associated cancer. For example, several studies on the risk of lymphoma in IBD have used this data source.[97-99] Sultan et al compared the prognosis of cancer in individuals with IBD with survival outcomes in SEER.[100] Other studies have compared IBD associated colorectal cancer (CRC) with sporadic CRC,[101, 102] or rates of early/missed cancers after colonoscopy in older patients with and without IBD.[103] While SEER itself could not be used as a surveillance system for IBD, this model could be extended to IBD related outcomes and complications, which are an important part of disease surveillance. Established IBD cohorts could be linked to SEER registries to capture IBD associated cancer. To develop a SEER type infrastructure to capture incident cases of IBD would be costly and would require that IBD be considered a reportable condition (i.e., mandatory reporting within a defined population). There are additional challenges as well, such as determination of who does the reporting and the fact that an IBD diagnosis can be more complicated than a cancer diagnosis, requiring multiple components such as clinical symptoms, endoscopic, pathologic and radiologic findings. As IBD is a rare condition, there is a lower economy of scale for establishing a SEER-like registry.

Internet Based Surveys—Most longitudinal cohort studies of CD and UC in the US have relied on administrative data or retrospective review of clinical records. Such studies often lack important clinical data that can only be collected directly from the patient. Examples include detailed smoking history, depression, sleep, physical activity, and diet. Traditionally, prospective cohort studies have been extraordinarily expensive to conduct due to the need for large personnel resources to recruit and follow patients over time. With the

advent of widespread availability of the internet, the opportunity exists to prospectively follow patients longitudinally using electronic methods at a dramatically reduced cost. The Crohn's & Colitis Foundation of America's (CCFA) Partners research program is a cohort study that utilizes internet-based recruitment and data collection to assemble detailed cross sectional and longitudinal data on large numbers of IBD patients.[104]

The objective of CCFA Partners is to establish a large and diverse registry of IBD patients that can be identified, enrolled and followed using the internet. To date, more than 12,000 subjects have been enrolled, the majority joined after recruitment from the nearly 400,000 person email roster maintained by the CCFA.

Detailed information is collected at baseline on disease type, location, medications, family history, disease activity, prevention activities (such as influenza vaccine), exercise, quality of life and patient reported outcomes. Every six months this information is updated. Reminders at 3 month intervals are designed to help maintain interest and deliver educational messages. Rotating modules are introduced to address specific questions developed by the data management team or by outside investigators who have peer-reviewed ancillary studies. A validation study comparing patient-reported disease type to physician-confirmed diagnosis has been completed and shows >95% specificity for IBD diagnosis (unpublished data).

Several studies have been published using data from this cohort, including a study on the role of diet in exacerbation of IBD symptoms, the perception of chronic illness care and sleep disturbance as a risk factors for disease relapse.[104-106] There are several advantages to this online cohort. The cohort was specifically designed to focus on patient-reported outcomes such as anxiety, depression, fatigue, pain, and sleep. There are large numbers of potential study subjects, low cost, and rapid results. Because study subjects are surveyed every six months and because questionnaires do not require printing, it is possible to introduce new questions and achieve results quickly. One central disadvantage to this cohort is that the population is highly selected and may not be representative of the general IBD population. The study is also threatened by drop-outs which will reduce study power and risk bias if the drop-outs are differential.

Unfortunately, this type of study cannot be used for traditional population surveillance. It is not possible to determine incidence and prevalence for example, as there is no underlying denominator to the population in its current design. However, the study can be used to conduct surveillance on enrolled subjects to determine the course of their disease over time and the impact of the disease on quality of life and function. This unique resource could certainly help to estimate the burden of disease within the IBD population.

Medical Informatics Approaches to Electronic Health Records—The current capabilities of electronic health records (EHRs) are ever changing and these records have evolved from storage of records to being able to track quality factors and generate reports. Data flow from the front end to the back end of the system using inputs in data elements. These systems can also facilitate order entry. Through order entry systems, systems can enact decision support features (for example, IBD-specific patient order sets to improve

compliance with preventive measures such as thromboembolism prophylaxis during hospitalizations). These systems can also utilize outpatient clinical reminders in the decision support system.

After information is entered into an EHR, it is stored as structured and unstructured data elements. Structured data elements include objective data such as medications, laboratory values, vital signs, demographics and diagnoses. Unstructured data consist of chart elements such as the chief complaint, or free text procedures and radiology reports. Currently, using data from EHRs, queries can be run to determine the total number of patients with a condition, such as IBD. Other structured data, such as number of visits, or medications are available. These systems can also be used to track quality process measures. There are also many things that EHRs cannot do. For example, a denominator cannot be generated from an EHR. In theory, if EHRs could eventually be combined throughout the US, this could form the start of a national surveillance program. However, there is limited penetration of EHRs throughout the US (only 60% in some estimates). Currently, there are 550 separate ambulatory vendors with complete EHRs approved by CMS. There are also not currently widely adopted standards for data exchange. Additionally, without a unique identifier, these EHRs do not have the ability to differentiate or uniquely identify patients seen in more than one practice.

There may be value in exploring the role of ancillary health information systems in disease surveillance. Other possible non-EHR systems with information about IBD care include laboratory information systems, imaging, pathology, endoscopy reporting systems, billing and communication between patients, pharmacies and insurance providers. Laboratory information systems are the most commonly used systems for infectious and microbiology surveillance. There is a well-defined system in place for mandatory reporting of certain diseases to state public health agencies and the CDC. This system would be inadequate for IBD as no specific lab is ordered in all IBD patients and no specific lab value can confirm the diagnosis. An imaging system would have limitations if used in surveillance, for example, there are variations in practice patterns involved in ordering imaging tests in IBD, there are not structured reports for IBD, and in general imaging is inadequate to diagnose isolated upper gastrointestinal or colonic disease. Surgical pathology systems in general do not utilize structured data, which can be a challenge when using these data for a surveillance system. There are a number of different electronic endoscopic record systems.[107] Data commonly available through these systems include factors such as scheduling data, monitoring of vital signs during a procedure, endoscopic record, correspondence with referring physicians and billing codes. Some of the advantages to using this system include the rich procedural details available and has excellent yield for upper GI tract and colonic IBD. The disadvantages of using this system include the low yield for isolated small bowel disease and the problem of unique identification.

There are numerous examples of studies on IBD from EHRs. Reddy et al examined EHR data to determine whether patients with IBD were receiving optimal care (such as addition of steroid sparing agents, or colon cancer surveillance).[108] There were high levels of noncompliance with important IBD quality measures. EHRs have also been used to perform studies of medication adherence in IBD. Kane demonstrated that medication non-adherence

correlated significantly with relapse of disease in patients with UC on 5-aminosalicylate (5-ASA) therapy.[109] There are many other retrospective studies in single centers describing disease outcomes and complications. These studies are greatly facilitated through EHRs.

How data from EHRs could add to a surveillance system in IBD remains to be seen. As technology advances, there may be the potential to embed cloud-based registries within the EHR. Any such efforts would also need to address the unique identification issue. One model for leveraging the health care system electronic records for the purposes of surveillance is the Federal Drug Administration (FDA) Mini-Sentinel program. This database contains information on over 130 million individuals from a combined total of 18 partner organizations. These data encompass nearly 400 million person-years of observation time. The Mini-Sentinel Common Data Model standardizes administrative and clinical information across the included Data Partners. Data Partners execute, within their own institutions' firewalls, standardized computer programs (e.g., modular programs) provided by the Operations Center or project workgroups. These data form one basis for surveillance of disease outbreaks in the US. Certainly, there are challenges to using these data in their current form as a tool to monitor surveillance in IBD, but as technologies emerge, the EHR may play an important role future IBD surveillance programs.

Possible Solutions to IBD surveillance in the US

The health care system in the US presents unique challenges to the development of a surveillance system. While we can learn important lesions from other countries that have been successful in this endeavor, the fragmentation of the US system and the lack of a unique personal identification number used for health care encounters makes this charge of IBD surveillance an even greater challenge. Unlike other disease surveillance programs in the US where a laboratory value alone can trigger a cascade of events, IBD diagnosis itself (requiring endoscopic, pathologic, and radiologic findings) also does not lend itself to the current surveillance model.

It becomes important to use a series of tools to accomplish the goals of a disease surveillance program (figure 2). In essence, the right tool must be used for the right job. In order to provide estimates of prevalence, Medicare/Medicaid data, closed panel HMO data and commercial US administrative claims data may be used, and with appropriate sampling and weighting, can produce national estimates that are reasonable enough. Other national surveys and data sources, such as the NHIS, could be leveraged as a means of determining incidence and prevalence if the sample size were large enough and representative of the overall population. IBD could be added to the screened conditions within this questionnaire at the next revision (2017), with preparations to validate the self-reported diagnoses of IBD. National-level data might particularly focus upon estimating the incidence and prevalence in those groups previously undercounted (such as minorities, immigrants, and children). For race in particular, these data would need to be captured under the auspices of a program such as NHIS, as administrative data are not sufficiently reliable in this regard. To complement these data, regional estimates of incidence, prevalence, and other factors such as natural history of disease, etiology and burden should also be abstracted from established population-based cohorts such as Olmsted County and OSCCAR. These data sources with a

richer level of clinical data could significantly enhance the counts available from administrative data sources and could answer important questions not amenable to administrative data sources. Novel regional population-based cohorts might be needed to achieve greater population sampling, including the southern and western regions and of African American, Latino, and Asian populations.

Perhaps it is time to move beyond assessments of incidence and prevalence. In effect, by rejecting the charge of only monitoring surveillance, much more can be accomplished. In general, the model of surveillance does not fit well with a disease such as IBD. IBD is not subject to outbreaks, as seen with infectious diseases where real time systems of disease surveillance are needed. Reported incidence and prevalence rates of IBD in various populations within the US, arrived at from several different data systems and approaches, remain in similar ranges. These estimates of disease may be good enough at this point. Future incidence and prevalence estimates should likely focus on previously under-counted populations.

Perhaps a broader definition of surveillance system could incorporate many other important aspects of disease (figure 2). Rather than merely estimating the incidence and prevalence of disease, we should be providing surveillance of the burden of disease to society. Factors involved in the burden of disease could be costs, loss of work productivity, morbidity, disease related cancers and other factors downstream from diagnosis. At this point it remains unknown whether the burden of disease is shared equally by the estimated 1.4 million Americans with IBD, or if this burden is concentrated in a portion of those affected. This surveillance program would be a new model for disease surveillance, in essence focusing upon "surveillance of the burden of disease." This system would focus separately upon 1) natural history of disease and 2) outcomes and complications of the disease and/or treatments.

Because we cannot prevent IBD, we should focus on preventing its potentially avoidable complications. A multi-faceted model of surveillance of the burden of disease (such as that seen in figure 2) could accomplish this important goal. However, research into the etiology of disease is still of great importance, particularly among populations with rapidly rising rates of IBD diagnosis. These populations, such as immigrants, may provide important etiologic clues as to disease development. It is possible that some of these etiologic clues are shared across autoimmune diseases, and may aid in future preventive efforts outside the direct arena of IBD.

Therefore, after a thorough review of epidemiologic data sources available in the US, and a review of successful international surveillance programs, we have concluded that the US should broaden the goals of surveillance to include surveillance of the burden of disease. As outlined in this document, this can only be accomplished through a combination of approaches, using unique data resources ranging from administrative claims data to national surveys to inception cohorts of IBD. With this model in place, we will be able to direct appropriate resources to the IBD community in order to begin to decrease the burden of disease.

References

1. Kappelman MD, Rifas-Shiman SL, Porter CQ, Ollendorf DA, Sandler RS, Galanko JA, Finkelstein JA. Direct health care costs of Crohn's disease and ulcerative colitis in US children and adults. Gastroenterology. 2008; 135:1907–13. [PubMed: 18854185]

2. Park KT, Bass D. Inflammatory bowel disease-attributable costs and cost-effective strategies in the United States: a review. Inflamm Bowel Dis. 2011; 17:1603–9. [PubMed: 21053357]

3. Kappelman MD, Rifas-Shiman SL, Kleinman K, Ollendorf D, Bousvaros A, Grand RJ, Finkelstein JA. The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. Clin Gastroenterol Hepatol. 2007; 5:1424–9. [PubMed: 17904915]

4. Loftus CG, Loftus EV Jr. Harmsen WS, Zinsmeister AR, Tremaine WJ, Melton LJ 3rd, Sandborn WJ. Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940-2000. Inflamm Bowel Dis. 2007; 13:254–61. [PubMed: 17206702]

5. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW, Kaplan GG. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology. 2012; 142:46–54. e42. quiz e30. [PubMed: 22001864]

6. Ng SC, Bernstein CN, Vatn MH, Lakatos PL, Loftus EV Jr. Tysk C, O'Morain C, Moum B, Colombel JF. Geographical variability and environmental risk factors in inflammatory bowel disease. Gut. 2013; 62:630–49. [PubMed: 23335431]

7. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. Inflamm Bowel Dis. 2011; 17:423–39. [PubMed: 20564651]

8. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology. 2004; 126:1504–17. [PubMed: 15168363]

9. Bernstein CN, Shanahan F. Disorders of a modern lifestyle: reconciling the epidemiology of inflammatory bowel diseases. Gut. 2008; 57:1185–91. [PubMed: 18515412]

10. Calkins BM, Lilienfeld AM, Garland CF, Mendeloff AI. Trends in incidence rates of ulcerative colitis and Crohn's disease. Dig Dis Sci. 1984; 29:913–20. [PubMed: 6478982]

11. Sonnenberg A, Wasserman IH. Epidemiology of inflammatory bowel disease among U.S. military veterans. Gastroenterology. 1991; 101:122–30. [PubMed: 2044901]

12. Sonnenberg A, McCarty DJ, Jacobsen SJ. Geographic variation of inflammatory bowel disease within the United States. Gastroenterology. 1991; 100:143–9. [PubMed: 1983816]

13. Kurata JH, Kantor-Fish S, Frankl H, Godby P, Vadheim CM. Crohn's disease among ethnic groups in a large health maintenance organization. Gastroenterology. 1992; 102:1940–8. [PubMed: 1587413]

14. Sonnenberg A. Demographic characteristics of hospitalized IBD patients. Dig Dis Sci. 2009; 54:2449–55. [PubMed: 19760158]

15. Sewell JL, Yee HF Jr. Inadomi JM. Hospitalizations are increasing among minority patients with Crohn's disease and ulcerative colitis. Inflamm Bowel Dis. 2010; 16:204–7. [PubMed: 19575353]

16. Carr I, Mayberry JF. The effects of migration on ulcerative colitis: a three-year prospective study among Europeans and first- and second- generation South Asians in Leicester (1991-1994). Am J Gastroenterol. 1999; 94:2918–22. [PubMed: 10520845]

17. Malaty HM, Fan X, Opekun AR, Thibodeaux C, Ferry GD. Rising incidence of inflammatory bowel disease among children: a 12-year study. J Pediatr Gastroenterol Nutr. 2010; 50:27–31. [PubMed: 19934770]

18. Adamiak T, Walkiewicz-Jedrzejczak D, Fish D, Brown C, Tung J, Khan K, Faubion W Jr. Park R, Heikenen J, Yaffee M, Rivera-Bennett MT, Wiedkamp M, Stephens M, Noel R, Nugent M, Nebel J, Simpson P, Kappelman MD, Kugathasan S. Incidence, Clinical Characteristics, and Natural History of Pediatric IBD in Wisconsin: a Population-based Epidemiological Study. Inflamm Bowel Dis. 2013

19. Abramson O, Durant M, Mow W, Finley A, Kodali P, Wong A, Tavares V, McCroskey E, Liu L, Lewis JD, Allison JE, Flowers N, Hutfless S, Velayos FS, Perry GS, Cannon R, Herrinton LJ.

Incidence, prevalence, and time trends of pediatric inflammatory bowel disease in Northern California, 1996 to 2006. J Pediatr. 2010; 157:233–239. e1. [PubMed: 20400099]

20. Herrinton LJ, Liu L, Lewis JD, Griffin PM, Allison J. Incidence and prevalence of inflammatory bowel disease in a Northern California managed care organization, 1996-2002. Am J Gastroenterol. 2008; 103:1998–2006. [PubMed: 18796097]

21. Weng X, Liu L, Barcellos LF, Allison JE, Herrinton LJ. Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern california-managed care organization. Am J Gastroenterol. 2007; 102:1429–35. [PubMed: 17437504]

22. Herrinton LJ, Liu L, Lafata JE, Allison JE, Andrade SE, Korner EJ, Chan KA, Platt R, Hiatt D, O'Connor S. Estimation of the period prevalence of inflammatory bowel disease among nine health plans using computerized diagnoses and outpatient pharmacy dispensings. Inflamm Bowel Dis. 2007; 13:451–61. [PubMed: 17219403]

23. Hutfless SM, Weng X, Liu L, Allison J, Herrinton LJ. Mortality by medication use among patients with inflammatory bowel disease, 1996-2003. Gastroenterology. 2007; 133:1779–86. [PubMed: 18054550]

24. Sands BE, LeLeiko N, Shah SA, Bright R, Grabert S. OSCCAR: Ocean State Crohn's and Colitis Area Registry. Med Health R I. 2009; 92:82–5, 88. [PubMed: 19385383]

25. Askling J, Dickman PW, Karlen P, Brostrom O, Lapidus A, Lofberg R, Ekbom A. Colorectal cancer rates among first-degree relatives of patients with inflammatory bowel disease: a population-based cohort study. Lancet. 2001; 357:262–6. [PubMed: 11214128]

26. Ekbom A, Helmick C, Zack M, Adami HO. Ulcerative colitis and colorectal cancer. A population-based study. N Engl J Med. 1990; 323:1228–33. [PubMed: 2215606]

27. Soderlund S, Granath F, Brostrom O, Karlen P, Lofberg R, Ekbom A, Askling J. Inflammatory bowel disease confers a lower risk of colorectal cancer to females than to males. Gastroenterology. 2010; 138:1697–703. [PubMed: 20167217]

28. Ma C, Crespin M, Proulx MC, DeSilva S, Hubbard J, Prusinkiewicz M, Nguyen GC, Panaccione R, Ghosh S, Myers RP, Quan H, Kaplan GG. Postoperative complications following colectomy for ulcerative colitis: a validation study. BMC Gastroenterol. 2012; 12:39. [PubMed: 22943760]

29. Kaplan GG, Seow CH, Ghosh S, Molodecky N, Rezaie A, Moran GW, Proulx MC, Hubbard J, MacLean A, Buie D, Panaccione R. Decreasing colectomy rates for ulcerative colitis: a population-based time trend study. Am J Gastroenterol. 2012; 107:1879–87. [PubMed: 23165448]

30. de Silva S, Ma C, Proulx MC, Crespin M, Kaplan BS, Hubbard J, Prusinkiewicz M, Fong A, Panaccione R, Ghosh S, Beck PL, Maclean A, Buie D, Kaplan GG. Postoperative complications and mortality following colectomy for ulcerative colitis. Clin Gastroenterol Hepatol. 2011; 9:972–80. [PubMed: 21806954]

31. Soon IS, Wrobel I, deBruyn JC, Sauve R, Sigalet DL, Kaplan BS, Proulx MC, Kaplan GG. Postoperative complications following colectomy for ulcerative colitis in children. J Pediatr Gastroenterol Nutr. 2012; 54:763–8. [PubMed: 22167014]

32. Rezaie A, Quan H, Fedorak RN, Panaccione R, Hilsden RJ. Development and validation of an administrative case definition for inflammatory bowel diseases. Can J Gastroenterol. 2012; 26:711–7. [PubMed: 23061064]

33. Kaplan GG, Hubbard J, Panaccione R, Shaheen AA, Quan H, Nguyen GC, Dixon E, Ghosh S, Myers RP. Risk of comorbidities on postoperative outcomes in patients with inflammatory bowel disease. Arch Surg. 2011; 146:959–64. [PubMed: 21844437]

34. Bernstein CN, Wajda A, Svenson LW, MacKenzie A, Koehoorn M, Jackson M, Fedorak R, Israel D, Blanchard JF. The epidemiology of inflammatory bowel disease in Canada: a population-based study. Am J Gastroenterol. 2006; 101:1559–68. [PubMed: 16863561]

35. Bernstein CN, Rawsthorne P, Cheang M, Blanchard JF. A population-based case control study of potential risk factors for IBD. Am J Gastroenterol. 2006; 101:993–1002. [PubMed: 16696783]

36. Bernstein CN, Singh S, Graff LA, Walker JR, Miller N, Cheang M. A prospective population-based study of triggers of symptomatic flares in IBD. Am J Gastroenterol. 2010; 105:1994–2002. [PubMed: 20372115]

37. Bernstein CN, Nugent Z, Blanchard JF. 5-aminosalicylate is not chemoprophylactic for colorectal cancer in IBD: a population based study. Am J Gastroenterol. 2011; 106:731–6. [PubMed: 21407180]

38. Graff LA, Vincent N, Walker JR, Clara I, Carr R, Ediger J, Miller N, Rogala L, Rawsthorne P, Lix L, Bernstein CN. A population-based study of fatigue and sleep difficulties in inflammatory bowel disease. Inflamm Bowel Dis. 2011; 17:1882–9. [PubMed: 21830266]

39. Rawsthorne P, Clara I, Graff LA, Bernstein KI, Carr R, Walker JR, Ediger J, Rogala L, Miller N, Bernstein CN. The Manitoba Inflammatory Bowel Disease Cohort Study: a prospective longitudinal evaluation of the use of complementary and alternative medicine services and products. Gut. 2012; 61:521–7. [PubMed: 21836028]

40. Singh H, Nugent Z, Demers AA, Bernstein CN. Increased risk of nonmelanoma skin cancers among individuals with inflammatory bowel disease. Gastroenterology. 2011; 141:1612–20. [PubMed: 21806945]

41. Singh H, Nugent Z, Demers AA, Bernstein CN. Screening for cervical and breast cancer among women with inflammatory bowel disease: a population-based study. Inflamm Bowel Dis. 2011; 17:1741–50. [PubMed: 21744429]

42. Targownik LE, Leslie WD, Carr R, Clara I, Miller N, Rogala L, Graff LA, Walker JR, Bernstein CN. Longitudinal change in bone mineral density in a population-based cohort of patients with inflammatory bowel disease. Calcif Tissue Int. 2012; 91:356–63. [PubMed: 22983212]

43. Targownik LE, Singh H, Nugent Z, Bernstein CN. The epidemiology of colectomy in ulcerative colitis: results from a population-based cohort. Am J Gastroenterol. 2012; 107:1228–35. [PubMed: 22613902]

44. Boualit M, Salleron J, Turck D, Fumery M, Savoye G, Dupas JL, Lerebours E, Duhamel A, Merle V, Cortot A, Colombel JF, Peyrin-Biroulet L, Gower-Rousseau C. Long-term Outcome After First Intestinal Resection in Pediatric-onset Crohn's Disease: A Population-based Study. Inflamm Bowel Dis. 2013; 19:7–14. [PubMed: 22573565]

45. Charpentier C, Salleron J, Savoye G, Fumery M, Merle V, Laberenne JE, Vasseur F, Dupas JL, Cortot A, Dauchet L, Peyrin-Biroulet L, Lerebours E, Colombel JF, Gower-Rousseau C. Natural history of elderly-onset inflammatory bowel disease: a population-based cohort study. Gut. 2013

46. Chouraki V, Savoye G, Dauchet L, Vernier-Massouille G, Dupas JL, Merle V, Laberenne JE, Salomez JL, Lerebours E, Turck D, Cortot A, Gower-Rousseau C, Colombel JF. The changing pattern of Crohn's disease incidence in northern France: a continuing increase in the 10- to 19-year-old age bracket (1988-2007). Aliment Pharmacol Ther. 2011; 33:1133–42. [PubMed: 21488915]

47. Crombe V, Salleron J, Savoye G, Dupas JL, Vernier-Massouille G, Lerebours E, Cortot A, Merle V, Vasseur F, Turck D, Gower-Rousseau C, Lemann M, Colombel JF, Duhamel A. Long-term outcome of treatment with infliximab in pediatric-onset Crohn's disease: a population-based study. Inflamm Bowel Dis. 2011; 17:2144–52. [PubMed: 21287665]

48. Gower-Rousseau C, Vasseur F, Fumery M, Savoye G, Salleron J, Dauchet L, Turck D, Cortot A, Peyrin-Biroulet L, Colombel JF. Epidemiology of inflammatory bowel diseases: new insights from a French population-based registry (EPIMAD). Dig Liver Dis. 2013; 45:89–94. [PubMed: 23107487]

49. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet. 2002; 70:845–57. [PubMed: 11875755]

50. Savoye G, Salleron J, Gower-Rousseau C, Dupas JL, Vernier-Massouille G, Fumery M, Merle V, Lerebours E, Cortot A, Turck D, Salomez JL, Lemann M, Colombel JF, Duhamel A. Clinical predictors at diagnosis of disabling pediatric Crohn's disease. Inflamm Bowel Dis. 2012; 18:2072–8. [PubMed: 22294515]

51. Vernier-Massouille G, Balde M, Salleron J, Turck D, Dupas JL, Mouterde O, Merle V, Salomez JL, Branche J, Marti R, Lerebours E, Cortot A, Gower-Rousseau C, Colombel JF. Natural history of pediatric Crohn's disease: a population-based cohort study. Gastroenterology. 2008; 135:1106–13. [PubMed: 18692056]

52. Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. Pharmacoepidemiol Drug Saf. 2009; 18:1086–93. [PubMed: 19672855]

53. http://www.cms.gov/Research-Statistics-Data-and-Systems/Research/ResearchGenInfo/CMSStatistics.html

54. Shaukat A, Virnig DJ, Salfiti NI, Howard DH, Sitaraman SV, Liff JM. Is inflammatory bowel disease an important risk factor among older persons with colorectal cancer in the United States? A population-based case-control study. Dig Dis Sci. 2011; 56:2378–83. [PubMed: 21409378]

55. Shaukat A, Virnig DJ, Howard D, Sitaraman SV, Liff JM, Lederle FA. Crohn's disease and small bowel adenocarcinoma: a population-based case-control study. Cancer Epidemiol Biomarkers Prev. 2011; 20:1120–3. [PubMed: 21467236]

56. Sanoff HK, Carpenter WR, Sturmer T, Goldberg RM, Martin CF, Fine JP, McCleary NJ, Meyerhardt JA, Niland J, Kahn KL, Schymura MJ, Schrag D. Effect of adjuvant chemotherapy on survival of patients with stage III colon cancer diagnosed after age 75 years. J Clin Oncol. 2012; 30:2624–34. [PubMed: 22665536]

57. Braithwaite D, Zhu W, Hubbard RA, O'Meara ES, Miglioretti DL, Geller B, Dittus K, Moore D, Wernli KJ, Mandelblatt J, Kerlikowske K. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst. 2013; 105:334–41. [PubMed: 23385442]

58. Baxter NN, Warren JL, Barrett MJ, Stukel TA, Doria-Rose VP. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol. 2012; 30:2664–9. [PubMed: 22689809]

59. Haynes K, Beukelman T, Curtis JR, Newcomb C, Herrinton LJ, Graham DJ, Solomon DH, Griffin MR, Chen L, Liu L, Saag KG, Lewis JD. Tumor necrosis factor alpha inhibitor therapy and cancer risk in chronic immune-mediated diseases. Arthritis Rheum. 2013; 65:48–58. [PubMed: 23055441]

60. Grijalva CG, Chen L, Delzell E, Baddley JW, Beukelman T, Winthrop KL, Griffin MR, Herrinton LJ, Liu L, Ouellet-Hellstrom R, Patkar NM, Solomon DH, Lewis JD, Xie F, Saag KG, Curtis JR. Initiation of tumor necrosis factor-alpha antagonists and the risk of hospitalization for infection in patients with autoimmune diseases. JAMA. 2011; 306:2331–9. [PubMed: 22056398]

61. Arora G, Singh G, Vadhavkar S, Shah SB, Mannalithara A, Mithal A, Triadafilopoulos G. Incidence and risk of intestinal and extra-intestinal complications in Medicaid patients with inflammatory bowel disease: a 5-year population-based study. Dig Dis Sci. 2010; 55:1689–95. [PubMed: 20428948]

62. Herrinton LJ, Curtis JR, Chen L, Liu L, Delzell E, Lewis JD, Solomon DH, Griffin MR, Ouellet-Hellstom R, Beukelman T, Grijalva CG, Haynes K, Kuriya B, Lii J, Mitchel E, Patkar N, Rassen J, Winthrop KL, Nourjah P, Saag KG. Study design for a comprehensive assessment of biologic safety using multiple healthcare data systems. Pharmacoepidemiol Drug Saf. 2011; 20:1199–209. [PubMed: 21919113]

63. Herrinton LJ, Liu L, Fireman B, Lewis JD, Allison JE, Flowers N, Hutfless S, Velayos FS, Abramson O, Altschuler A, Perry GS. Time trends in therapies and outcomes for adult inflammatory bowel disease, Northern California, 1998-2005. Gastroenterology. 2009; 137:502–11. [PubMed: 19445944]

64. Herrinton LJ, Liu L, Weng X, Lewis JD, Hutfless S, Allison JE. Role of thiopurine and anti-TNF therapy in lymphoma in inflammatory bowel disease. Am J Gastroenterol. 2011; 106:2146–53. [PubMed: 22031357]

65. Velayos FS, Liu L, Lewis JD, Allison JE, Flowers N, Hutfless S, Abramson O, Perry GS, Herrinton LJ. Prevalence of colorectal cancer surveillance for ulcerative colitis in an integrated health care delivery system. Gastroenterology. 2010; 139:1511–8. [PubMed: 20659470]

66. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol. 2011; 173:1059–68. [PubMed: 21430193]

67. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, Pankratz JJ, Brue SM, Rocca WA. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. Int J Epidemiol. 2012; 41:1614–24. [PubMed: 23159830]

68. St Sauver JL, Grossardt BR, Leibson CL, Yawn BP, Melton LJ 3rd, Rocca WA. Generalizability of epidemiological findings and public health decisions: an illustration from the Rochester Epidemiology Project. Mayo Clin Proc. 2012; 87:151–60. [PubMed: 22305027]

69. Rocca WA, Yawn BP, St Sauver JL, Grossardt BR, Melton LJ 3rd. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population. Mayo Clin Proc. 2012; 87:1202–13. [PubMed: 23199802]

70. Loftus EV Jr. Silverstein MD, Sandborn WJ, Tremaine WJ, Harmsen WS, Zinsmeister AR. Crohn's disease in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. Gastroenterology. 1998; 114:1161–8. [PubMed: 9609752]

71. Loftus EV Jr. Silverstein MD, Sandborn WJ, Tremaine WJ, Harmsen WS, Zinsmeister AR. Ulcerative colitis in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. Gut. 2000; 46:336–43. [PubMed: 10673294]

72. Silverstein MD, Loftus EV, Sandborn WJ, Tremaine WJ, Feagan BG, Nietert PJ, Harmsen WS, Zinsmeister AR. Clinical course and costs of care for Crohn's disease: Markov model analysis of a population-based cohort. Gastroenterology. 1999; 117:49–57. [PubMed: 10381909]

73. Jess T, Loftus EV Jr. Velayos FS, Harmsen WS, Zinsmeister AR, Smyrk TC, Schleck CD, Tremaine WJ, Melton LJ 3rd, Munkholm P, Sandborn WJ. Risk of intestinal cancer in inflammatory bowel disease: a population-based study from olmsted county, Minnesota. Gastroenterology. 2006; 130:1039–46. [PubMed: 16618397]

74. Schwartz DA, Loftus EV Jr. Tremaine WJ, Panaccione R, Harmsen WS, Zinsmeister AR, Sandborn WJ. The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. Gastroenterology. 2002; 122:875–80. [PubMed: 11910338]

75. Loftus EV Jr. Crowson CS, Sandborn WJ, Tremaine WJ, O'Fallon WM, Melton LJ 3rd. Long-term fracture risk in patients with Crohn's disease: a population-based study in Olmsted County, Minnesota. Gastroenterology. 2002; 123:468–75. [PubMed: 12145800]

76. Loftus EV Jr. Achenbach SJ, Sandborn WJ, Tremaine WJ, Oberg AL, Melton LJ 3rd. Risk of fracture in ulcerative colitis: a population-based study from Olmsted County, Minnesota. Clin Gastroenterol Hepatol. 2003; 1:465–73. [PubMed: 15017646]

77. Faubion WA Jr. Loftus EV Jr. Harmsen WS, Zinsmeister AR, Sandborn WJ. The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology. 2001; 121:255–60. [PubMed: 11487534]

78. Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV Jr. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. Gastroenterology. 2010; 139:1147–55. [PubMed: 20637205]

79. Peyrin-Biroulet L, Harmsen WS, Tremaine WJ, Zinsmeister AR, Sandborn WJ, Loftus EV Jr. Surgery in a population-based cohort of Crohn's disease from Olmsted County, Minnesota (1970-2004). Am J Gastroenterol. 2012; 107:1693–701. [PubMed: 22945286]

80. Jess T, Loftus EV Jr. Harmsen WS, Zinsmeister AR, Tremaine WJ, Melton LJ 3rd, Munkholm P, Sandborn WJ. Survival and cause specific mortality in patients with inflammatory bowel disease: a long term outcome study in Olmsted County, Minnesota, 1940-2004. Gut. 2006; 55:1248–54. [PubMed: 16423890]

81. Ananthakrishnan AN, Higuchi LM, Huang ES, Khalili H, Richter JM, Fuchs CS, Chan AT. Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. Ann Intern Med. 2012; 156:350–9. [PubMed: 22393130]

82. Ananthakrishnan AN, Khalili H, Higuchi LM, Bao Y, Korzenik JR, Giovannucci EL, Richter JM, Fuchs CS, Chan AT. Higher predicted vitamin D status is associated with reduced risk of Crohn's disease. Gastroenterology. 2012; 142:482–9. [PubMed: 22155183]

83. Ananthakrishnan AN, Khalili H, Pan A, Higuchi LM, de Silva P, Richter JM, Fuchs CS, Chan AT. Association between depressive symptoms and incidence of Crohn's disease and ulcerative colitis: results from the Nurses' Health Study. Clin Gastroenterol Hepatol. 2013; 11:57–62. [PubMed: 22944733]

84. Khalili H, Ananthakrishnan AN, Higuchi LM, Richter JM, Fuchs CS, Chan AT. Early life factors and risk of inflammatory bowel disease in adulthood. Inflamm Bowel Dis. 2013; 19:542–7. [PubMed: 23429446]

85. Khalili H, Higuchi LM, Ananthakrishnan AN, Richter JM, Feskanich D, Fuchs CS, Chan AT. Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. Gut. 2012

86. Thirumurthi S, Chowdhury R, Richardson P, Abraham NS. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. Dig Dis Sci. 2010; 55:2592–8. [PubMed: 20033847]

87. Bansal P, Sonnenberg A. Risk factors of colorectal cancer in inflammatory bowel disease. Am J Gastroenterol. 1996; 91:44–8. [PubMed: 8561142]

88. Delco F, Sonnenberg A. Military history of patients with inflammatory bowel disease: an epidemiological study among U.S. veterans. Am J Gastroenterol. 1998; 93:1457–62. [PubMed: 9732925]

89. Hou JK, Chang M, Nguyen T, Kramer JR, Richardson P, Sansgiry S, D'Avolio LW, El-Serag HB. Automated identification of surveillance colonoscopy in inflammatory bowel disease using natural language processing. Dig Dis Sci. 2013; 58:936–41. [PubMed: 23086115]

90. Hou JK, Kramer JR, Richardson P, Sansgiry S, El-Serag HB. Myelosuppression monitoring after immunomodulator initiation in veterans with inflammatory bowel disease: a national practice audit. Aliment Pharmacol Ther. 2012; 36:1049–56. [PubMed: 23061548]

91. Khan N, Abbas AM, Bazzano LA, Koleva YN, Krousel-Wood M. Long-term oral mesalazine adherence and the risk of disease flare in ulcerative colitis: nationwide 10-year retrospective cohort from the veterans affairs healthcare system. Aliment Pharmacol Ther. 2012; 36:755–64. [PubMed: 22882428]

92. Khan N, Abbas AM, Moehlen M, Balart L. Methotrexate in Ulcerative Colitis: A Nationwide Retrospective Cohort from the Veterans Affairs Health Care System. Inflamm Bowel Dis. 2013

93. Sonnenberg A, Richardson PA, Abraham NS. Hospitalizations for inflammatory bowel disease among US military veterans 1975-2006. Dig Dis Sci. 2009; 54:1740–5. [PubMed: 19259815]

94. Khan N, Abbas AM, Lichtenstein GR, Loftus EV Jr. Bazzano LA. Risk of Lymphoma in Patients with Ulcerative Colitis Treated with Thiopurines-a Nationwide Retrospective Cohort Study. Gastroenterology. 2013

95. Khan N, Abbas AM, Almukhtar RM, Khan A. Prevalence and predictors of low bone mineral density in males with ulcerative colitis. J Clin Endocrinol Metab. 2013; 98:2368–75. [PubMed: 23596137]

96. Khan N, Abbas A, Williamson A, Balart L. Prevalence of Corticosteroids Use and Disease Course After Initial Steroid Exposure in Ulcerative Colitis. Dig Dis Sci. 2013

97. Deepak P, Sifuentes H, Sherid M, Stobaugh D, Sadozai Y, Ehrenpreis ED. T-cell non-Hodgkin's lymphomas reported to the FDA AERS with tumor necrosis factor-alpha (TNF-alpha) inhibitors: results of the REFURBISH study. Am J Gastroenterol. 2013; 108:99–105. [PubMed: 23032984]

98. Chiorean MV, Pokhrel B, Adabala J, Helper DJ, Johnson CS, Juliar B. Incidence and risk factors for lymphoma in a single-center inflammatory bowel disease population. Dig Dis Sci. 2011; 56:1489–95. [PubMed: 20953706]

99. Siegel CA, Marden SM, Persing SM, Larson RJ, Sands BE. Risk of lymphoma associated with combination anti-tumor necrosis factor and immunomodulator therapy for the treatment of Crohn's disease: a meta-analysis. Clin Gastroenterol Hepatol. 2009; 7:874–81. [PubMed: 19558997]

100. Sultan K, Korelitz BI, Present D, Katz S, Sunday S, Shapira I. Prognosis of lymphoma in patients following treatment with 6-mercaptopurine/azathioprine for inflammatory bowel disease. Inflamm Bowel Dis. 2012; 18:1855–8. [PubMed: 22241664]

101. Gearhart SL, Nathan H, Pawlik TM, Wick E, Efron J, Shore AD. Outcomes from IBD-associated and non-IBD-associated colorectal cancer: a Surveillance Epidemiology and End Results Medicare study. Dis Colon Rectum. 2012; 55:270–7. [PubMed: 22469793]

102. Shaukat A, Salfiti NI, Virnig DJ, Howard DH, Sitaraman SV, Liff JM, Lederle FA. Is ulcerative colitis associated with survival among older persons with colorectal cancer in the US? A population-based case-control study. Dig Dis Sci. 2012; 57:1647–51. [PubMed: 22113428]

103. Wang YR, Cangemi JR, Loftus EV Jr. Picco MF. Rate of early/missed colorectal cancers after colonoscopy in older patients with or without inflammatory bowel disease in the United States. Am J Gastroenterol. 2013; 108:444–9. [PubMed: 23295277]

104. Long MD, Kappelman MD, Martin CF, Lewis JD, Mayer L, Kinneer PM, Sandler RS. Development of an internet-based cohort of patients with inflammatory bowel diseases (CCFA Partners): methodology and initial results. Inflamm Bowel Dis. 2012; 18:2099–106. [PubMed: 22287300]

105. Cohen AB, Lee D, Long MD, Kappelman MD, Martin CF, Sandler RS, Lewis JD. Dietary Patterns and Self-Reported Associations of Diet with Symptoms of Inflammatory Bowel Disease. Dig Dis Sci. 2012

106. Ananthakrishnan AN, Long MD, Martin CF, Sandler RS, Kappelman MD. Sleep Disturbance and Risk of Active Disease in Patients with Crohn's Disease and Ulcerative Colitis. Clin Gastroenterol Hepatol. 2013

107. Atreja A, Achkar JP, Jain AK, Harris CM, Lashner BA. Using technology to promote gastrointestinal outcomes research: a case for electronic health records. Am J Gastroenterol. 2008; 103:2171–8. [PubMed: 18844611]

108. Reddy SI, Friedman S, Telford JJ, Strate L, Ookubo R, Banks PA. Are patients with inflammatory bowel disease receiving optimal care? Am J Gastroenterol. 2005; 100:1357–61. [PubMed: 15929770]

109. Kane S, Huo D, Aikens J, Hanauer S. Medication nonadherence and the outcomes of patients with quiescent ulcerative colitis. Am J Med. 2003; 114:39–43. [PubMed: 12543288]

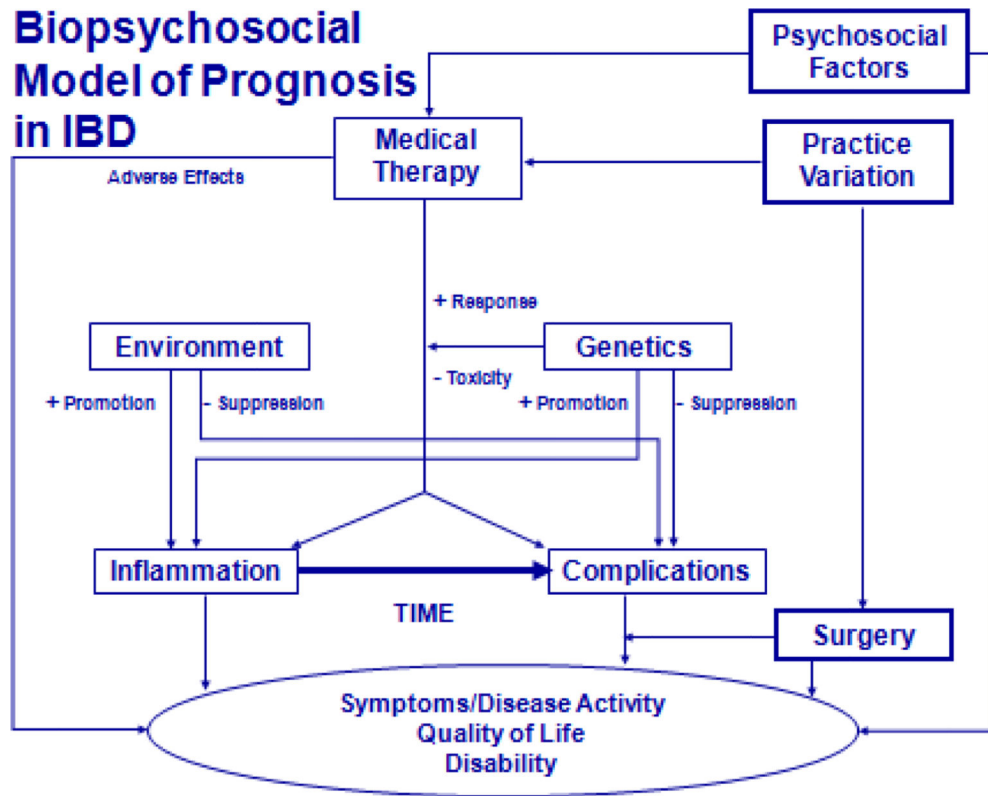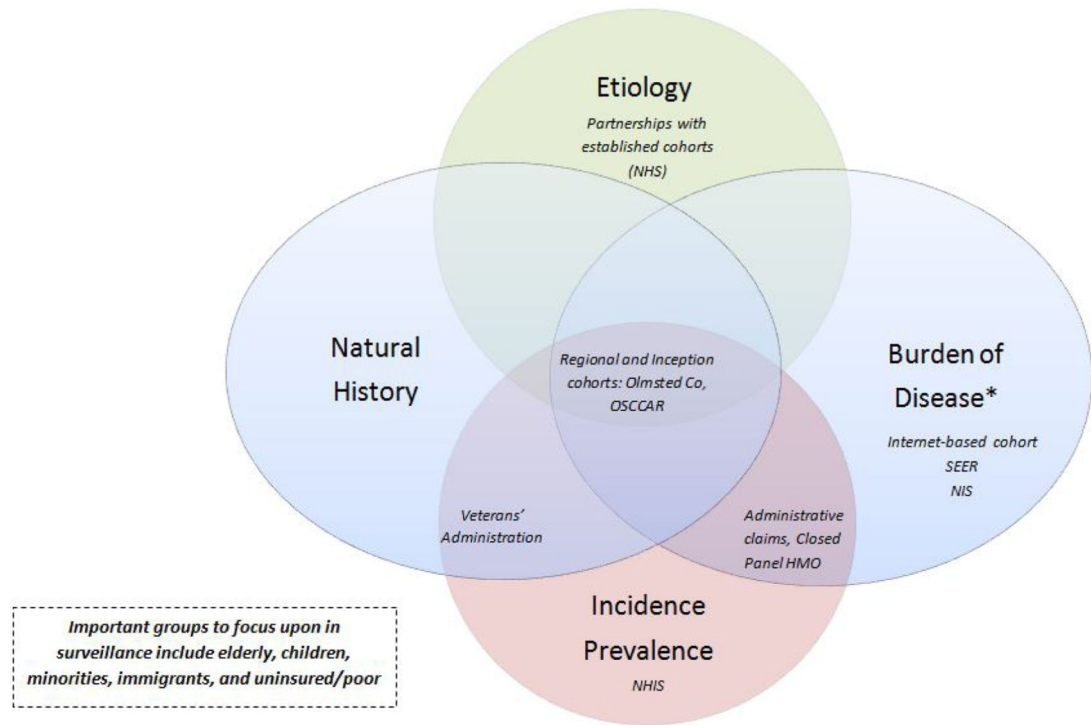**Figure 1.**
Biopsychosocial model of inflammatory bowel disease prognosis

*Measurements could include costs, loss of work productivity, other morbidity or disease-related malignancy

Abbreviations: NHS: Nurses' Health Study, OSCCAR: Ocean State Crohn's and Coltis Area Registry, HMO: Health Maintenance Organization, NHIS: National Health Interview Survey, NIS: Nationwide Inpatient Sample, SEER: Surveillance Epidemiology and End Results

Figure 2.
Important Components of an Inflammatory Bowel Disease Surveillance System in the United States

Table 1

Underrepresented groups in epidemiological studies of inflammatory bowel disease in the United States

• African Americans
• Native Americans
• Asian Americans
• Hispanic ethnicity
• 1st and 2nd generation immigrants
• Children
• Elderly
• Residents of southeastern United States