



Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease

Julien Kirchgesner, Rishi Desai, Laurent Beaugerie, Seoyoung Kim, Sebastian Schneeweiss

► To cite this version:

Julien Kirchgesner, Rishi Desai, Laurent Beaugerie, Seoyoung Kim, Sebastian Schneeweiss. Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease. *Clinical Pharmacology and Therapeutics*, 2022, 111 (1), pp.179-186. 10.1002/cpt.2304 . hal-03657125

HAL Id: hal-03657125

<https://hal.sorbonne-universite.fr/hal-03657125>

Submitted on 2 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Calibrating real-world evidence studies against randomized trials: treatment effectiveness of infliximab in Crohn's disease

Julien Kirchgesner, MD, PhD ^{1,2}, Rishi J. Desai, PhD ¹, Laurent Beaugerie, MD, PhD ², Seoyoung C. Kim, MD, ScD ^{1,3}, Sebastian Schneeweiss, MD, ScD ¹

(1) Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

(2) Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpital Saint-Antoine, Department of gastroenterology, Paris, France

(3) Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

Corresponding author:

Julien Kirchgesner

Division of Pharmacoepidemiology and Pharmacoeconomics,
Department of Medicine, Brigham and Women's Hospital and Harvard Medical School
1620 Tremont Street, Suite 3030, Boston, MA 02120

Phone: 617-278-0930

Fax: 617-232-8602

E-mail: jkirchgesner@bwh.harvard.edu

Body text word count: 2807 words

Conflict of Interest Disclosures:

Declaration of personal interests: The authors disclose the following: Julien Kirchgesner received research support from the French National Society of Gastroenterology (SNFGE). Rishi J. Desai has received research support to the Brigham and Women's Hospital from Novartis, Vertex, and Bayer for unrelated studies. Laurent Beaugerie has received consulting fees from Janssen, Pfizer and Allergan, lecture fees from AbbVie, Janssen, MSD, Ferring Pharmaceuticals, Mayoly-Spendler, Takeda, and Tillots, and research support from Abbott, Ferring Pharmaceuticals, Hospira-Pfizer, Janssen, MSD, Takeda, and Tillots for unrelated studies. Sebastian Schneeweiss is participating in investigator-initiated grants to the Brigham and Women's Hospital from Bayer, Vertex, and Boehringer Ingelheim unrelated to the topic of this study. He is a consultant to Aetion Inc., a software manufacturer of which he owns equity. Seoyoung C. Kim has received research support to the Brigham and Women's Hospital from Pfizer, Roche, AbbVie, and Bristol-Myers Squibb for unrelated studies.

Funding information:

The investigators conducted the research independently through funding from internal sources of the Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School.

Keywords:

Pharmacoepidemiology; gastrointestinal; biologics; immunosuppressants.

Abstract

Real-world evidence (RWE) on the effectiveness of treatments in Crohn's disease (CD) derived from clinical practice data will help fill many evidence gaps left by randomized controlled trials (RCTs). Emulating RCTs with healthcare database studies may calibrate RWE studies in CD.

We aimed to emulate the SONIC trial on the effectiveness of infliximab in patients with CD using U.S. and French healthcare claims data. SONIC had shown improved remission with combination therapy (i.e., infliximab plus thiopurines) compared to infliximab monotherapy.

Using claims data (2004-2019) from commercially insured patients in the U.S. (IBM 'MarketScan' and Optum) and France (SNDS), we conducted a cohort study of patients with CD who initiated combination therapy and compared them with patients who initiated infliximab alone. The primary outcome was a composite endpoint of treatment failure including hospitalization or surgery related to CD, treatment switch, or continuation of corticosteroids 26 weeks after infliximab initiation. Risk ratios (RRs) with 95% confidence intervals (CIs) were estimated in propensity score (PS) matched cohorts. We identified 1,437 PS-matched pairs of combination therapy versus infliximab monotherapy users. Like in SONIC, the risk of treatment failure was decreased with combination therapy in the overall cohort (RR, 0.71; 95%CI 0.62–0.82; RR, 0.78; 95%CI 0.62-0.97 in SONIC). Findings were consistent across MarketScan, Optum, and SNDS databases: RR (95%CI), 0.83 (0.63–1.10), 0.66 (0.46–0.93), and 0.68 (0.57–0.82), as well as component endpoints.

These robust findings highlight opportunities of RWE analysis to study treatment effectiveness in patients with CD in clinical practice.

Introduction

Real world evidence (RWE) has been used for decades to inform decisions about treatment safety, while it has been used only in limited circumstances to inform regulatory decisions about effectiveness complementing randomized controlled trials (RCTs).¹⁻⁴ Following the mandate of the 21st Century Cures Act of 2016, FDA proposed a framework to conduct studies from data generated from the routine operation of the healthcare system, or real world data (RWD), for effectiveness regulatory decision.⁵ The framework acknowledged challenges of RWE including that clinically meaningful measures of effectiveness outcomes may not be consistently captured in electronic health records or medical claims data. The emulation of RCTs by RWE studies may help to calibrate effectiveness outcome measurable in RWD, that could be applied to other treatment comparison within the same indication. Additionally, calibration of RWE against RCTs faces the challenge of interpreting differences between treatment effect estimates from the two study types, that can be driven by residual bias in RWE but also emulation differences.^{6,7}

The therapeutic armamentarium in immune-mediated inflammatory diseases, notably Crohn's disease (CD), dramatically increased in the last decade, highlighting the need of RWE to complement RCT evidence. As in many chronic conditions, evidence of head-to-head comparisons, add-on therapies, and various ways to stage treatments are lacking and it is unlikely that we will have RCT evidence for all these clinically important questions any time soon. The optimal position of each drug remains largely unknown,⁸ while RWE could provide head-to-head comparison; however, for RWE to be actionable we need to gain confidence that such studies can measure clinically meaningful endpoints reliably and come to causal conclusions on the treatment effectiveness.

The aim of this study was to emulate an RCT, the Study of Biologic and Immunomodulator Naive Patients in Crohn's Disease (SONIC) trial, studying the effectiveness of a biologic agent in patients with CD using data from two large U.S. and a French nationwide healthcare

claims database. The SONIC trial was a parallel-group RCT assessing treatment failure in patients treated with infliximab and azathioprine combined compared to patients treated with either infliximab or thiopurines alone for CD.⁹

Methods

Data source

This study was conducted by using two U.S. health care claims databases, IBM MarketScan (MarketScan) 2004-2018 and Optum Clinformatics (Optum) 2005-2019, and the French administrative health database 2009-2018 (*Système National des Données de Santé*, SNDS). Patients enrolled in the MarketScan and Optum databases are representative of a commercially insured population in the U.S.; the SNDS insures 95% of the French population. The two U.S. databases are de-identified and contain demographic data and longitudinal information on all encounters with the professional healthcare system, including outpatient visits, ER visits, and hospitalizations with diagnosis and procedure information, as well as pharmacy drug dispensing. Similarly, the SNDS contains data on all drug reimbursements, inpatient and outpatient medical care prescribed or provided by health-care professionals.¹⁰ It does not provide any information on the medical indication for each reimbursement, but contains the patient's status with respect to full reimbursement of care for long-term diseases (LTDs), which includes CD and allows to assess the date of CD diagnosis.¹¹ The study was approved by the institutional review board of the Brigham and Women's Hospital, and the French Data Protection Authority.

Study population

Infliximab is a tumor necrosis factor antagonist (anti-TNF) and was approved for the treatment of CD in 1998 in the US and in 1999 in Europe,¹² while thiopurines have been used for the treatment of CD since the nineteen seventies.¹³ The SONIC trial was an RCT conducted between 2005 and 2008 aimed to assess the effectiveness of infliximab and thiopurines combined (i.e., the combination therapy) compared to either infliximab or thiopurines alone in patients with CD naïve to immunosuppressants or anti-TNF. It concluded that the combination therapy of infliximab and thiopurines was superior to treatment with its component alone.

The emulation study's inclusion/exclusion criteria were tailored to replicate SONIC whenever possible.⁹ The SONIC trial included adult patients with moderate to severe CD with a Crohn's Disease Activity Index (CDAI) of 220 to 450 points, despite corticosteroids and/or aminosalicylates. Patients were naïve to thiopurines, methotrexate, and anti-TNF. To emulate this design, we identified all patients aged 18 years or older initiating infliximab and with at least one visit for CD using the International Classification of Diseases 9th (ICD-9) or 10th (ICD-10). In the SNDS database, CD diagnosis was based on previous published algorithms,^{11,14–16} and the date of CD diagnosis was defined as the earliest diagnosis date either from hospital discharge diagnosis or from LTD diagnosis. Cohort entry date was defined as the date of infliximab initiation. Patients in MarketScan and Optum databases were required to have continuous enrolment during the baseline period of 180 days before cohort entry date. Patients were followed from the day after cohort entry date to 180 days after cohort entry date (week 26).

In analogy to SONIC, we excluded patients previously exposed to methotrexate, and to any anti-TNF agents. Treatment with thiopurines was only allowed in the month prior infliximab initiation, and patients exposed to thiopurines more than one month before infliximab initiation were excluded. Patients with an ostomy, stricture, abscess, abdominal surgery, tuberculosis, opportunistic infections, hepatitis B and C within the previous 6 months were excluded. We also excluded patients with HIV infection, multiple sclerosis, and previous history of cancer (excluding non-melanoma skin cancer). Additional exclusion criteria compared to the SONIC trial were considered. First, patients with a diagnosis code related to ulcerative colitis were excluded in order to only include patients with CD. Second, patients with a diagnosis of rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and psoriasis were excluded in order to only include patients among whom anti-TNF was prescribed for CD. Lastly, exposure to biologics and immunosuppressant that were available after 2010 were considered as exclusion criteria, namely, natalizumab, vedolizumab, ustekinumab, and tofacitinib.

Treatment groups

Two treatment groups were considered. Since thiopurines may be started before the first infliximab infusion and thiopurines may take up to 3 months to be effective as monotherapy, combination therapy was defined as starting thiopurines within 30 days before infliximab initiation. Infliximab monotherapy was defined as infliximab initiation without thiopurines exposure any time prior. We did not consider patients only exposed to thiopurines monotherapy, since recent guidelines recommend against its use as monotherapy for the induction of remission in patients with CD.⁸

Outcome

In the SONIC trial, the primary end point was corticosteroid-free clinical remission at week 26, defined by a CDAI less than 150 points without any systemic corticosteroid use. Since the CDAI is not available in the US or French healthcare databases, we developed a composite effectiveness outcome measure based on three surrogate endpoints for treatment failure: (1) hospitalization or surgery related to CD; (2) treatment switch to another biologics (adalimumab, certolizumab pegol, golimumab, natalizumab, vedolizumab, ustekinumab) or small molecules (tofacitinib); or (3) exposure to systemic corticosteroids at week 26. Related codes are summarized in Table S1.

Follow-up started the day after cohort entry and ended on the earliest occurrence of 180 days of follow-up, outcome occurrence, or death, for all subjects started on either one of the treatments.

Patient characteristics

Baseline patient characteristics and markers of CD phenotype and severity were considered, including demographics and comorbidities (previous serious infections, *Clostridioides difficile* infection, cardiovascular disease, chronic kidney failure, chronic liver disease, chronic pulmonary disease, venous thromboembolism, and diabetes). Comorbidities were assessed in the 180 days before cohort entry in MarketScan and Optum and in all data available in the

SNDS. CD phenotype included Montreal phenotype,¹⁷ including inflammatory (B1), stricturing (B2), and penetrating phenotype (B3), as well as perianal involvement. Hospitalization for complicated CD course (stricture or abscess) and surgery related to CD occurring more than 180 days before cohort entry were also assessed. CD severity and healthcare use intensity were assessed in the 180 days before cohort entry in all databases. CD severity was assessed by corticosteroids and aminosalicylates exposure, occurrence of CD-related hospitalization or surgery, abdominal imaging, gastrointestinal endoscopy, fecal pathogen, and the number of C-reactive protein tests ordered. Healthcare use was assessed by hospitalizations not related to CD and the number of gastroenterologist visits.

Statistical analysis

To control for confounding, we calculated a propensity score (PS) for each patient predicting the probability of initiating combination therapy with infliximab and thiopurines versus infliximab monotherapy with logistic regression, including all measured baseline covariates without further variable selection and the year of cohort entry.¹⁸ We matched treatment groups 1:1 on their PS using optimal-matching within a caliper of 0.02.¹⁹ After matching, standardized differences matching were calculated to assess balance between patients exposed to combination therapy and their controls exposed to infliximab monotherapy.²⁰ After PS matching we estimated the 180 day risk of the outcomes in percent including 95% confidence intervals (CIs). Log-binomial regression models were used to estimate adjusted risk ratios (RR) with their 95% CIs comparing the risk of treatment failure associated with combination therapy versus infliximab monotherapy.²¹ Cohort-specific RRs were combined by an inverse variance-weighted, fixed-effects model.

Additional prespecified analyses included secondary analyses assessing the risk of each individual component of the composite endpoint, and subgroups analyses stratified on Montreal phenotype (B1 and B2-B3 combined), the presence of perianal involvement, and exposure to corticosteroids at infliximab initiation. Several sensitivity analyses were performed to test the robustness of our results. First, variables with an absolute standardized

difference greater than 0.1 after PS-matching were included as covariate in log-binomial regression models. Second, cox proportional hazards regression models were used to estimate hazard ratios (HR). Lastly, we performed a PS matched analysis with a variable ratio up to 1:4 with a caliper of 0.02.

Analyses were performed using the validated Aetion Evidence Platform (V 4.10) using R²²⁻²⁴ and SAS (version 9.4) statistical software (SAS Institute).

Results

Patient characteristics

A total of 43 844 patients with CD who initiated infliximab were identified across the three databases. After applying exclusion criteria (Figure 1), 15 007 immunosuppressants and anti-TNF naive patients with CD were included, initiating either combination therapy with infliximab and thiopurines (n=1452) or infliximab monotherapy (n=13 555).

More than 70% of patients had an inflammatory phenotype according to Montreal classification (B1). In the SNDS cohort, the mean (SD) duration of CD was 3.9 (6.0) and 2.9 (5.4) years in the combination therapy and infliximab monotherapy groups, respectively (Table S2). After PS-matching, the cohorts contained 1437 pairs, with mean age between 33 and 38 years, which was similar to SONIC.⁹ Conversely, around half of the patients were treated with corticosteroids at infliximab initiation, compared to 27.4% of patients in SONIC (Table 1). Only two covariates had an absolute standardized difference greater than 0.1 (Figure S1).

Risk of treatment failure

After PS-matching, treatment failure occurred in 263 (18.3%; 95% CI 16.2-20.2) and 369 (25.7%; 95% CI 24.6-30.1) patients initiating combination therapy and infliximab monotherapy, respectively (Table 2). Among patients with treatment failure, hospitalization or surgery related to CD was the most frequent outcome, accounting for 59% and 63% of outcomes in patients initiating combination therapy and infliximab monotherapy. The proportion of patients with treatment failure according to databases before PS-matching is provided in Table S3.

Patients initiating combination therapy had a 29% decreased risk of treatment failure compared to patients initiating infliximab monotherapy (RR, 0.71, 95% CI 0.62-0.82). Results were similar across the three databases (RR, 0.83, 95% CI 0.63–1.10, RR, 0.66, 95% CI, 0.46–0.93, and RR, 0.68, 95% CI, 0.57–0.82 in MarketScan, Optum, and SNDS,

respectively). Risk ratios of treatment failure in the emulated cohort and in the SONIC trial are provided in Figure 2. The reduced risk of treatment failure with combination therapy was observed for all secondary endpoints, with a 25% (9-37%) reduced risk of hospitalization or surgery related to CD, a 44% (21-60%) reduced risk of treatment switch, and a 43% (25-57%) reduced risk of persistent corticosteroids exposure at week 26 (Figure 3).

Subgroup and sensitivity analyses

Overall, there was no statistically significant heterogeneity in the risk of treatment failure across subgroups (Figure 4 and Table S4). The combined RR for treatment failure associated with combination therapy versus infliximab monotherapy was 0.75 (95%CI, 0.64-0.89) and 0.86 (95%CI, 0.64-1.15) in patients with inflammatory phenotype and patients with stricturing or penetrating phenotypes, respectively. In patients treated with corticosteroids at infliximab initiation, the combined RR was 0.69 (95%CI, 0.58-0.83).

The various sensitivity analyses yielded consistent results (Table S5). The risk of treatment failure was notably decreased by 27% (16-37%) with combination therapy after propensity score matching with a 1:4 variable ratio.

Discussion

Based on three large population-based cohorts in the U.S. and France, we emulated a RCT assessing the effectiveness of combination therapy with infliximab and thiopurines compared to infliximab monotherapy in patients with CD. Our analysis replicated the findings of the SONIC trial by using a composite effectiveness outcome based on hospitalization or surgery related to CD, treatment switch, or corticosteroids continuation. Findings were consistent across databases and for each individual components of the composite outcome.

Calibration of RWE studies against treatment effect assessed in RCTs allows to evaluate whether RWE can support causal conclusions if conducted using robust methodology.⁷ Our study design was as close as possible to the design of the SONIC trial, including exclusion and inclusion criteria, and outcome definition. Our findings support the concept of using database analysis to provide evidence for comparisons that are not assessed in RCTs. In CD, all recent phase III RCTs compared the active treatment with placebo, whereas standard first line treatment is available and the European Medicines Agency recommended a direct comparison with current generally accepted standard first line treatment in this setting.²⁵ RWE may be used to assess effectiveness in head to head comparison in addition to safety studies.¹⁶

The main strength of our study is the generalizability and large size as we used two large U.S. and one French nationwide population-based cohorts, which allows to assess treatment effectiveness in different healthcare schemes and potential prescribing patterns. A multi-database study contributes to assess the impact of differences in population selection, data collection and follow-up between databases. While patients are followed from birth to emigration or death in the SNDS database, patients could have a shorter enrollment period to assess covariates due to health insurance enrollment changes in Optum and MarketScan. The U.S databases also only included commercially insured patients, while the SNDS database included all French residents, since universal healthcare is guaranteed for all

French residents. Results were consistent across databases, which suggest that these differences had a minimal impact on the treatment estimate observed. Additionally, the use of a new-user design, the inclusion of an active comparator, and the assessment of CD disease activity allowed us to rigorously control for confounding. Lastly, findings were consistent for each individual component of the composite outcome, which strengthens the robustness of the results.

Some limitations should be noted. The study population differs from the population included in the SONIC trial, notably regarding the exposure to corticosteroids at infliximab initiation. Higher rates of corticosteroids exposure reported in our study population may suggest that patients included in the SONIC trial had a less severe disease compared to real-life setting. It might be related to the selected profile of patients included in RCTs, with more than 60% of patients with CD followed in tertiary referral centers who would have not been eligible to participate in recent RCTs, mainly due to disease severity.²⁶ We also performed stratified analyses according to baseline disease severity (stricturing or penetrating phenotype, exposure to corticosteroid at cohort entry), and results were consistent with the main analysis. The CDAI score used in SONIC includes clinical and biological parameters that are not collected in claims data. We decided to define treatment failure based on hospitalization or surgery related to CD, treatment switch, and corticosteroids continuation at week 26. Our definition may select higher degree of disease activity, as we reported a lower rate of treatment failure compared to SONIC. Additionally, a CDAI score above 150, which defined treatment failure in SONIC, may be not specific to CD activity and has been reported in patients with irritable bowel syndrome.²⁷ Effectiveness outcome measure developed in real world data may be only appropriate in the setting where they were developed. Furthermore, as we did not assess effectiveness in non-users of immunosuppressive treatment,²⁸ this should be further assessed. However, add-on strategies with thiopurines are used in the therapeutic management of CD, notably with recently approved biologics such as vedolizumab and ustekinumab. The effectiveness outcome measure developed in this study

may be used in future studies assessing the impact of adding thiopurines with other treatment than infliximab.

In summary, this study based on three large population-based cohorts of patients with CD in both the U.S. and France provided evidence that RWE studies can contribute to the assessment of treatment effectiveness for CD in clinical practice. These findings support the concept of using RWE analysis to complement evidence gained from RCTs.

Study Highlights

What is the current knowledge on the topic?

Real world evidence (RWE) is increasingly used to assess treatment effectiveness, emulating randomized clinical trials (RCTs) with healthcare database studies may calibrate RWE studies.

What question did this study address?

We aimed to calibrate RWE in Crohn's disease (CD) by replicating a RCT (SONIC) using US and French administrative health databases. The SONIC trial assessed in 2010 the effectiveness of the combination of infliximab and thiopurines compared to infliximab monotherapy in patients with CD at week 26 (primary end point; CDAI score). Based on the inclusion and exclusion criteria of SONIC, we assessed the risk of treatment failure at week 26 in patient initiating combination therapy or infliximab monotherapy. The primary endpoint was a composite endpoint based on hospitalization or surgery related to CD, treatment switch, and corticosteroids exposure at week 26.

What does this study add to our knowledge?

Our comparison replicated the effectiveness of combination therapy with infliximab and thiopurines in SONIC. Results were consistent across the three databases and for all individual components of the composite outcome.

How might this change clinical pharmacology or translational science?

Calibrating RWE against RCT using a robust methodology and multiple databases supports the use of RWE to assess effectiveness in real life setting.

Author Contributions:

A list of each author's contributions. The standard contributions include: J.K. and S.S. wrote manuscript, J.K., R.J.D., L.B., S.C.K., and S.S. designed research, J.K. and S.S. performed research, JK. analyzed data.

References

1. Schneeweiss, S. & Avorn, J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* **58**, 323–337 (2005).
2. Makady, A. *et al.* Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. *Value Health* **20**, 520–532 (2017).
3. Franklin, J. M., Glynn, R. J., Martin, D. & Schneeweiss, S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin. Pharmacol. Ther.* **105**, 867–877 (2019).
4. Schneeweiss, S. Real-World Evidence of Treatment Effects: The Useful and the Misleading. *Clin. Pharmacol. Ther.* **106**, 43–44 (2019).
5. US Food and Drug Administration. Framework for FDA's real- world evidence program. (2018).at <<https://www.fda.gov/media/120060/download>>
6. Franklin, J. M., Glynn, R. J., Suissa, S. & Schneeweiss, S. Emulation Differences vs. Biases When Calibrating Real-World Evidence Findings Against Randomized Controlled Trials. *Clin. Pharmacol. Ther.* (2020).doi:10.1002/cpt.1793
7. Franklin, J. M. *et al.* Emulating Randomized Clinical Trials with Nonrandomized Real-World Evidence Studies: First Results from the RCT DUPLICATE Initiative. *Circulation* (2020).doi:10.1161/CIRCULATIONAHA.120.051718
8. Torres, J. *et al.* ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment. *J. Crohns Colitis* **14**, 4–22 (2020).
9. Colombel, J. F. *et al.* Infliximab, azathioprine, or combination therapy for Crohn's disease. *N. Engl. J. Med.* **362**, 1383–1395 (2010).
10. Bezin, J. *et al.* The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol. Drug Saf.* **26**, 954–962 (2017).

11. Kirchgesner, J. *et al.* Therapeutic management of inflammatory bowel disease in real-life practice in the current era of anti-TNF agents: analysis of the French administrative health databases 2009-2014. *Aliment. Pharmacol. Ther.* **45**, 37–49 (2017).
12. Hanauer, S. B. *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *The Lancet* **359**, 1541–1549 (2002).
13. Summers, R. W. *et al.* National Cooperative Crohn's Disease Study: results of drug treatment. *Gastroenterology* **77**, 847–869 (1979).
14. Lemaitre, M. *et al.* Association Between Use of Thiopurines or Tumor Necrosis Factor Antagonists Alone or in Combination and Risk of Lymphoma in Patients With Inflammatory Bowel Disease. *JAMA* **318**, 1679–1686 (2017).
15. Kirchgesner, J. *et al.* Risk of acute arterial events associated with treatment of inflammatory bowel diseases: nationwide French cohort study. *Gut* **69**, 852–858 (2020).
16. Kirchgesner, J., Desai, R. J., Beaugerie, L., Schneeweiss, S. & Kim, S. C. Risk of serious infections with vedolizumab versus tumor necrosis factor antagonists in patients with inflammatory bowel disease. *Clin. Gastroenterol. Hepatol.* (2020).doi:10.1016/j.cgh.2020.12.030
17. Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J.-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **55**, 749–753 (2006).
18. Glynn, R. J., Schneeweiss, S. & Stürmer, T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin. Pharmacol. Toxicol.* **98**, 253–259 (2006).
19. Rassen, J. A. *et al.* One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol. Drug Saf.* **21**, 69–80 (2012).
20. Franklin, J. M., Rassen, J. A., Ackermann, D., Bartels, D. B. & Schneeweiss, S. Metrics for covariate balance in cohort studies of causal effects. *Stat. Med.* **33**, 1685–1699 (2014).

21. Robbins, A. S., Chao, S. Y. & Fonseca, V. P. What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Ann. Epidemiol.* **12**, 452–454 (2002).
22. Aetion Evidence Platform® (2020). Software for real-world data analysis. Aetion, Inc. *Aetion* at <<https://aetion.com/>>
23. Wang, S. V. *et al.* Transparency and Reproducibility of Observational Cohort Studies Using Large Healthcare Databases. *Clin. Pharmacol. Ther.* **99**, 325–332 (2016).
24. Patorno, E., Schneeweiss, S., Gopalakrishnan, C., Martin, D. & Franklin, J. M. Using Real-World Data to Predict Findings of an Ongoing Phase IV Cardiovascular Outcome Trial: Cardiovascular Safety of Linagliptin Versus Glimepiride. *Diabetes Care* **42**, 2204–2210 (2019).
25. European Medicines Agency. Guideline on the development of new medicinal products for the treatment of Crohn's Disease. (2019).at <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-crohns-disease-revision-2_en.pdf>
26. Ha, C., Ullman, T. A., Siegel, C. A. & Kornbluth, A. Patients Enrolled in Randomized Controlled Trials Do Not Represent the Inflammatory Bowel Disease Patient Population. *Clin. Gastroenterol. Hepatol.* **10**, 1002–1007 (2012).
27. C, L. *et al.* The Crohn's disease activity index (CDAI) is similarly elevated in patients with Crohn's disease and in patients with irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **37**, (2013).
28. Spoenclin, J. *et al.* Using Healthcare Databases to Replicate Trial Findings for Supplemental Indications: Adalimumab in Patients with Ulcerative Colitis. *Clin. Pharmacol. Ther.* (2020).doi:10.1002/cpt.1861

Figure Legends:

Figure 1. Study Population Flowchart

Figure 2. Risk ratios for treatment failure associated with combination therapy with infliximab and thiopurines compared to infliximab monotherapy

Figure 3. Risk ratios for each individual component of the composite outcome associated with combination therapy with infliximab and thiopurines compared to infliximab monotherapy (hospitalization or surgery related to CD [A], treatment switch [B], and corticosteroids continuation [C])

Figure 4. Risk ratios for treatment failure associated with combination therapy with infliximab and thiopurines compared to infliximab monotherapy: subgroup analysis

Table Legends:

Table 1. Characteristics of study patients, propensity score-matched with a 1:1 fixed ratio

Table 2. Proportion of patients with treatment failure after propensity score matching

Supplementary Material:

Table S1. Effectiveness outcome measure with related codes

Table S2. Characteristics of study patients treated with infliximab and thiopurines or infliximab monotherapy, before propensity score-matching

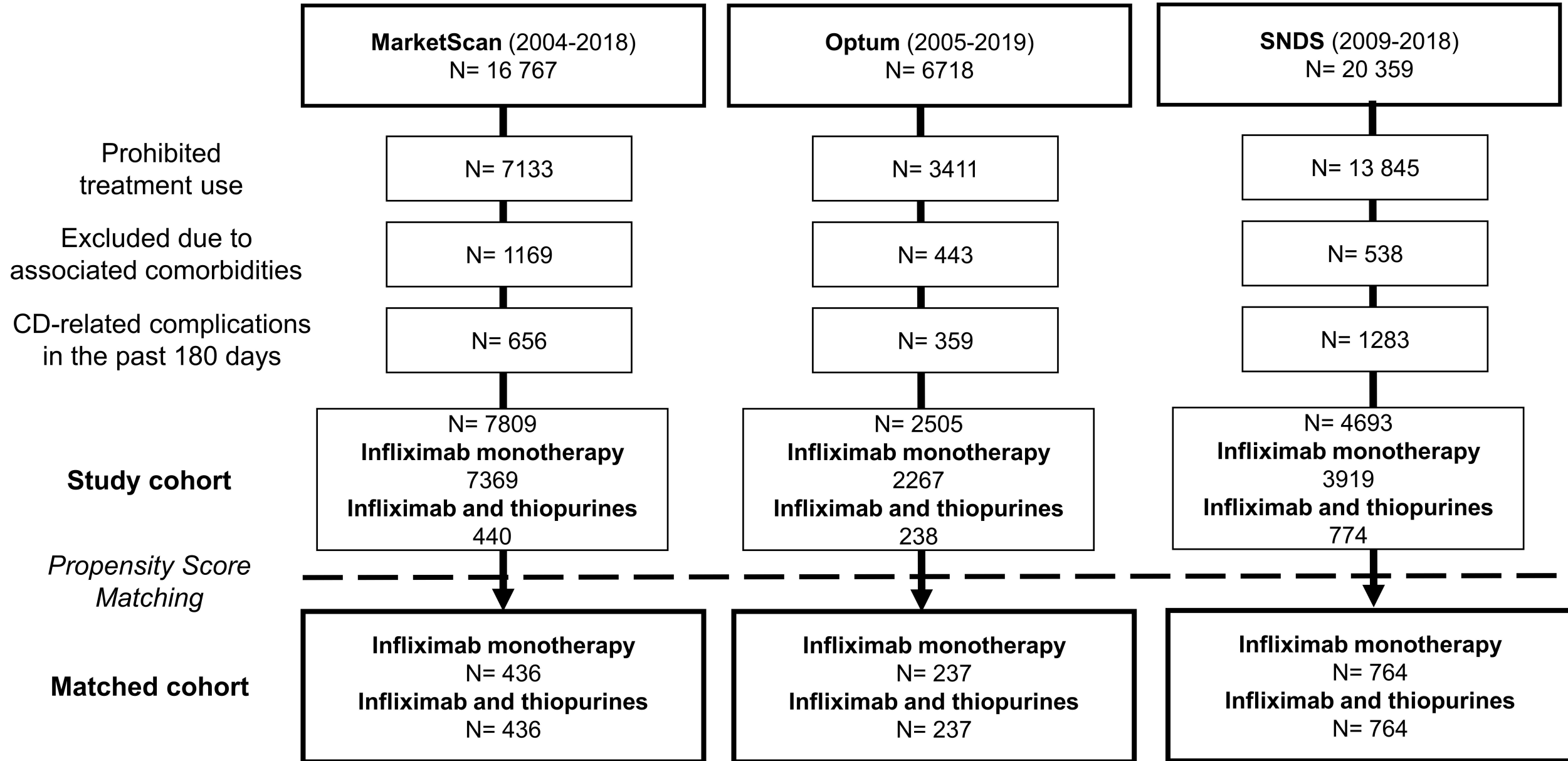
Table S3. Outcomes, before propensity-score matching

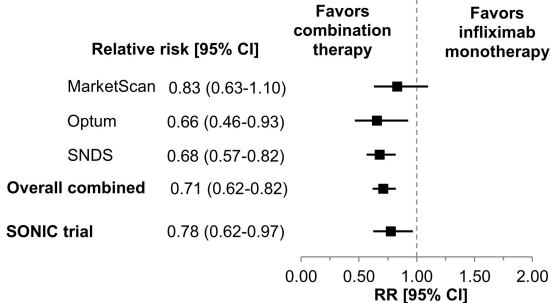
Table S4. Relative risks (95% confidence intervals) of treatment failure in patients treated with combination therapy versus infliximab monotherapy, subgroups analyses

Table S5. Relative risks (95% confidence intervals) of treatment failure in patients treated with combination therapy versus infliximab monotherapy, sensitivity analyses

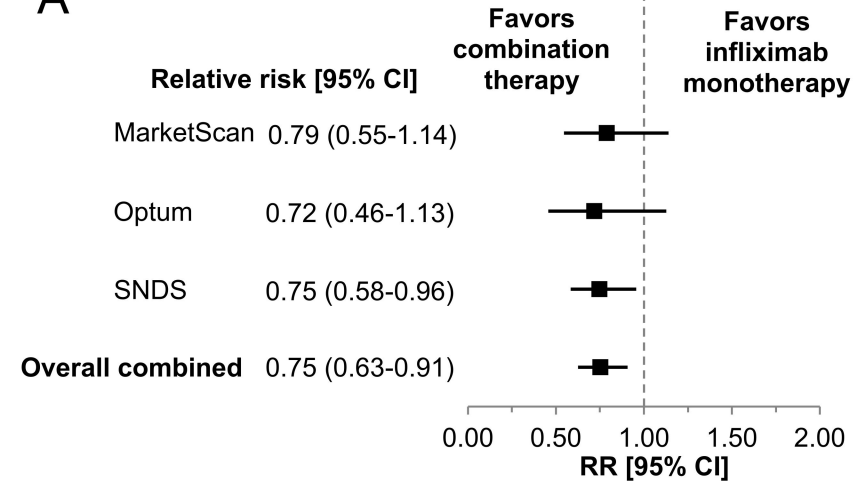
Figure S1. Standardized differences before and after propensity score matching (MarketScan [A], Optum[B], and SNDS [C])

Patients aged 18 years or older with Crohn's disease starting infliximab

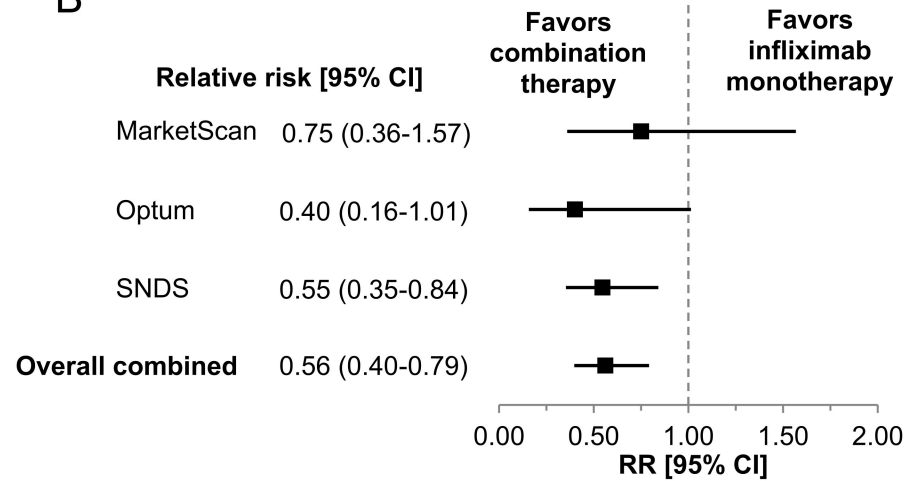




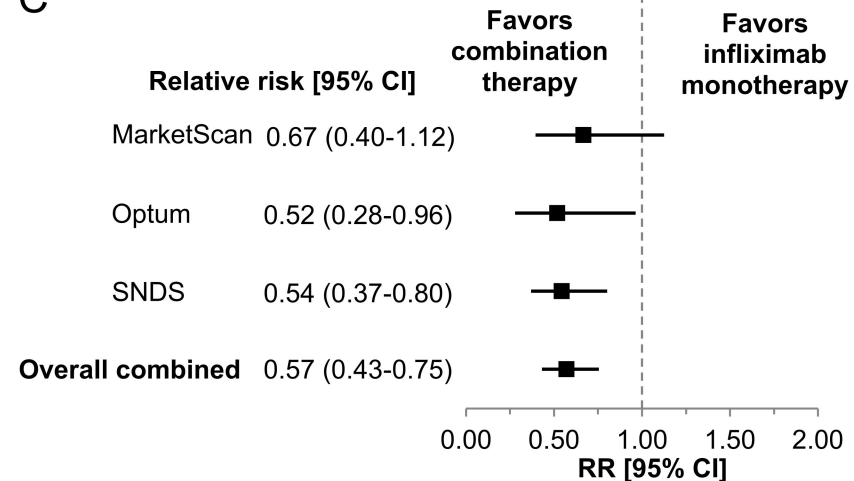
A



B



C



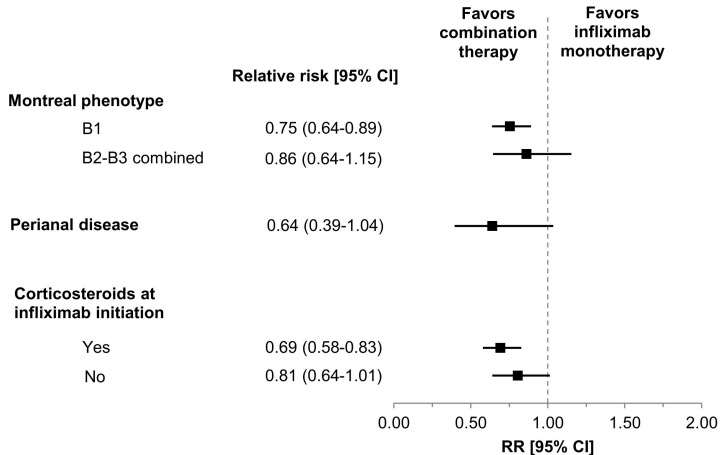


Table 1. Characteristics of study patients treated with infliximab and thiopurines or infliximab monotherapy, propensity score-matched with a 1:1 fixed ratio

Variable	MarketScan, n (%)		Optum, n (%)		SNDS, n (%)	
	Infliximab monotherapy (n=436)	Infliximab and thiopurines (n=436)	Infliximab monotherapy (n=237)	Infliximab and thiopurines (n=237)	Infliximab monotherapy (n=764)	Infliximab and thiopurines (n=764)
Age, mean (SD), years	38.2 (15.2)	37.6 (13.7)	36.4 (13.7)	36.4 (12.5)	33.6 (14.2)	33.0 (12.8)
Sex						
Male	246 (56.4)	249 (57.1)	131 (55.3)	126 (53.2)	312 (40.8)	313 (41.0)
Female	190 (43.6)	187 (42.9)	106 (44.7)	111 (46.8)	452 (59.2)	451 (59.0)
Crohn's disease duration, mean (SD), years	-	-	-	-	2.85 (4.82)	2.90 (5.42)
Montreal phenotype						
B1	328 (75.2)	323 (74.1)	164 (69.2)	166 (70.0)	660 (86.4)	655 (85.7)
B2	70 (16.1)	72 (16.5)	38 (16.0)	39 (16.5)	61 (8.0)	58 (7.6)
B3	38 (8.7)	41 (9.4)	35 (14.8)	32 (13.5)	43 (5.6)	51 (6.7)
Perianal Crohn's disease	76 (17.4)	72 (16.5)	38 (16.0)	42 (17.7)	102 (13.4)	96 (12.6)
Hospitalization for Crohn's disease complication more than 180 days before cohort entry	4 (0.9)	11 (2.5)	2 (0.8)	2 (0.8)	37 (4.8)	41 (5.4)
Surgery related to Crohn's disease more than 180 days before cohort entry	13 (3.0)	16 (3.7)	12 (5.1)	10 (4.2)	57 (7.5)	57 (7.5)
Corticosteroids (oral)						
within 180 days before cohort entry	299 (68.6)	298 (68.3)	161 (67.9)	154 (65.0)	478 (62.6)	473 (61.9)
at cohort entry	210 (48.2)	222 (50.9)	112 (47.3)	110 (46.4)	360 (47.1)	364 (47.6)
Aminosalicylates (oral) at cohort entry	241 (55.3)	249 (57.1)	115 (48.5)	115 (48.5)	380 (49.7)	387 (50.7)
Opioids ^a	221 (50.7)	223 (51.1)	110 (46.4)	106 (44.7)	121 (15.8)	117 (15.3)
Crohn's disease activity assessment ^a						
Hospitalization related to Crohn's disease	116 (26.6)	110 (25.2)	51 (21.5)	54 (22.8)	313 (41.0)	307 (40.2)
Abdominal imaging	228 (52.3)	222 (50.9)	134 (56.5)	138 (58.2)	564 (73.8)	541 (70.8)
Lower GI endoscopy	290 (66.5)	288 (66.1)	158 (66.7)	151 (63.7)	572 (74.9)	574 (75.1)
Upper GI endoscopy	107 (24.5)	101 (23.2)	60 (25.3)	55 (23.2)	389 (50.9)	392 (51.3)
CRP tests ordered, mean (SD)	0.82 (1.15)	0.75 (1.08)	1.21 (1.78)	1.06 (1.55)	2.76 (2.61)	2.86 (2.46)
Fecal pathogen tests ordered	87 (20.0)	79 (18.1)	55 (23.2)	57 (24.1)	153 (20.0)	166 (21.7)
Comorbidities						
<i>Clostridioides difficile</i> infection	6 (1.4)	9 (2.1)	3 (1.3)	4 (1.7)	8 (1.0)	9 (1.2)
Serious infection	4 (0.9)	5 (1.1)	1 (0.4)	2 (0.8)	7 (0.9)	8 (1.0)
Cardiovascular disease	33 (7.6)	30 (6.9)	19 (8.0)	15 (6.3)	29 (3.8)	29 (3.8)
Chronic kidney failure	8 (1.8)	5 (1.1)	6 (2.5)	4 (1.7)	3 (0.4)	2 (0.3)
Chronic liver disease	9 (2.1)	10 (2.3)	11 (4.6)	9 (3.8)	4 (0.5)	4 (0.5)
Chronic pulmonary disease	62 (14.2)	48 (11.0)	21 (8.9)	23 (9.7)	187 (24.5)	184 (24.1)
Venous thromboembolism	12 (2.8)	12 (2.8)	7 (3.0)	5 (2.1)	15 (2.0)	14 (1.8)
Diabetes	20 (4.6)	17 (3.9)	15 (6.3)	13 (5.5)	35 (4.6)	29 (3.8)
Healthcare use characteristics						
Hospitalizations not related to Crohn's disease ^a	48 (11.0)	43 (9.9)	17 (7.2)	18 (7.6)	101 (13.2)	95 (12.4)
Gastroenterologist visits, mean (SD) ^a	5.56 (5.83)	5.67 (5.25)	8.76 (15.6)	7.87 (8.07)	5.35 (6.07)	5.37 (6.16)

^a Assessed within 180 days before cohort entry. Abbreviation: CRP, C-reactive protein

Table 2. Outcomes, propensity score-matched with a 1:1 fixed ratio

Variable	MarketScan, n (%)		Optum, n (%)		SNDS, n (%)		Overall cohort, n (%)	
	Infliximab monotherapy	Infliximab and thiopurines	Infliximab monotherapy	Infliximab and thiopurines	Infliximab monotherapy	Infliximab and thiopurines	Infliximab monotherapy	Infliximab and thiopurines
	(n=436)	(n=436)	(n=237)	(n=237)	(n=237)	(n=764)	(n=764)	(n=764)
Treatment failure	89 (20.4)	74 (17.0)	64 (27.0)	42 (17.7)	216 (28.3)	147 (19.2)	369 (25.7)	263 (18.3)
Hospitalization or surgery related to CD ^a	57 (13.1)	45 (10.3)	39 (16.5)	28 (11.8)	127 (16.6)	95 (12.4)	223 (15.5)	168 (11.7)
Switch to another biologics or tofacitinib ^a	16 (3.7)	12 (2.8)	15 (6.3)	6 (2.5)	55 (7.2)	30 (3.9)	86 (6.0)	48 (3.3)
Corticosteroids exposure at week 26 ^a	33 (7.6)	22 (5.0)	27 (11.4)	14 (5.9)	68 (8.9)	37 (4.8)	128 (8.9)	73 (5.1)

^a Secondary outcomes are assessed without censoring at the first outcome occurrence in case of multiple outcome occurrences during follow-up. Abbreviation: CD, Crohn’s disease