Best Practice & Research Clinical Gastroenterology 38-39 (2019) 101619

Contents lists available at ScienceDirect

Best Practice & Research Clinical Gastroenterology

journal homepage: https://ees.elsevier.com/ybega/default.asp

In the absence of head-to-head trials, what do real world studies tell us about the comparative effectiveness of biologics in Crohn's disease

Joseph Meserve [a], Maria Barsky [a], Parambir S. Dulai [a, *]

[a] University of California San Diego, Division of Gastroenterology and Hepatology, 9500, Gilman Drive, #0956, La Jolla, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 19 February 2019
Accepted 23 May 2019

Keywords:
Crohn's disease
Real-world data
Real-world evidence
Comparative effectiveness

ABSTRACT

Head-to-head clinical trials are the highest quality of evidence to support comparative effectiveness. However, there are currently no head-to-head phase 3 clinical trials of biologics in Crohn's Disease. With a need for direct comparisons but lagging RCTs, Real World Data (RWD) can provide evidence on the comparative effectiveness of biologics for a diverse population that is more generalizable to routine practice. The majority of available real-world comparative analyses show no significant difference in effectiveness outcomes-primarily clinical remission and CD related complications. Real-world data is limited by its susceptibility to bias and clinicians must critically evaluate the methods and data sources utilized. Moving forward, it is important to note that comparisons including newer biologics may be limited by significant prior biologic exposure. Additionally, shared decision making incorporates efficacy, safety, and tolerability with patient preference and clinicians should use data from real-world comparative analyses as a part of this equation.

Introduction

Clinical trials represent the backbone of drug investigation, and head-to-head clinical trials are the highest quality of evidence to support comparative effectiveness and relative positioning of therapies across diseases. The fields of oncology, rheumatology, and cardiology have created a wealth of high quality head-to-head trials to inform relative positioning and treatment algorithms. In Crohn's disease (CD), despite substantial advancements in therapeutic options and our understanding of strategies to optimize disease outcomes, there are currently no head-to-head phase 3 clinical trials to inform treatment positioning of biologics.

Attempts have been made to overcome these deficiences through indirect comparisons of placebo controlled phase 3 trials for biologics in CD [1–4]. These network meta-analyses rely on the placebo groups from all studies being pooled and representing a common comparator to then make indirect comparisons across agents to inform the relative comparative effectiveness. This approach therefore has several limitations, particularly for CD. First, they are unable to accurately account for variations in patient populations, treatment strategies, and end-points over time for trials spanning over 2 decades of research in CD. Thus, the ability to appropriately account for and balance unmeasurable confounders is lost through the indirectness of comparisons. Second, the strength of network meta-analyses relies heavily on the presence of at least one head-to-head comparison to inform the overall network. Given there are currently no head-to-head trials in CD the comparisons are all made through indirect comparisons with placebo and therefore the strength of the network is somewhat limited. Finally, a substantial proportion of CD patients seen in routine practice do not qualify for these pivotal phase 3 biologic trials which brings into question the translatability of clinical trial data and network meta-analyses using clinical trial data to routine practice expectations [5]. This is of particular importance given the growing emphasis being placed on generalizability of data to routine practice for isurance coverage and payment policies.(6).

Real world data (RWD) is defined as data used for decision-making that is not collected in conventional randomized controlled trials [6]. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has provided a framework for consideration when relying on evidence generated from RWD (real world evidence, RWE) for coverage and payment decisions, and this framework lends itself to informing how RWE can be used to guide the use of RWD for studying the comparative effectiveness of biologics in CD [6]. In this review article we highlight existing

* Corresponding author.
E-mail addresses: jmeserve@ucsd.edu (J. Meserve), mmbarsky@ucsd.edu (M. Barsky), pdulai@ucsd.edu (P.S. Dulai).

https://doi.org/10.1016/j.bpg.2019.05.006
1521-6918/

RWE for the comparative effectiveness of biologics in CD and relative strengths or weaknesses of these approaches. With the growing number of biologics and small molecule inhibitors coming to market for CD, it will become impractical to do head-to-head trials for all comparisons, and we will over time need to progressively rely on RWD and RWE to inform treatment positioning in routine practice.

Considerations when interpreting real-world evidence

RCTs represent the gold standard for causal inference through rigorous study design and internal validity. By utilizing data that exists outside of RCTs, RWD represent a sample more typical of clinical practice and is able to inform clinicians on factors important for biologic comparison in CD (effectiveness, safety, affordability, etc). Generating high-quality RWE and comparative effectiveness analyses from RWD heavily relies on the quality of source data and methodology applied to data ascertainment and analyses.

Data sources

Data sources for RWD include: supplementary data collection alongside RCTs, large simple trials, patient registries, administrative claims database, surveys, and medical records. RCT supplemental data has the same limited population and short duration as its source RCT data; but it can be evaluated outside of its otherwise primary focus of clinical outcome. Large simple trials are typically larger than their RCT counterparts but are also randomized to reduce bias while including a more diverse population. Registries are prospective observational cohorts of patients with a known disease, outcome, or intervention which is amenable to long-term monitoring of outcomes in a routine population; however, they are not randomized. Administrative data, such as claims databases, are amenable to retrospective or real-time evaluation of large cohorts within the database. This data is particularly prone to data quality issues by nature of its collection, such as missing or miscoded data. Health surveys can be utilized to collect a variety of information from representative patients, however, they are vulnerable to subjectivity and recall bias by participants. Medical records and electronic health record platforms have the potential to reduce the work for chart reviews and can provide detailed information at the patient level; but transferring this clinical data into a research tool relies on assumptions and often suffers from variability on provider documentation, medical record coding, and interface of various platforms.

Endpoints and adjudication

Endpoints for RWD studies include clinical, economic, and/or patient reported outcomes. The type of endpoint utilized is often determined by the source of data. Claims databases often implore outcomes related to healthcare utilization and economic outcomes (hospitalization or surgery) given the inability to ascertain patient level effectiveness. These are informative for health policy but limited by the inability to adjudicate the reason for the event or linking of the event to actual patient level response to therapy. Other data sources will often heavily rely on patient level effectiveness, either through direct prospective measurements using validated disease activity indices, or retrospective evaluations through patient self reported surveys or medical record review. These data sources provide an ability to understand how the effectiveness of a drug is in routine practice and how this compares to results from RCTs which implore strict inclusion criteria. Irrespective of the source or type of endpoint assessed, it is imperative to understand whether an adjudication process was implored to overcome inconsistency in measurement which is a crucial component of the recently created FDA framework for evaluating RWD/RWE for regulatory decisions [7]. Well validated ICD coding alogrithms, central adjudication with redacted medical records, and/or linking of multiple data sources to confirm consistency in endpoint measurement are among the most commonly utilized adjudication processes.

Statistical approaches

When evaluating RWE design it is important to assess the investigator's approach towards handing routine practice variability in follow-up, and addressed both measured and unmeasured confounders. With large simple trials or prospective registries it is possible to create consistent follow-up intervals for measurements of effectiveness, however, with retrospective data or patient reported outcomes this is less feasible. Several approaches exist to handling this issue of variability in timing of assessment, but no single approach is uniformly correct. Therefore, studies should be assessed for their consistency in results across methods such as non-response imputation which assumes non-response in patients lost to follow-up, time-to-event which accounts for variability in censoring, and last observation carried forward which uses the last observed measure of effectiveness as the overall effectiveness. When dealing with comparative effectiveness using RWD, further consideration needs to be given to addressing both measured and unmeasured confounders which might have influenced the choice of therapy and/or the achievement of the outcome. Regression analyses adjust for measurable confounders but the use of this approach is often limited by sample size due to the increasing number of events needed to justify the number of variables accounted for in the regression model. Propensity score weighting or matching is a method in which baseline variables felt to be potential confounders are transformed into a single propensity score for each patient and then this propensity score is used for all further regression modeling or weighting and matching. Its advantages are the ability to overcome sample size issues when wanting to account for multiple baseline confounders, but its main limitation is the inability to account for unmeasurable confounders, which may actually be unleashed during propensity weighting.

Randomization is the perfect instrument for adjusting for measured and unmeasured confounders, but in the absence of randomization investigators must be aware and evaluate for possible effect of unmeasured bias on outcomes. Sensitivity analyses for sub-groups felt to be most vulnerable to unemasurable confounders can help assess the consistency of findings and is often utilized in most propensity scoring studies. Instrument variable analysis allows for pseudoramization of a cohort based on measured confounders while allowing for random variation which occurs independently of unmeasured confounders. This approach often takes into account variations in practice patterns that can not be directly measured on a patient level but can be accounted for at a population level. Although informative, instrumental variable analyses are rarely done due their statistical complexity and need for large population level data sources.

With this framework in mind we will now review the currently available RWE for comparative effectiveness of biologics in CD. Table 1 provides a summary of this.

Comparative effectiveness

TNF antagonists

Infliximab (IFX) and Adalimumab (ADA) are the two most commonly used TNF antagonists for CD, so it is not surprising that

Table 1
Bibliography.

Study	Biologics	Source Type	Adjusting for Measurable Confounders	Main Finding of Study
Singh et al. [7]	ADA vs IFX	National Registry (Dutch)	Propensity Score Matching	No difference in risk CD-related hospitalization, abdominal surgery, or repeat steroid
Singh et al. [8]	ADA vs IFX	Administrative Claims Database (Optum Labs)	Propensity Score Matching	Lower risk CD-related hospitalization, abdominal surgery, or repeat steroid use in IFX
Osterman et al. [9]	ADA vs IFX	Administrative Claims Database (Medicare)	Propensity Score Matching	No difference in treatment persistance, CD-related surgery, or hospitalization
Ananthakrishnan et al. [10]	ADA vs IFX	Tertirary Medical Center Cohorts	Propensity Score Matching	No difference in CD−related surgery, hospitalization, or steroid use
Bohm et al. [13]	VEDO vs IFX + ADA	Tertiary Medical Center Cohorts	Propensity Score Matching	No difference in clinical remission or steroid-free remission
Biemans et al. [14]	UST vs VEDO	National Registry (Dutch)	Propensity Score Matching	No difference in clinical remission or steroid-free remission

these encompass the majority of data on comparative effectiveness among this class of biologic. Even among these commonly used biologics with over 20 years of clinical experience, there is limited data directly comparing them.

Singh et al. evaluated two nationwide registry-based propensity score-matched cohorts Denmark and the US [7,8]. A total of 2908 Danish adults with biologic-naive CD were evaluated and 315 ADA-treated and 512 IFX-treated patients were included after propensity score matching. Effectiveness outcomes evaluated were all-cause hospitalization, CD-related hospitalization, major abdominal surgery, and corticosteroid prescription at least 60 days after TNF antagonist. Over a median follow-up 2.3 years after starting biological therapy there were no significant differences in rate of CD-related hospitalization (hazard ratio [HR], 0.81 [95% CI, 0.55−1.20]), major abdominal surgery (HR 1.24 [0.66−2.33]), or new steroid use (HR 1.13 [0.61−2.09]) between ADA and IFX-treated patients [7]. The US-based cohort was from a national claims database utilizing Medicare and private insurance claims from 2006 to 2014. They used the same effectiveness outcomes as the Dutch study above for 2040 biologic-naive CD patients treated with either IFX (n = 1020) or ADA (n = 1020) and found, after a median follow-up of 19 months, that IFX-treated patients had lower risk of CD-related hospitalization [adjusted HR (aHR) 0.8 (0.66−0.98)], abdominal surgery [aHR 0.76 (0.58−0.99)], and corticosteroid use [aHR 0.85 (0.75−0.96)] compared with ADA-treated patients [8].

Osterman et al. evaluated US Medicare data from 2006 to 2010 for biologic-naive CD patients. 1459 IFX-treated and 871 ADA-treated CD patients were propensity-score adjusted. Effectiveness outcomes were determined by disease persistance at 26 weeks, CD-related surgery, and CD-related hospitalation. At the 26 week mark, there was no difference in treatment persistance with 49% remaining on IFX and 47% remaining on ADA and no significant difference in CD-related surgeries [HR 0.79, (0.60−1.05)] or hospitalization [HR, 0.88 (0.72−1.07)] [9].

Ananthakrishnan et al. conducted a real-world study in which patients with CD or ulcerative colitis at two academic tertiary referral centers in Boston, MA initiated either infliximab (IFX) or adalimumab (ADA) between 1998 and 2010. The final cohort included 1060 new initiations of IFX (68% for CD) and 391 of ADA (79% for CD). Effectiveness was primarily defined as likelihood of nonresponse based on classification score of narrative mentions of symptoms in the EHR at 1 year. In CD, the likelihood of nonresponse was higher in ADA than IFX [odds ratio (OR), 1.62 (1.21−2.17)]. Similar differences favoring efficacy of IFX were observed for the individual symptoms of diarrhea, pain, bleeding, and fatigue. This method may be more prone to bias and error as it relies strictly on EHR recall of narrative mentions. There was no difference in outcomes of CD−related surgery, hospitalizations, or steroid use within 1 year after initiation of IFX or ADA in CD [10].

Integrin antagonists vs TNF antagonists

As understanding of the pathogenesis of IBD improves, novel and more selective therapies have resulted. Integrin antagonists were the first non-TNF antagonist biologics to be approved for the treatment of CD. Their increased selectivity is due to the mechanism of action and location of targeted integrins. Natalizumab and Vedolizumab inhibit the a4 integrin subunit in the CNS and gut and the a4b7 integrin heterodimer primarily in the gut and oropharynx, respectively. This results in inhibition of leukocyte adhesion at the vascular endothelium thus inhibiting migration. Natalizumab has a limited role currently since the approval of Vedolizumab (VEDO) whidh does not share the risk of JC virus reactivation and potential for Progressive Multifocal Leukoencephalopathy. Because of this limited role, we will focus on Vedolizumab in the integrin antagonist class.

Current expert consensus guidelines endorse both integrin and TNF antagonists for moderate to severely active CD. VEDO is currently FDA approved for use in CD patients with moderate to severe disease who have failed prior first line therapy with TNF antagonists. It's use as a first line therapy is still controversial but due to comparable efficacy and a safety profile more favorable than TNF antagonists it may be considered in select populations; such as those with current, prior, or high risk of malignancy, heart failure, demyelinating disease, or high risk of infection. It is important to note that evidence from both clinical trial post-hoc analysis and RWE show that biologic-naive patients receiving VEDO have higher clinical remission rates than total VEDO treated populations [11,12].

No peer-reviewed manuscripts exist to provide evidence of the comparative effectiveness of TNF-antagonists vs Integrin Antagonists in real-world data of CD. However, in review of published abstracts, there is a large propensity score-matched analysis of CD patients from the multi-center VICTORY Consortium. 538 CD patients treated with VEDO were matched with TNF antagonist treated CD patients. When adjusted for disease and patient-specific factors, VEDO-treated CD patients had a trend towards higher rates of clinical remission (HR 1.27, 95% CI 0.91−1.78) and steroid-free remission (HR 1.75, 95% CI 0.90−3.43) but was not statistically significant. Interestingly, among those with colonic CD, VEDO-treated patients had significantly higher rates of clinical and steroid free remission [13].

The lack of significant evidence can make decisions on treatment difficult, especially if considering integrin antagonists as first line therapy. Patient preferences and shared decision making are important factors in this process and could also result in VEDO being used as a first line biologic, though it may be limited by

insurance providers.

Integrin antagonists vs IL23 antagonists

Similar to integrin antagonists, IL23 inhibition for therapeutic use in IBD was the result of improved understanding of the underlying mechanism of disease and genome-wide association studies identifying IL23 variants as protective vs susceptibility foci. The only FDA approved IL23 antagonist is Ustekinumab (UST) and it is one of the newest biologics available for CD treatment. Because of its relative novelty, there is not a significant amount of data on comparative effectiveness of UST among biologics. However, the comparison between integrin and IL23 antagonists would be particularly useful in the setting of TNF antagonist failure as these are most commonly second or third-line agents in CD.

There are no peer-reviewed manuscripts comparing UST and VEDO but an abstract by Biemans et al. used propensity-score matching of a Dutch national cohort to compare 42 UST-treated and 42 VEDO-treated patients. Clinical and steroid-free remission were used as effectiveness outcomes. There was no difference between clinical remission at 24 weeks (mean −3.9 VDZ (±6.0) vs. −4.6 UST (6.5), $p = 0.72$) or in steroid-free remission at 24 weeks [46.2% VDZ and 57.9% UST, $p = 0.44$) [14].

Summary

The clinical effectiveness of all FDA-approved biologics for CD has been well-established and extends from the controlled environment of clinical trials to the more generalizable realm of real-world experience. However, there is a paucity of data comparing biologics directly to inform clinicians' treatment choices. Here we have reviewed the existing real-world evidence of comparative analyses above, not including network meta-analyses.

In summary, the majority of these studies show no significant difference in effectiveness outcomes-primarily clinical remission and CD related complications. It should be noted that while the direct comparison of TNF antagonists evaluated biologic naive CD patients, the majority of VEDO treated patients from the VICTORY consortium and 100% of VEDO and 97.6% of UST patients from Biemans' Dutch cohort were TNF exposed. As these therapies are increasingly intregrated into routine practice and their positioning within the treatment algorithm become more clear, it will be important to continue to assess the comparative effectiveness of these therapies in an ongoing basis.

Practice Points:

- The majority of current Real-World Comparative Effectiveness Research (RWCER) suggests there is no significant difference across biologics in CD
- Consider the source, statistical methodology, and limitations of RWE when applying current CER data
- RWCER should be interpreted in addition to RCT data as a part of shared decision making within the hierarchy of patient preference

Research Agenda:

- Phase 3 RCTs comparing efficacy of biologics are forthcoming but will likely be limited by cost and time
- RWCER is a less costly and time-consuming source of evidence to help inform clinicians and should be a priority with increasing therapeutic options coming to market

Conflicts of interest

- Joseph Meserve: No disclosures
- Maria Barsky: No disclosures
- Parambir S Dulai: Research support, honorarium, and travel support from Takeda, research support from Pfizer, advisory board for Janssen.

References

[1] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;14(4):417–28.
[2] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. Value Health 2011;14(4):429–37.
[3] Stidham RW, Lee TC, Higgins PD, Deshpande AR, Sussman DA, Singal AG, et al. Systematic review with network meta-analysis: the efficacy of anti-TNF agents for the treatment of Crohn's disease. Aliment Pharmacol Ther 2014;39(12):1349–62.
[4] Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic review and network meta-analysis: first- and second-line biologic therapies for moderate-severe Crohn's disease. Aliment Pharmacol Ther 2018;48(4):394–409 .
[5] Ha C, Ullman TA, Siegel CA, Kornbluth A. Patients enrolled in randomized controlled trials do not represent the inflammatory bowel disease patient population. 9. In: Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association, vol. 10; 2012. p. 1002–7. quiz e78.
[6] Garrison Jr LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health 2007;10(5):326–35.
[7] Singh S, Andersen NN, Andersson M, Loftus EV, Jess T. Comparison of infliximab with adalimumab in 827 biologic-naïve patients with Crohn's disease: a population-based Danish cohort study. Aliment Pharmacol Ther 2018;47:596–604. https://doi.org/10.1111/apt.14466.
[8] Singh S, Heien HC, Sangaralingham LR, Schilz SR, Kappelman MD, Shah ND, et al. Comparative effectiveness and safety of anti–tumor necrosis factor Agents in biologic-naive patients with Crohn's disease. Clin Gastroenterol Hepatol, 14(8). doi:10.1016/j.cgh.2016.03.038
[9] Osterman MT, Haynes K, Delzell E, Zhang J, Bewtra M, Brensinger C, et al. Comparative effectiveness of infliximab and adalimumab for Crohn's disease. Clin Gastroenterol Hepatol, 12(5), 811-817. doi:10.1016/j.cgh.2013.06.010
[10] Ananthakrishnan AN, Cagan A, Cai T, Gainer VS, Shaw SY, Savova G, et al. Comparative effectiveness of infliximab and adalimumab in crohn's disease and ulcerative colitis. Inflamm Bowel Dis 1 April 2016;ume 22(Issue 4):880–5. https://doi.org/10.1097/MIB.0000000000000754.
[11] Feagan B, Lasch K, Lissoos T, Cao C, Wojtowicz A, Khalid J, et al. Rapid response to vedolizumab therapy in biologic-naive patients with inflammatory bowel diseases. Clin Gastroenterol Hepatol 2019;17(1):130–8. e7.
[12] Kopylov U, Verstockt B, Biedermann L, Sebastian S, Pugliese D, Sonnenberg E, et al. Effectiveness and safety of vedolizumab in anti-TNF-naïve patients with inflammatory bowel disease—a multicenter retrospective European study. Inflamm Bowel Dis, 24(11), pp.2442-2451.
[13] Bohm M, Sagi S, Fischer M, Kadire S, Tran G, Rahal M, et al. OP025 Comparative effectiveness of vedolizumab and tumour necrosis factor-antagonist therapy in Crohn's disease: a multicentre consortium propensity score-matched analysis. J Crohn's Colitis, 12(supplement_1), pp.S018-S018.
[14] Biemans V, Van der Woude C, Van der Meulen A, Dijkstra G, De Boer N, Oldenburg B, et al. Dutch Initiative on Crohn and Colitis (ICC); DOP052 Vedolizumab vs. ustekinumab for Crohn's disease: comparative effectiveness in a real-life observational cohort study (ICC case series). J Crohn's Colitis 16 January 2018;12:S066–7. Issue supplement_1, https://doi.org/10.1093/ecco-jcc/jjx180.089.