

# REVIEWS IN BASIC AND CLINICAL GASTROENTEROLOGY AND HEPATOLOGY

## Integrating Evidence to Guide Use of Biologics and Small Molecules for Inflammatory Bowel Diseases



Parambir S. Dulai,<sup>1</sup> Siddharth Singh,<sup>2</sup> Vipul Jairath,<sup>3</sup> Emily Wong,<sup>4</sup> and Neeraj Narula<sup>4</sup>

<sup>1</sup>Division of Gastroenterology and Hepatology, Northwestern University, Chicago, Illinois; <sup>2</sup>Division of Gastroenterology and Hepatology, University of California San Diego, San Diego, California; <sup>3</sup>Division of Gastroenterology and Hepatology, Western University, London, Ontario, Canada; and <sup>4</sup>Division of Gastroenterology and Hepatology, McMaster University, Hamilton, Ontario, Canada

Advances in science have led to the development of multiple biologics and small molecules for the treatment of inflammatory bowel diseases (IBDs). This growth in advanced medical therapies has been accompanied by an increase in methodological innovation to study and compare therapies. Guidelines provide an evidence-based approach to integrating therapies into routine practice, but they are often unable to provide timely recommendations as new therapies come to market, and they have limited incorporation of real-world evidence when making recommendations. This limits the scope and usability of guidelines, and a gap remains in defining how best to position and integrate advanced medical therapies for IBD. In this review, we provide a framework for clinicians and researchers to understand key differences in sources of evidence, how different methodologies are applied to study the comparative effectiveness of advanced medical therapies in IBD, and considerations for how these sources of evidence can be used to better integrate current guideline recommendations. Over time, we anticipate this framework will allow for a transition to living guidelines and/or practice recommendations.

**Keywords:** Comparative Effectiveness Research; Causal Inference; Real-World Data.

Inflammatory bowel diseases (IBDs)—ulcerative colitis (UC) and Crohn’s disease (CD)—are chronic immune-mediated disorders that often require long-term use of biologics and/or small molecules to achieve sustained disease remission. Substantial advances have been made in our understanding of disease mechanisms and the development of novel advanced medical therapies. With this increased therapeutic armamentarium comes a need to better understand how different levels of evidence should be used to integrate therapies in routine practice. Randomized controlled trials (RCTs) represent the most scientifically robust source of data to define treatment efficacy and comparative efficacy. RCTs form the basis for guidelines and societal recommendations, however, they have limitations for IBD specifically related to generalizability of study populations,<sup>1</sup> and feasibility for timely completion to inform utilization in the ever-evolving practice setting. Real-world evidence (RWE), the real-world setting investigation of

therapies already found to be efficacious in RCTs, provides an understanding of effectiveness and safety when prescribed by licensed providers with varying degrees of expertise and practice patterns.<sup>2</sup> RWE is generated more rapidly within the first years of drug availability, offers an assessment of routine practice effectiveness and safety, and may help to identify populations in which the greatest potential benefit or risk may be seen. However, the uncontrolled environment within which these data are generated creates difficulties in making inferences about the validity of results. A critical gap remains in how we approach recommendations for advanced medical therapy positioning, and there is a need to bridge guideline recommendations informed by RCTs with practice-level data. In this review, we will discuss relative advantages and disadvantages of data sources for evidence generation, how these data sources can and have been applied to guide the integration of biologics and small molecules for IBD, and we will provide considerations for future efforts to assimilate the necessary evidence more rapidly for advanced medical therapies as they come to market for IBD. Finally, we provide an updated framework for advanced medical therapy positioning from the current American Gastroenterological Association (AGA) guidelines for both CD and UC for provider use.

### Randomized Clinical Trials vs Real-World Evidence

The underlying conceptual framework of “measurable” and “unmeasurable” confounders defines the core debate around RWE. Measurable confounders are any factors that can be directly quantified or assessed, and unmeasurable

**Abbreviations used in this paper:** AGA, American Gastroenterological Association; CD, Crohn’s disease; ER, endoscopic remission; IBD, inflammatory bowel disease; IPD, individual participant data; MAIC, matching-adjusted indirect comparison; OR, odds ratio; RCT, randomized controlled trial; RWE, real-world evidence; TNF, tumor necrosis factor; UC, ulcerative colitis.

Most current article

© 2024 The Author(s). Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0016-5085

<https://doi.org/10.1053/j.gastro.2023.10.033>

confounders are all other unknown factors that may or may not influence outcomes. The concept of measurable varies by data source, such as endoscopic disease activity, which might be a measurable confounder in an observational cohort in which endoscopy assessments are available but in a claims data set where the data cannot be ascertained, it becomes an unmeasurable confounder. Thus, the presence of unmeasurable confounders cannot be consistently or adequately accounted for and there are data set-specific limitations that need to be considered. In contrast, the act of randomly assigning participants to different interventions balances all confounders, measurable and unmeasurable, if sufficiently powered; this is why RCTs are felt to be the highest-quality evidence for comparisons<sup>3-5</sup> (Figure 1<sup>4-6</sup>). Accordingly, it has been suggested that the ideal approach to bridging gaps in evidence synthesis is not to increasingly rely on observational data sets where data gaps may introduce bias, but instead to remove obstacles to randomized trial completion and ideally design and conduct large, simple trials that can be integrated with routine clinical care.<sup>7</sup>

In the field of IBD, there are specific limitations that must be considered before discounting RWE as an important source of evidence synthesis to guide advanced medical therapy integration (Supplementary Table 1<sup>8-14</sup>). IBD is not simply 2 diseases—CD and UC—but rather 2 categories (CD and UC) with multiple permutations for sub-groups (based on demographic characteristics, treatment exposures, and disease extent and location, or prior complications) and disease phenotypes (eg, perianal disease and pouchitis) that cannot be feasibly studied individually in RCTs. The most common limitation cited for RCTs, which has supported the promotion of RWE for advanced medical therapies in IBD, is the lack of generalizability in findings due to excessively stringent eligibility criteria for trials. The strict eligibility criteria are used for a number of reasons, including the following: to minimize risk of trial failure by enrolling a homogeneous population in which outcomes measures are relatively well defined and without introducing safety risks through inclusion of patients with significant comorbidity; to minimize placebo effects through use of centralized reading of endoscopy, which introduces additional procedures and processes; and lack of validity for current patient-reported outcome and endoscopic indices for subpopulations cared for in routine practice who are usually excluded from trials (eg, pouch, ostomy, and postoperative). RWE offers a distinct advantage here in its ability to rapidly assimilate effectiveness and safety data post market approval for advanced medical therapies to guide treatment integration among subtypes typically excluded from clinical trials, vulnerable populations (eg, older adult or pediatric patients, active or recent cancer, concomitant infectious diseases, such as HIV and hepatitis C or B), underserved and underrepresented populations often not enrolled (eg, racial and ethnic minority groups, low socioeconomic status, or those with disabilities), and difficult to treat or high-risk populations (eg, recent hospitalizations, multiple surgical procedures, or multiple biologic failures). It is impractical to consider performing RCTs in all of these populations to

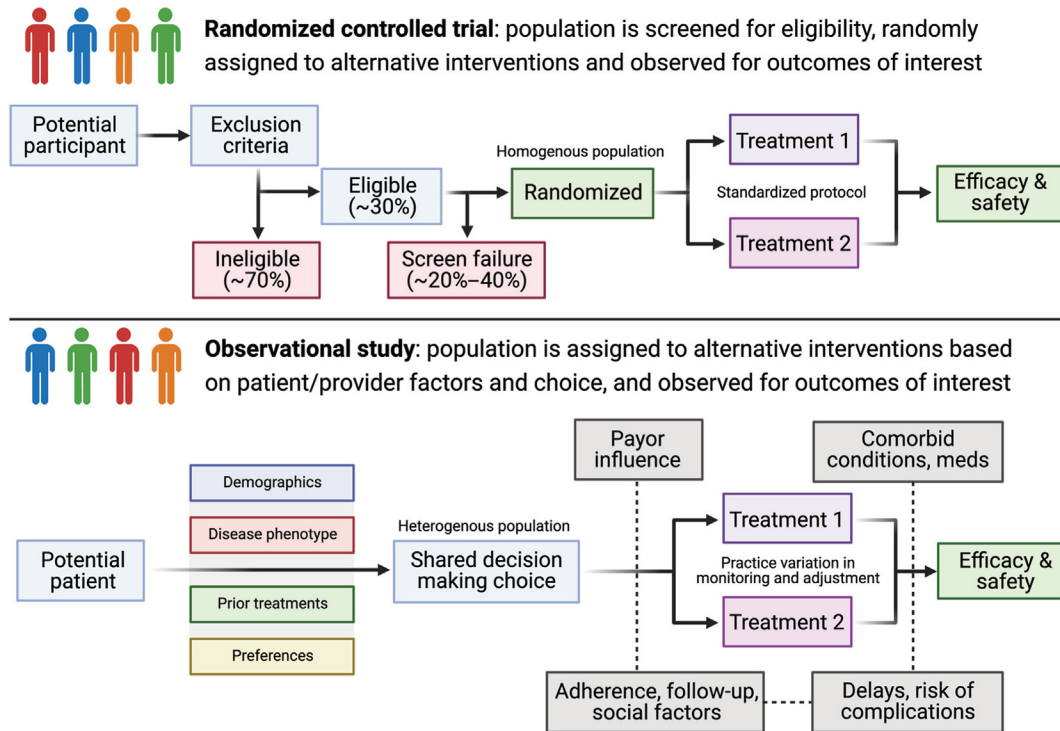
either confirm efficacy and safety or define comparative effectiveness and safety. Therefore, we must critically appraise current data synthesis approaches in the field to understand how RCTs and RWE can be used together to allow for strengths of one to overcome weaknesses of the other and define the optimal integration of advanced medical therapies in routine practice.

## Approaches to Use of Data to Guide Integration

To date, there have been 2 head-to-head clinical trials for regulatory approved advanced medical therapies completed in the field of IBD, the VARSITY trial, which compared vedolizumab with adalimumab in UC, and the SEAVUE trial, which compared ustekinumab with adalimumab in CD.<sup>15,16</sup> The VARSITY trial demonstrated superiority for vedolizumab vs adalimumab for clinical and endoscopic remission, however, criticism remains due to a lack of uniform steroid-tapering protocol and its potential influence on the outcomes. Results from the SEAVUE trial did not show superiority for ustekinumab vs adalimumab in achieving symptom-based remission, and criticisms remain that the high rates of clinical remission in this early-disease, moderately active, CD cohort are not representative of routine practice expectations for effectiveness. There are very few examples of real-world clinical trials (ie, explanatory trials) conducted to date in the field of IBD. Typically, these trials are considered pragmatic trials, as defined by the Pragmatic-Explanatory Continuum Indicator Summary 2 criteria, which uses a 9-domain spoked wheel approach for trial design decisions to ensure applicability. These domains are scored from 1 (very explanatory) to 5 (very pragmatic) to facilitate discussion, consensus, and ensure trial design is consistent with intended use of trial results.<sup>17</sup> The 2 trials in the field of IBD considered closest to the definition of pragmatic trials are the REACT trial and REACT2 trial.<sup>18</sup> These were both designed as cluster randomized trials comparing the effectiveness of treatment algorithms in CD when implemented into routine care in community practices. Aside from these studies, all other considerations for comparative effectiveness and safety of regulatory approved advanced therapies have been made through indirect assessments via the following main approaches: network meta-analyses and either post-hoc comparisons of individual participant data (IPD) from RCTs or observational comparative studies, using various statistical approaches to adjust for measurable confounders. In the following sections, we outlined the rationale, methodology, advantages, and limitations, as well as use cases, for each of these methodologies.

### Network Meta-Analyses

**Rationale.** Network meta-analyses involve the simultaneous analysis of direct evidence (from RCTs directly comparing treatments of interest) and indirect evidence (from RCTs comparing treatments of interest with a common comparator), to calculate a mixed-effect estimate as the



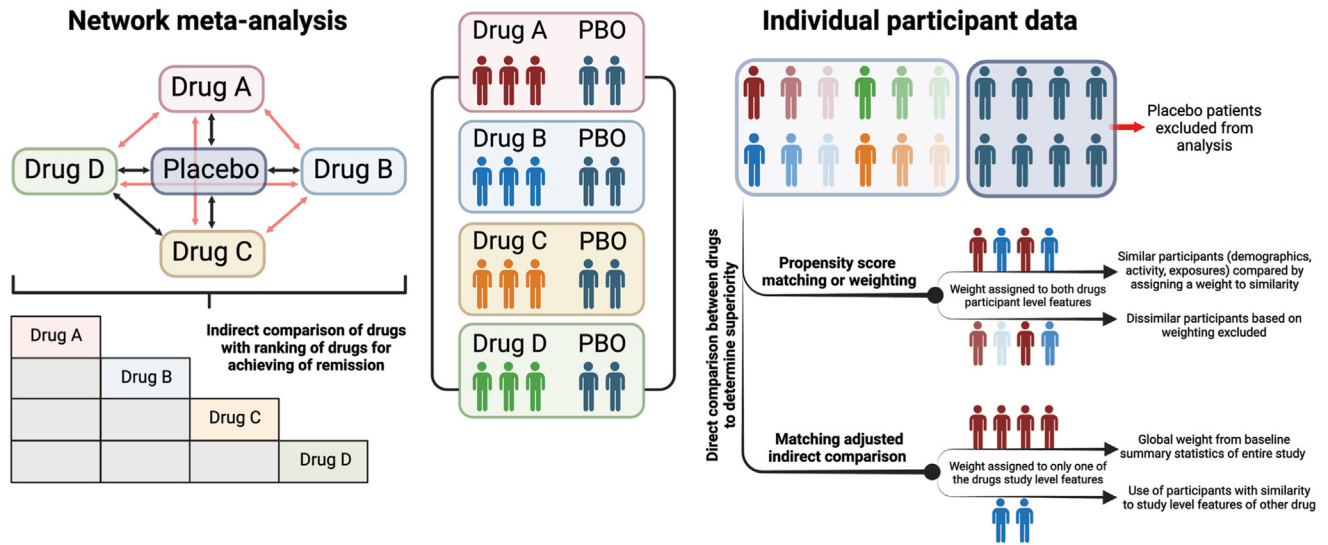
**Figure 1.** Comparison of RCTs with RWE observational studies studying comparative effectiveness and safety of 2 therapies in IBD. In contrast to highly selected participants in RCTs, where the act of randomization balances both measurable and unmeasurable confounders and both groups have an equal opportunity to receive either therapy and/or presumed equal baseline probability to experience outcomes of interest, observational studies include broader populations and have multiple levels of influence on treatment choice, adherence, probability of experiencing outcomes of interest, and variability in care delivery or observational period.

weighted average of the 2, as shown in Figure 2.<sup>19,20</sup> Such a technique can help improve the precision of the estimate (compared with direct evidence alone), and also allows estimation of the comparative efficacy of 2 active treatments, even if no studies compare them directly. These indirect comparisons of competing interventions, adjusted by a common control such as placebo, can help to partially account for prognostic characteristics of patients in different trials, and may allow for more rapid and ongoing assimilation of data to inform stakeholders regarding the optimal integration of advanced medical therapies in clinical practice.

**Approach.** When performing and interpreting a network meta-analysis, there are 2 key assumptions: transitivity (ie, similarity of trials and populations) and coherence (ie, consistency of findings).<sup>19,21,22</sup> Transitivity implies that trials are sufficiently similar in important clinical and methodological characteristics. When designing and interpreting a network meta-analysis, it is critical to understand whether included trials are conceptually very similar in terms of key effect modifiers, such as patients (similar disease characteristics and severity, including prior failure of therapies); interventions (standard dose and schedule of approved therapies from phase 3 trials); co-interventions (such as steroid dose and tapering protocols, which can influence treatment efficacy); outcome assessment (similar definitions and measures of outcome, such as criteria for endoscopy scoring and methodology used for

scoring—central vs local); and trial design and quality (treat straight-through design vs re-randomization of responders, open-label feeders into maintenance phase, timing of outcome assessment, and low risk of bias). These judgments are made *a priori* by the investigators and are subjective. After conducting a network meta-analysis, coherence (statistical agreement) should be assessed between direct and indirect comparison for the same treatment comparison pair. If incoherence is observed, potential reasons should be examined, including genuine diversity in findings due to differences in enrolled participants, interventions, background treatment and management, and definition and measurement of outcomes, or bias in direct (or head-to-head trials) or indirect comparisons due to limitations in trial design or patient populations.

There are 2 main statistical approaches to performing a network meta-analysis: Bayesian and Frequentist methods.<sup>21</sup> Bayesian models use the deviance information criterion to compare models and assess overall goodness of fit, whereas non-Bayesian models often use hypothesis tests based on deviance statistics. Regardless of the approach selected, it is recommended that the model fits the data well. Inconsistency or heterogeneity in network meta-analyses is assessed using global measures of inconsistency and loop-specific approaches for each comparison where direct and indirect evidence exists. Network meta-analyses are best presented as risk estimates comparing different interventions with each other (relative risk or odds ratio [OR],



**Figure 2.** Approach to use of clinical trial data to generate different levels of evidence. Network meta-analyses take group-level estimates to make inferences on comparative effectiveness between 2 interventions using a common comparator (often placebo [PBO]) or available data from head to head trials. Post-hoc analyses of individual participant data take patient-level data and perform either propensity score modeling, which assigns a propensity score to sum up dependent variables and allow for matching or weighting of participants from each treatment group to make comparisons, and matching-adjusted indirect comparison, which reweights baseline summary statistics from individual participant data in one study to baseline summary statistics of another study.

with 95% CIs or 95% credible intervals in case of Bayesian analysis). In addition, ranking probabilities best presented as surface under the cumulative ranking curve, or P-scores (based on point estimates and SEs of the network estimates under the normality assumption<sup>23</sup>) can help inform relative efficacy, although these are frequently overinterpreted.<sup>24</sup> Finally, findings from network meta-analysis, like findings from traditional meta-analysis, should be critically appraised using standard approaches, such as Grading of Recommendations Assessment, Development and Evaluation for network meta-analysis.<sup>25</sup>

**Use case.** A recent network meta-analysis that examined the comparative efficacy of different advanced medical therapies for the management of tumor necrosis factor (TNF)-naïve and TNF-exposed patients with moderate to severe UC informed the AGA clinical guidelines for the management of moderate to severe UC.<sup>26</sup> The AGA guidelines suggested using infliximab or vedolizumab rather than adalimumab for induction of remission in adult outpatients with moderate to severe UC who are naïve to biologic agents.<sup>27,28</sup> In adult outpatients with moderate to severe UC who have previously been exposed to infliximab, particularly those with primary nonresponse, the AGA suggests using ustekinumab or tofacitinib, rather than vedolizumab or adalimumab for induction of remission.

**Strengths and limitations.** The following are the key strengths of network meta-analyses: ability to compare interventions that have not been compared head to head; being able to deal with multiple treatments and synthesize comparative efficacy of the entire body of RCT evidence for a specific condition; and potentially increasing precision, especially with connected networks.<sup>29</sup> However, network meta-analyses often fail to provide a clear choice (ie, either infliximab or vedolizumab as first-line therapy in the AGA

guidelines) and there are several important limitations in interpreting network meta-analyses and their applicability to integrating advanced medical therapies, which are outlined in [Supplementary Table 2](#).<sup>30,31</sup> First, there has been a substantial evolution in how disease activity is assessed over time. For example, there has been an evolution in how endoscopy is assessed (local readers vs central readers) and how friability is used in the Mayo endoscopic subscore from the original infliximab trials, which separated mild (Mayo 1) and moderate (Mayo 2) friability, to modern-day trials where any friability is now considered moderate (Mayo 2). In CD, there has been a shift from using the full Crohn's Disease Activity Index to using stool frequency and abdominal pain subscores only, while awaiting fully validated disease activity measures.<sup>32</sup> This was recommended by the US Food and Drug Administration in 2015 due to the high placebo rates, but this shift cannot be fully accounted for when comparing CD trials pre 2015 to post 2015 in network meta-analyses.<sup>33</sup> Second, judgments on transitivity or comparability of trial participants are subjective. In patients with IBD, prior exposure to biologic therapies, the impact of different biologic exposures (TNF-antagonist vs others) on efficacy of subsequent therapies, reasons for treatment failure (eg, primary nonresponse and secondary loss of response), disease duration, disease distribution, and burden of inflammation, can significantly influence treatment efficacy and these have evolved considerably over time in cohorts enrolled. Third, co-interventions are not standardized across trials with consideration for steroid dosing at entry and steroid-tapering regimens, which likely influence efficacy estimates for active therapy and placebo arms. Through study-level synthesis, network meta-analyses are unable to account for these differences in trial populations or co-interventions. Fourth, network

meta-analyses are unable to adequately account for differences in trial design and conduct. Although some approaches have been suggested, there are additional layers of assumptions involved and this is particularly important for the evolution in trial design with re-randomization of responders in maintenance being commonplace. Yet, network meta-analyses continue to compare trials without consideration for how long-lasting pharmacodynamic impacts may influence placebo rates during maintenance, thereby influencing ranking. An example of this is seen in the recently completed risankizumab trials for CD.<sup>34,35</sup> Fifth, some network meta-analyses that include early-phase clinical trials can introduce instability and be overly optimistic for newer therapies. There is frequent over-reliance on ranking probabilities, without a critical appraisal of the quality of evidence. Finally, network meta-analyses are unable to adequately inform comparative safety of therapies, due to low frequency of major adverse events in short-term clinical trials of IBD. For these reasons, network meta-analyses cannot be solely relied on for decisions on how advanced medical therapies should be integrated into clinical practice.

### *Post-Hoc Comparisons of Individual Participant Data From Randomized Trials*

**Rationale.** It is impractical to expect head-to-head trials for all possible comparisons of advanced medical therapies in IBD, and the limitations of fully relying on network meta-analyses to inform these comparisons has been outlined above. RCT data sets still represent the most well-phenotyped and monitored cohorts in existence for advanced medical therapies, thereby providing the highest quality assessment for potential measurable confounders. One strategy to overcome the limitations noted for study level comparisons through network meta-analyses is to use IPD from separate RCTs to make more direct participant comparisons after adjusting for key potential confounders. Although the potential for confounding is still an important consideration, given the absence of randomization between the interventions being compared, evolutions in techniques to account for unevenly distributed confounding factors have allowed for more optimal use of IPD to guide comparative effectiveness research in IBD and inform integration of advanced medical therapies.

**Approach.** IPD is often obtained through repositories, such as the Yale University Open Data Access Project and Vivli, which act as custodians for these data for pharmaceutical companies that have conducted the RCTs. Although these central data repositories have proven to be extremely valuable, particularly within the context of recent efforts to optimize data transparency, open science, and data sharing<sup>36</sup>; they are not without limitations in accessibility, usability, and timeliness in data availability. A substantial proportion of requests for data never yield completed projects due to complexities in data analyses and realizations by investigators that their expertise was insufficient to perform the proposed work, as outlined by the co-directors of the Yale University Open Data Access Project.<sup>37</sup> Furthermore, variation in data file types, inability to

readily integrate data across platforms that separately store pivotal IBD trial data sets, and delays in data upload by sponsors, impact timeliness in data synthesis. Nonetheless, post-hoc analyses of IPD from these RCTs in IBD when feasible offer several benefits.

Several approaches have now been applied to statistically adjust IPD when comparing interventions from different studies in IBD, with propensity score approaches being the most widely used (Figure 2). The main limitation with traditional regression (logistic or proportional hazard) modeling is the need for a larger number of events to allow for accurate adjustment of potential confounders. Methods using propensity scores can mitigate the potential for selection bias by improving balance of baseline covariates between groups without being restricted by number of covariates accounted for in cohorts. The assignment of propensity scores for all included patients acts as a proxy between treatment assignment and potential confounders, thereby maximizing comparability. Growing evidence supports propensity score approaches to be superior to regression modeling,<sup>38</sup> and propensity score approaches for IPD comparisons across trials help to address the following key limitations in effect modifiers of network meta-analyses: patients—an ability to account for disease characteristics, severity, and prior treatment failures as covariates; co-interventions—an ability to account for variation in doses used between induction and maintenance and/or concomitant medications, such as steroids; and outcome assessment and trial design—an ability to make definitions uniform across cohorts through recalculation of end points and/or modification of timing of assessment for comparisons from those used in the original RCTs.

**Use case.** The network meta-analysis for UC, which informed the AGA guidelines, recommended either infliximab or vedolizumab for first-line use in UC, however, it could not comment on any potential superiority of these drugs to each other, individual patients most ideal for either therapy, or 1-year outcomes due to variations in trial designs that could not be accounted for through study-level comparisons. A post-hoc analysis of IPD for biologic-naïve patients with UC from these pivotal phase 3 programs for infliximab and vedolizumab was able to overcome these barriers and observed that although both drugs were comparable for achieving improvements in symptomatic improvement at 1 year, infliximab was superior for achieving corticosteroid-free clinical and endoscopic remission at 1-year in biologic-naïve patients with UC.<sup>39</sup> This demonstrates an ability for IPD to extend observations from network meta-analyses to inform integration of advanced medical therapies in routine practice. In CD, IPD has had a particular advantage when considering more objective and robust outcomes. Although several network meta-analyses have compared the relative efficacy of various biologics for achieving clinical remission in CD,<sup>40–42</sup> no comparisons have been made for 1-year endoscopic remission (ER), given the varying use of ER as an end point, variations in definitions of ER, and timing of assessment across trials, small sample sizes for endoscopy secondary outcomes, and differences in trial design due to evolutions

in use of re-randomization of responders approaches. Using patient-level clinical trial data obtained from the Yale University Open Data Access Project and Vivli, a comparison of adalimumab, infliximab, ustekinumab, and vedolizumab found that adalimumab and infliximab had rates of ER superior to those of vedolizumab, and among patients with ileal CD with large (0.5 cm) ulcers, infliximab achieved the highest rates of ER at 1 year.<sup>43</sup> A major strength of this study was the ability to use a uniform definition for study outcomes, which were global ER (Simple Endoscopic Score for Crohn's disease score < 3) and segment-specific ER (Simple Endoscopic Score for Crohn's disease score = 0). In this regard, IPD comparisons can be extremely informative and consideration should be given to a stepwise triage of questions to be addressed with network meta-analyses defining key comparisons needing assessments and then post-hoc IPD analyses being used to provide more granular refinement for integration in practice.

**Strengths and limitations.** Advantages of conducting indirect patient comparisons using patient-level clinical trial data include the use of robust prospective collected data from phase 3 and 4 clinical trials, use of standardized clinical symptom indices to measure disease activity, an ability to deconstruct disease activity measures (such as endoscopic sub-scores) to create uniform definitions across trials, and the ability to compare outcomes uniformly defined across common time points. Several limitations should also be acknowledged. Indirect comparisons cannot replace the need for randomization due to the inability to balance for unknown confounders. As these are not randomized comparisons, results from indirect comparisons cannot be interpreted as such. Given the observational nature of indirect comparisons, there remains the potential for several biases, which may be introduced if the assumptions underlying indirect comparison are violated,<sup>44</sup> giving rise to other methodological issues. Finally, availability and accessibility of data make it difficult to readily apply this methodology to inform decision making, as advanced medical therapies come to market. To specifically overcome this limitation, a greater emphasis has been placed recently on a new technique—matching-adjusted indirect comparison (MAIC)<sup>45</sup> (Figure 2). This approach reweights IPD from one study to the baseline summary statistics of another, to provide greater adjustment for observed trial differences compared with conventional meta-analytic methods.<sup>46</sup> This allows for more timely comparative evidence than propensity score approaches with IPD, but helps to partially overcome limitations of network meta-analyses by incorporating trial level IPD characteristics from one group. Early analyses using this approach have been applied for therapies yet to obtain US Food and Drug Administration approval (filgotinib<sup>47,48</sup>) or those very recently approved (upadacitinib<sup>49</sup> and ozanimod<sup>50</sup>), but caution should be taken with widely adopting this methodology and relying heavily on these results for informing treatment integration, given you only have IPD from one group (vs IPD from both groups for propensity score approaches) and therefore the weighting adjusted may be biased due to large weights being applied to a small proportion of the index population to

balance the groups.<sup>51</sup> An example of this is seen in a recent article comparing multivariate regression modeling, MAIC, and propensity score matching for the comparison of vedolizumab with adalimumab in biologic-naïve participants from 2 independently conducted trials GEMINI (vedolizumab) and ULTRA (adalimumab), with the head-to-head trial (VARSITY; vedolizumab vs adalimumab) serving as ground truth. In this analysis, MAIC favored adalimumab over vedolizumab, which differed from the propensity score approach using IPD from both studies and the head-to-head VARSITY trial, which both favored vedolizumab.<sup>52</sup> MAIC is an attractive analytic approach for sponsors to rapidly compare emerging therapies using internal IPD data sets with external cohort-level summary statistics to inform positioning as therapies come to market, but caution needs to be taken when interpreting the results and further work is needed to refine how these analyses are done for IBD therapies.

### *Observational Comparative Studies*

**Rationale.** RWE has several limitations, including selection bias, unsystematic data collection, inconsistency in outcome measurements or data availability, and retrospective design, which must be considered when interpreting comparative effectiveness data generated from observational cohort studies. However, there are multiple advantages to the use of RWE, which have driven the dramatic increase in studies over the past decade. First, the use of these therapies in patient populations or cohorts not included in the pivotal clinical trial programs creates a greater generalizability for efficacy estimates and comparative effectiveness findings in routine practice. Second, the way these therapies are used in routine practice often allows for an exploration of off-label dosing (ie, dose optimization) and clinical benefits that can lead to novel insights into optimally integrating these therapies into practice. Third, the large size of these cohorts and use in patients with comorbid conditions often allows for an assessment of safety, particularly for events of key importance such as cancer, cardiovascular events, or serious infections, which are often not feasible in clinical trial data sets due to the short follow-up period and/or highly restrictive eligibility criteria. Finally, there is an increased interest in using RWE to inform treatment positioning for payors with increased adoption and consideration by international societies such as ISPOR (Professional Society for Health Economics and Outcomes Research) on how RWE may be used by regulatory agencies through the 21<sup>st</sup> Century Cures Act. Most importantly, RWE gives providers an opportunity to assess the “face validity” of treatment efficacy in the patients they treat in routine practice and provides them with confidence in expectations, as well as evidence to appeal payor decisions on positioning and/or dosing of advanced medical therapies. For these reasons, RWE plays a pivotal role in defining how therapies are integrated into clinical practice.

**Approach.** Methodological approaches to RWE generation with observational cohorts follow statistical principles like post-hoc analyses of IPD from RCTs. The primary consideration within observational cohort data RWE

generation is the cohort creation and methodology for data integration. There are 2 broad approaches to obtaining real-world data—grass roots data collection and integration or systematic data storage with careful curation using available metadata to guide cohort creation. Examples of grass roots approaches include the GETAID (Groupe d'Étude Thérapeutique des Affections Inflammatoires du Tube Digestif) collaboration in France,<sup>53-58</sup> I-CARE (IBD Cancer and Serious Infections in Europe) European collaboration,<sup>12</sup> and the IBD Health Outcomes Consortium in the United States.<sup>9-11,59-61</sup> These groups reported systematic data collection, prospectively or retrospectively, using pre-defined data coding sheets, definitions, and quality-control measures to collate deidentified data into common data repositories through pre-established data sharing agreements to comply with data privacy. Although this approach often provides a higher degree of confidence in data quality, it requires considerable time and resources investment, which limits feasibility. Furthermore, it remains dependent on investigator engagement and data availability, which does not adequately overcome barriers related to inclusion of underserved or underrepresented populations often not seen in these centers. An alternative approach to using real-world data for RWE generation in IBD is to use available metadata from closed or open health care systems through natural language processing and privacy-preserving data linkage approaches. Successful examples of these approaches include the Veterans Affairs Health System,<sup>62-65</sup> and the University of California Research eXchange (UCReX).<sup>66-68</sup> These large health system-based data integrations rely on common enterprise data warehouses of existing medical records to collate patient-level information in deidentified manners, and then create opportunities for use of honest data brokers and/or remote data integration to study large populations across diverse geographic regions with varying provider care approaches, access, and/or socioeconomic barriers influencing outcomes. Although neither of these approaches can overcome limitations seen with RWE using real-world data, they do offer unique opportunities to extend observations from network meta-analyses and/or post-hoc analyses of IPD to determine consistency.

**Use case.** Continuing with our theme use case of first-line biologic therapy in moderate to severe UC, the US IBD Health Outcomes Consortium studied the comparative effectiveness of vedolizumab and TNF-antagonist therapy in UC. They observed vedolizumab to be superior to TNF-antagonist therapy for the achievement of clinical remission, steroid-free clinical remission, and steroid-free deep remission (clinical plus endoscopic remission), which was mostly driven by the sub-group comparison in biologic-naïve patients and for vedolizumab vs adalimumab. The point estimate for comparative effectiveness of clinical remission from this study (hazard ratio, 1.65 in favor of vedolizumab; 95% CI, 1.23–2.22) was very similar to that seen in the subsequently completed head-to-head RCT of vedolizumab and adalimumab (OR, 1.57 in favor of vedolizumab; 95% CI, 1.14–2.16).<sup>15,69</sup> This is in contrast to the network meta-analysis completed before the VARSITY trial (OR, 2.41 in favor of vedolizumab; 95% CI, 0.75–7.79),<sup>70</sup> and

the subsequent network meta-analysis completed shortly after with VARSITY data included (OR, 1.31 in favor of vedolizumab; 95% CI, 0.88–1.95).<sup>26</sup> Taken together, the totality of data from network meta-analyses, post-hoc analyses of IPD from clinical trials, and RWE from observational cohort data, demonstrated consistency in choice of first-line biologics for moderate to severe UC with differences in relative strengths and limitations that help to further bolster confidence when the totality of data are assessed together.

**Strengths and limitations.** The strengths and limitations of real-world data from observational cohorts has been outlined throughout this review, but one specific limitation worth noting is the absence of centralized endoscopic video recording, which could be readily addressed in the coming years. There has been a substantial increase in uptake of video-recording solutions for routine endoscopic procedures, including “always on” capture devices, which allow for assessment of colonoscopy quality and are used for colon cancer screening augmentation.<sup>71</sup> These endoscopic video-recording solutions have been embedded for the goal of ensuring quality and potential applications for artificial intelligence, primarily for adenoma detection. However, over time, if these video-recording solutions are linked with large collaborative consortiums in open or closed health system models, a tremendous opportunity is created to harness real-world data for rapid assessment of therapies as they come to market with rigorous data measures. The promise of real-world data and use of real-world data to guide advanced medical therapies in IBD is therefore at a critical junction and we anticipate substantial successes in coming years.

### *Personalized Medicine and Prediction Models: Integrating Individual Participant Data From Randomized Controlled Trials With Real World Data Sets to Guide Integration of Advanced Medical Therapies in Inflammatory Bowel Disease*

Within the entire framework we have outlined, we must acknowledge that these data sources provide only group-wise comparisons for individual therapies. They do not offer concrete opportunities to provide patient-level assessments for optimal advanced medical therapy selection. In this regard, prediction modeling will likely be the final data point to consider when fully defining how advanced medical therapies should be integrated. A case use example of this, which leveraged the multiple data sources outlined above, is the US IBD Health Outcomes Consortium Clinical Decision Support Tool for UC and CD ([www.CDSTforIBD.com](http://www.CDSTforIBD.com)).<sup>72</sup> Using IPD from the pivotal phase 3 clinical trial program for vedolizumab in UC, a regression model was built and subsequently validated in the routine practice cohort from the US IBD Health Outcomes Consortium and demonstrated good discriminative performance for steroid-free clinical and endoscopic remission with vedolizumab in UC. Most notably, the prediction was specific to vedolizumab, and the tool did not predict outcomes for TNF-antagonist therapy in routine practice. A subsequent validation of the tool in the

VARSITY trial data set for predicting histoendoscopic remission using blinded central reads of videos and histology demonstrated the tool was able to predict outcomes for vedolizumab but not adalimumab in this RCT.<sup>73</sup> Although the overall trial demonstrated superiority of vedolizumab to adalimumab, the decision support tool was able to identify a sub-group of patients who would benefit most from vedolizumab and a sub-group who might benefit from adalimumab over vedolizumab within this trial data set. Finally, a recent extension of the tool applied to the tofacitinib trial data sets obtained through Vivli demonstrated that the tool identified patients most likely to respond to vedolizumab relative to adalimumab and tofacitinib, despite this sub-group of high-probability patients having factors that would traditionally be considered associated with high probability of response to any advanced medical therapy. There was an inverse probability of response to tofacitinib relative to vedolizumab and adalimumab, which further confirmed that prediction modeling may help to further refine groupwise comparative estimates to fully define how advanced medical therapies should be integrated into routine practice. Similar observations have been made for decision support tools from the consortium built for vedolizumab and ustekinumab in CD, which have undergone validation by other groups internationally and have been found to have drug specificity in prediction across broad populations.<sup>74–79</sup> As clinical trial data become increasingly available through open access data platforms, and routine practice observational cohorts become increasingly higher in data quality, opportunities exist to expand decision support tool creation and validation to guide treatment positioning.

## State of Affair for Advanced Medical Therapy Positioning

As we have outlined in this review, current guidelines cannot fully inform treatment positioning and integration, and no single source of evidence can be relied on to make these decisions. Therefore, a combined approach is needed where all sources of evidence are considered together in a rank-order manner with RCTs, network meta-analyses, post-hoc analyses of individual participant data, and then RWE. In this regard, we provide suggested recommendations that extend the current AGA guidelines for both UC and CD. These recommendations should not be considered as formal endorsements, but they represent the opinion of the authors based on best-available evidence. These extended recommendations can be found in [Table 1](#)<sup>15,39,41,43,80–88</sup>

## Conclusions and Future Considerations

RCTs remain the gold standard for evidence, and a shift toward more comparative studies is developing with the use of multiple active comparator arms to reduce probability of randomization to placebo, and trials would ideally be designed with uniform end-point assessments and measurement of core outcome sets to allow for more uniform

comparison within network meta-analyses.<sup>89</sup> An additional shift toward more pragmatic trial designs is also needed where studies have minimum eligibility criteria, include simple interventions that could be incorporated with routine care, and measure nontraditional outcomes or rare events (safety) that are most important to patients and enable simple outcome data to be extracted from routine health records.<sup>90</sup>

When considering observational data and causal inference, target trial emulation may offer a unique opportunity to optimize study quality.<sup>91</sup> Target trial emulation is a 2-step process. The first step is articulating the causal question in the form of the protocol of a hypothetical randomized trial that would provide the answer. This protocol specifies key elements as a clinical trial, including eligibility criteria, treatment strategies, treatment assignment, the start and end of follow-up, outcomes, causal contrasts, as well as data analysis plan. The randomized trial described in the protocol becomes the target study for the causal inference of interest. The second step is explicitly emulating the components of that protocol using the observational data: finding eligible individuals, assigning them to a treatment strategy compatible with their data, following them up from assignment until outcome or end of follow-up, and conducting the same analysis as the corresponding target trial, except that there is adjustment for baseline confounders to emulate random treatment assignment through propensity score methods. This approach was used recently to compare vedolizumab monotherapy vs vedolizumab in combination with thiopurines for the management of IBD. Using observational data from 3 administrative claims databases, mirroring the SONIC and SUCCESS trials comparing infliximab monotherapy vs infliximab in combination with thiopurines, Kirchgesser and colleagues<sup>86</sup> inferred that combination therapy with vedolizumab and thiopurines was associated with lower treatment failure compared with vedolizumab monotherapy in CD but not UC.

Finally, irrespective of how the evidence is generated, there remains a need for rapid integration and lessons can be learned from the COVID-19 pandemic using living guidelines. A living guideline is defined as “optimization of the standard guideline process, as it allows updating of individual recommendations as soon as new relevant evidence becomes available.”<sup>92,93</sup> The aim of a living guideline is to provide accurate, relevant, and trustworthy advice to clinicians, policy makers, and other decision makers in a timely manner, particularly in the context of rapidly evolving evidence and an urgent need to inform decision making. The unit of update in living guidelines is an individual recommendation, and not necessarily the whole guideline. Living guidelines were immensely useful to inform pharmacologic management of COVID-19.<sup>94</sup> Given the rapidity with which new advanced medical therapies are becoming available in IBD, which quickly make current guidelines obsolete, living guidelines may offer a unique consideration to ensure applicability to routine care.

**Table 1.** Positioning of Biologics and Small Molecules for Moderate to Severe Ulcerative Colitis and Crohn's Disease

AGA recommendations	Gaps in guidance	Updated guidance
<b>UC</b>		
<b>Biologic-naïve</b>		
Infliximab or vedolizumab over adalimumab for biologic-naïve patients with UC	No recommendation on superiority of infliximab vs vedolizumab	Infliximab superior to vedolizumab, adalimumab, ustekinumab, ozanimod <sup>39,80,81</sup>
	No clear recommendation for ustekinumab	Vedolizumab superior to adalimumab <sup>15</sup>
	Unable to assess ozanimod	No demonstrated superiority among vedolizumab, ustekinumab, and ozanimod
<b>Biologic-exposed</b>		
Ustekinumab or tofacitinib over vedolizumab for infliximab-exposed patients with UC	Recommendation limited to infliximab-exposed patients with UC	Upadacitinib superior to tofacitinib, ustekinumab, vedolizumab, adalimumab, ozanimod in TNF-antagonist exposed patients with UC <sup>80,81</sup>
	No recommendation on superiority of ustekinumab vs tofacitinib	Ustekinumab superior to vedolizumab or tofacitinib in TNF-antagonist exposed patients with UC <sup>82</sup>
	No recommendation for upadacitinib or ozanimod in TNF-antagonist exposed	No demonstrated superiority among ozanimod, ustekinumab, and vedolizumab
<b>Combination therapy</b>		
Combining TNF-antagonist, vedolizumab, and ustekinumab therapy with thiopurines or methotrexate over biologic monotherapy in patients with UC	Limited comparative data for thiopurine or methotrexate combination therapy with vedolizumab and ustekinumab vs biologic monotherapy to inform statement	No evidence to support the routine use of thiopurines or methotrexate with vedolizumab or ustekinumab compared with monotherapy of biologics for UC <sup>83,84</sup>
<b>CD (luminal)</b>		
<b>Biologic-naïve</b>		
Infliximab, adalimumab, and ustekinumab are recommended and vedolizumab suggested over certolizumab for TNF-antagonist-naïve CD patients	Recommendations limited to TNF-antagonist exposure due to trial design No recommendation on superiority of infliximab or ustekinumab Unable to assess risankizumab or upadacitinib	Infliximab and risankizumab favored over ustekinumab, adalimumab, vedolizumab, and certolizumab <sup>43,85,88</sup> No demonstrated superiority between infliximab and risankizumab
<b>Biologic-exposed</b>		
In TNF-antagonist primary nonresponders ustekinumab is recommended and vedolizumab is suggested. In secondary TNF antagonist nonresponders, ustekinumab or adalimumab recommended and vedolizumab suggested	Recommendations limited to TNF-antagonist exposure due to trial design  Unable to assess risankizumab or upadacitinib	Risankizumab and upadacitinib favored over ustekinumab, adalimumab, and vedolizumab <sup>85</sup>  No demonstrated superiority between risankizumab and upadacitinib
<b>Combination therapy</b>		
In biologic and immunosuppressant-naïve CD, combination thiopurine or methotrexate recommended with TNF-antagonist therapy	No recommendation made for vedolizumab or ustekinumab	Addition of thiopurine to vedolizumab may reduce risk of treatment failure <sup>86</sup> No evidence to support the routine use of thiopurines or methotrexate with ustekinumab compared with ustekinumab monotherapy <sup>84</sup>

NOTE. JAK inhibitors (tofacitinib and upadacitinib) are not considered in biologic-naïve populations, given US Food and Drug Administration and European Medicines Agency labeling restricting the use of JAK inhibitors to biologic/TNF-antagonist-exposed individuals.

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at [www.gastrojournal.org](http://www.gastrojournal.org), and at <http://doi.org/10.1053/j.gastro.2023.10.033>.

### References

- Ha C, Ullman TA, Siegel CA, et al. Patients enrolled in randomized controlled trials do not represent the inflammatory bowel disease patient population. *Clin Gastroenterol Hepatol* 2012;10:1002–1007; quiz e78.
- Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health* 2017;20:858–865.
- Collins R, Bowman L, Landray M, et al. The magic of randomization versus the myth of real-world evidence. *N Engl J Med* 2020;382:674–678.
- Armstrong K. Methods in comparative effectiveness research. *J Clin Oncol* 2012;30:4208–4214.
- Eichler HG, Pignatti F, Schwarzer-Daum B, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. *Clin Pharmacol Ther* 2021;109:1212–1218.
- Uzzan M, Bouhnik Y, Abreu M, et al. Declining enrolment and other challenges in IBD clinical trials: causes and potential solutions. *J Crohns Colitis* 2023;17:1066–1078.
- Dubinsky MC, Collins R, Abreu MT. Challenges and opportunities in IBD clinical trial design. *Gastroenterology* 2021;161:400–404.
- Click B, Barnes EL, Cohen BL, et al. Methodology and initial results from a real-world observational cohort of patients with inflammatory bowel disease: TARGET-IBD. *Crohns Colitis* 360 2021;3(3):otab023.
- Dulai PS, Singh S, Jiang X, et al. The real-world effectiveness and safety of vedolizumab for moderate-severe Crohn's disease: results from the US VICTORY Consortium. *Am J Gastroenterol* 2016;111:1147–1155.
- Narula N, Peerani F, Meserve J, et al. Vedolizumab for ulcerative colitis: treatment outcomes from the VICTORY Consortium. *Am J Gastroenterol* 2018;113(9):1345.
- Johnson AM, Barsky M, Ahmed W, et al. The real-world effectiveness and safety of ustekinumab in the treatment of Crohn's disease: results from the SUCCESS Consortium. *Am J Gastroenterol* 2023;118:317–328.
- Peyrin-Biroulet L, Rahier JF, Kirchgessner J, et al. I-CARE, a European prospective cohort study assessing safety and effectiveness of biologics in inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2023;21:771–788.e10.
- Cohen-Mekelburg S, Wallace BI, Van T, et al. Association of anti-tumor necrosis factor therapy with mortality among veterans with inflammatory bowel disease. *JAMA Netw Open* 2021;4(3):e210313.
- Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Assoc* 2014;311:578–582.
- Sands BE, Peyrin-Biroulet L, Loftus EV Jr, et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. *N Engl J Med* 2019;381:1215–1226.
- Sands BE, Irving PM, Hoops T, et al. Ustekinumab versus adalimumab for induction and maintenance therapy in biologic-naïve patients with moderately to severely active Crohn's disease: a multicentre, randomised, double-blind, parallel-group, phase 3b trial. *Lancet* 2022;399(10342):2200–2211.
- Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147.
- Khanna R, Bressler B, Levesque BG, et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015;386(10006):1825–1834.
- Cipriani A, Higgins JP, Geddes JR, et al. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159:130–137.
- Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017;358:j3932.
- Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
- Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162:777–784.
- Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58.
- Chaimani A, Higgins JP, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013;8(10):e76654.
- Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol* 2018;93:36–44.
- Singh S, Murad MH, Fumery M, et al. First- and second-line pharmacotherapies for patients with moderate to severely active ulcerative colitis: an updated network meta-analysis. *Clin Gastroenterol Hepatol* 2020;18:2179–2191.e6.
- Feuerstein JD, Isaacs KL, Schneider Y, et al. AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology* 2020;158:1450–1461.
- Singh S, Allegretti JR, Siddique SM, et al. AGA technical review on the management of moderate to severe ulcerative colitis. *Gastroenterology* 2020;158:1465–1496.e17.
- Jackson D, White IR, Price M, et al. Borrowing of strength and study weights in multivariate and network meta-analysis. *Stat Methods Med Res* 2017;26:2853–2868.
- Cooper NJ, Peters J, Lai MC, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health* 2011;14:371–380.
- Caldwell DM, Dias S, Welton NJ. Extending treatment networks in health technology assessment: how far should we go? *Value Health* 2015;18:673–681.

32. Ma C, Guizzetti L, Jairath V. Improving clinical trial efficiency in gastroenterology. *Gastroenterology* 2019; 157:892–893.
33. Jairath V, Zou G, Parker CE, et al. Systematic review with meta-analysis: placebo rates in induction and maintenance trials of Crohn's disease. *Aliment Pharmacol Ther* 2017;45:1021–1042.
34. D'Haens G, Panaccione R, Baert F, et al. Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials. *Lancet* 2022;399(10340):2015–2030.
35. Ferrante M, Panaccione R, Baert F, et al. Risankizumab as maintenance therapy for moderately to severely active Crohn's disease: results from the multicentre, randomised, double-blind, placebo-controlled, withdrawal phase 3 FORTIFY maintenance trial. *Lancet* 2022; 399(10340):2031–2046.
36. Bauchner H, McDermott MM, Butte AJ. Data sharing enters a new era. *Ann Intern Med* 2023;176:400–401.
37. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Sci Data* 2018;5:180268.
38. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003;158:280–287.
39. Narula N, Wong ECL, Marshall JK, et al. Comparative efficacy for infliximab vs vedolizumab in biologic naive ulcerative colitis. *Clin Gastroenterol Hepatol* 2022; 20:1588–1597.e3.
40. Singh S, Fumery M, Sandborn WJ, et al. Systematic review and network meta-analysis: first- and second-line biologic therapies for moderate-severe Crohn's disease. *Aliment Pharmacol Ther* 2018;48:394–409.
41. Singh S, Murad MH, Fumery M, et al. Comparative efficacy and safety of biologic therapies for moderate-to-severe Crohn's disease: a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol* 2021;6:1002–1014.
42. Stidham RW, Lee TC, Higgins PD, 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;39:1349–1362.
43. Narula N, Wong ECL, Dulai PS, et al. Comparative effectiveness of biologics for endoscopic healing of the ileum and colon in Crohn's disease. *Am J Gastroenterol* 2022;117:1106–1117.
44. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26):1–134; iii–iv.
45. Saure D, Schacht A, Kadziola Z, et al. Combination of several matching adjusted indirect comparisons (MAICs) with an application in psoriasis. *Pharm Stat* 2020; 19:532–540.
46. Thom H, Jugl SM, Palaka E, et al. Matching adjusted indirect comparisons to assess comparative effectiveness of therapies: usage in scientific literature and health technology appraisals. *Value Health* 2016;19(3):A100–A101.
47. Lu X, Zhou Z-Y, Xin Y, et al. Matching-adjusted indirect comparisons of filgotinib vs vedolizumab, tofacitinib, and ustekinumab for moderately to severely active ulcerative colitis. *Inflamm Bowel Dis* 2024;30:64–77.
48. Loftus EV, Vermeire S, Feagan BG, et al. Corticosteroid-sparing effects of filgotinib in moderately to severely active ulcerative colitis: data from the Phase 2b/3 SELECTION Study. *J Crohns Colitis* 2023;17:211–220.
49. Reinisch W, Panaccione R, Duijvestein M, et al. P546 Comparative efficacy and safety of upadacitinib versus ustekinumab as induction therapy in patients with moderately to severely active ulcerative colitis: a matching-adjusted indirect comparison. *J Crohns Colitis* 2023;17(Suppl 1):i672–i673.
50. Dubinsky MC, Betts KA, LaPensee K, et al. S694 Comparative efficacy and safety of ozanimod vs adalimumab and vedolizumab in patients with moderately to severely active ulcerative colitis. *Am J Gastroenterol* 2021;116:S314.
51. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012; 15:940–947.
52. Narula N, Wong ECL, Dulai PS, et al. Vedolizumab and adalimumab in biologic-naïve ulcerative colitis: comparison of patient-level clinical trial data and VARSITY for week 6 clinical remission. *Dig Dis Sci* 2023; 68:2658–2666.
53. Fumery M, Defrance A, Roblin X, et al. Effectiveness and safety of risankizumab induction therapy for 100 patients with Crohn's disease: a GETAID multicentre cohort study. *Aliment Pharmacol Ther* 2023;57:426–434.
54. Meyer A, Fumery M, Peyrin-Biroulet L, et al. Comparative real-world effectiveness of vedolizumab and ustekinumab for patients with ulcerative colitis: a GETAID multicentre cohort study. *Scand J Gastroenterol* 2022; 57:1454–1462.
55. Koudsi M, Martinez-Vinson C, Pigneur B, et al. Ustekinumab use in pediatric inflammatory bowel disease: a French multicenter study from the pediatric GETAID. *J Pediatr Gastroenterol Nutr* 2023;76:763–770.
56. Boschetti G, Nachury M, Laharie D, et al. Efficacy and safety of infliximab retreatment in Crohn's disease: a multicentre, prospective, observational cohort (REGAIN) study from the GETAID. *Am J Gastroenterol* 2022; 117:1482–1490.
57. Fumery M, Filippi J, Abitbol V, et al. Effectiveness and safety of ustekinumab maintenance therapy in 103 patients with ulcerative colitis: a GETAID cohort study. *Aliment Pharmacol Ther* 2021;54:944–951.
58. Amiot A, Filippi J, Abitbol V, et al. Effectiveness and safety of ustekinumab induction therapy for 103 patients with ulcerative colitis: a GETAID multicentre real-world cohort study. *Aliment Pharmacol Ther* 2020;51:1039–1046.
59. Schmidt E, Kochhar G, Hartke J, et al. Predictors and management of loss of response to vedolizumab in inflammatory bowel disease. *Inflamm Bowel Dis* 2018; 24:2461–2467.
60. Meserve J, Aniwaniwan S, Koliiani-Pace JL, et al. Retrospective analysis of safety of vedolizumab in patients with inflammatory bowel diseases. *Clin Gastroenterol Hepatol* 2019;17:1533–1540.e2.

61. Faleck DM, Winters A, Chablaney S, et al. Shorter disease duration is associated with higher rates of response to vedolizumab in patients with Crohn's disease but not ulcerative colitis. *Clin Gastroenterol Hepatol* 2019;17:2497–2505.e1.
62. Kim HS, Hernaez R, Sangsriy S, et al. Comparative effectiveness of surveillance colonoscopy intervals on colorectal cancer outcomes in a national cohort of patients with inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2022;20:2848–2857.e2.
63. Hou JK, Imler TD, Imperiale TF. Current and future applications of natural language processing in the field of digestive diseases. *Clin Gastroenterol Hepatol* 2014;12:1257–1261.
64. Stidham RW, Liu Y, Enchakalody B, et al. The use of readily available longitudinal data to predict the likelihood of surgery in Crohn disease. *Inflamm Bowel Dis* 2021;27:1328–1334.
65. Cohen-Mekelburg S, Rosenblatt R, Gold S, et al. Fragmented care is prevalent among inflammatory bowel disease readmissions and is associated with worse outcomes. *Am J Gastroenterol* 2019;114:276–290.
66. Singh S, Kim J, Luo J, et al. Comparative safety and effectiveness of biologic therapy for Crohn's disease: a CA-IBD cohort study. *Clin Gastroenterol Hepatol* 2023;21:2359–2369.e5.
67. Gu P, Luo J, Kim J, et al. Effect of obesity on risk of hospitalization, surgery, and serious infection in biologic-treated patients with inflammatory bowel diseases: a CA-IBD cohort study. *Am J Gastroenterol* 2022;117:1639–1647.
68. Nguyen NH, Luo J, Paul P, et al. Effectiveness and safety of biologic therapy in hispanic vs non-hispanic patients with inflammatory bowel diseases: a CA-IBD cohort study. *Clin Gastroenterol Hepatol* 2023;21:173–181.e5.
69. Lukin D, Faleck D, Xu R, et al. Comparative safety and effectiveness of vedolizumab to tumor necrosis factor antagonist therapy for ulcerative colitis. *Clin Gastroenterol Hepatol* 2022;20:126–135.
70. Singh S, Fumery M, Sandborn WJ, et al. Systematic review with network meta-analysis: first- and second-line pharmacotherapy for moderate-severe ulcerative colitis. *Aliment Pharmacol Ther* 2018;47:162–175.
71. Duloy A, Yadlapati RH, Benson M, et al. Video-based assessments of colonoscopy inspection quality correlate with quality metrics and highlight areas for improvement. *Clin Gastroenterol Hepatol* 2019;17:691–700.
72. Dulai PS, Singh S, Vande Casteele N, et al. Development and validation of clinical scoring tool to predict outcomes of treatment with vedolizumab in patients with ulcerative colitis. *Clin Gastroenterol Hepatol* 2020;18:2952–2961.e8.
73. Dulai PS, Wong ECL, Reinisch W, et al. Decision support tool identifies ulcerative colitis patients most likely to achieve remission with vedolizumab vs adalimumab. *Inflamm Bowel Dis* 2022;28:1555–1564.
74. Dulai PS, Amiot A, Peyrin-Biroulet L, et al. A clinical decision support tool may help to optimise vedolizumab therapy in Crohn's disease. *Aliment Pharmacol Ther* 2020;51:553–564.
75. Dulai PS, Boland BS, Singh S, et al. Development and validation of a scoring system to predict outcomes of vedolizumab treatment in patients with Crohn's disease. *Gastroenterology* 2018;155:687–695.e10.
76. Dulai PS, Peyrin-Biroulet L, Demuth D, et al. Early intervention with vedolizumab on longer term surgery rates in Crohn's disease: post hoc analysis of the GEMINI phase 3 and long-term safety programs. *J Crohns Colitis* 2020;15:195–202.
77. Dulai PS, Wan Y, Huang Z, et al. Probability of response as defined by a clinical decision support tool is associated with lower healthcare resource utilization in vedolizumab-treated patients with Crohn's disease. *Crohns Colitis* 2022;4(4):otac048.
78. Park J, Chun J, Yoon H, et al. Feasibility of a clinical decision support tool for ustekinumab to predict clinical remission and relapse in patients with Crohn's disease: a multicenter observational study. *Inflamm Bowel Dis* 2023;29:548–554.
79. Alric H, Amiot A, Kirchgessner J, et al. Vedolizumab clinical decision support tool predicts efficacy of vedolizumab but not ustekinumab in refractory Crohn's disease. *Inflamm Bowel Dis* 2022;28:218–225.
80. Lasa JS, Olivera PA, Danese S, et al. Efficacy and safety of biologics and small molecule drugs for patients with moderate-to-severe ulcerative colitis: a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol* 2022;7:161–170.
81. Burr NE, Gracie DJ, Black CJ, et al. Efficacy of biological therapies and small molecules in moderate to severe ulcerative colitis: systematic review and network meta-analysis. *Gut* 2022;71:1976–1987.
82. Kochhar GS, Desai A, Farraye FA, et al. Efficacy of biologic and small molecule agents as second-line therapy after exposure to TNF inhibitors in patients with ulcerative colitis: a propensity-matched cohort study. *Aliment Pharmacol Ther* 2023;58:297–308.
83. Yzet C, Diouf M, Singh S, et al. No benefit of concomitant immunomodulator therapy on efficacy of biologics that are not tumor necrosis factor antagonists in patients with inflammatory bowel diseases: a meta-analysis. *Clin Gastroenterol Hepatol* 2021;19:668–679.e8.
84. Hu A, Kotze PG, Burgevin A, et al. Combination therapy does not improve rate of clinical or endoscopic remission in patients with inflammatory bowel diseases treated with vedolizumab or ustekinumab. *Clin Gastroenterol Hepatol* 2021;19:1366–1376.e2.
85. Barberio B, Gracie DJ, Black CJ, et al. Efficacy of biological therapies and small molecules in induction and maintenance of remission in luminal Crohn's disease: systematic review and network meta-analysis. *Gut* 2023;72:264–274.
86. Kirchgessner J, Desai RJ, Schneeweiss MC, et al. Decreased risk of treatment failure with vedolizumab and thiopurines combined compared with vedolizumab monotherapy in Crohn's disease. *Gut* 2022;71:1781–1789.
87. Feuerstein JD, Ho EY, Shmidt E, et al. AGA clinical practice guidelines on the medical management of moderate to severe luminal and perianal fistulizing Crohn's disease. *Gastroenterology* 2021;160:2496–2508.
88. Wong ECL, Dulai PS, Marshall JK, et al. Comparative efficacy of infliximab vs ustekinumab for maintenance of clinical response in biologic naïve Crohn's disease. *Inflamm Bowel Dis* 2023;29:1015–1023.

89. Ma C, Hanzel J, Panaccione R, et al. CORE-IBD: a multidisciplinary international consensus initiative to develop a core outcome set for randomized controlled trials in inflammatory bowel disease. *Gastroenterology* 2022;163:950–964.
90. Hawkey C, Avery A, Coupland CAC, et al. *Helicobacter pylori* eradication for primary prevention of peptic ulcer bleeding in older patients prescribed aspirin in primary care (HEAT): a randomised, double-blind, placebo-controlled trial. *Lancet* 2022;400(10363):1597–1606.
91. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 2022;328:2446–2447.
92. Akl EA, Meerpohl JJ, Elliott J, et al. Living systematic reviews: 4. living guideline recommendations. *J Clin Epidemiol* 2017;91:47–53.
93. El Mikati IK, Khabsa J, Harb T, et al. A framework for the development of living practice guidelines in health care. *Ann Intern Med* 2022;175:1154–1160.
94. Lamontagne F, Agarwal A, Rochweg B, et al. A living WHO guideline on drugs for covid-19. *BMJ* 2020; 370:m3379.

---

Received April 24, 2023. Accepted October 27, 2023.

#### Correspondence

Address correspondence to: Parambir S. Dulai, MD, Division of Gastroenterology and Hepatology, Feinberg School of Medicine,

Northwestern University, Arkes Pavilion, 676 N St Clair Street, 14<sup>th</sup> Floor, Chicago, Illinois 60611. e-mail: [Parambir.dulai@northwestern.edu](mailto:Parambir.dulai@northwestern.edu).

#### Credit Authorship Contributions

Parambir S. Dulai, MD (Conceptualization: Lead; Methodology: Lead; Supervision: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Siddharth Singh, MD (Conceptualization: Supporting; Methodology: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Vipul Jairath, MD, PhD (Methodology: Supporting; Writing – review & editing: Supporting)

Emily Wong, MS (Methodology: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Neeraj Narula, MD (Conceptualization: Supporting; Methodology: Supporting; Supervision: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting)

#### Conflicts of interest

These authors disclose the following: Parambir S. Dulai: consulting for Abbvie, Abivax, Adiso, Bristol Myers Squibb, GSK, Janssen, Lilly, Pfizer, and Takeda; grant support from Bristol-Myers Squibb, Janssen, Pfizer, and Takeda; Licensing royalties from PreciDiag. Siddharth Singh: research grants from AbbVie and Pfizer; personal compensation from Pfizer for ad hoc grant review. Vipul Jairath: consulting/advisory board: AbbVie, Alimentiv Inc, Arena Pharmaceuticals, Asahi Kasei Pharma, Asieris, Astra Zeneca, Avoro Capital, Bristol Myers Squibb, Celltrion, Eli Lilly, Endpoint Health, Ferring, Flagship Pioneering, Fresenius Kabi, Galapagos, Gilde Healthcare, GlaxoSmithKline, Genentech, Gilead, Innomar, JAMP, Janssen, Merck, Metacrine, Mylan, Pandion, Pendopharm, Pfizer, Protagonist, Prometheus Biosciences, Reistone Biopharma, Roche, Sandoz, SCOPE, Second Genome, Sorriso pharmaceuticals, Takeda, Teva, Topivert, Ventyx, and Vividion; speaker; Abbvie, Ferring, Bristol Myers Squibb, Galapagos, Janssen Pfizer Shire, Takeda, and Fresenius Kabi; financial support: salary support from the John and Susan McDonald Endowed Chair at Western University, London, Ontario, Canada. Neeraj Narula has been a speaker or advisor for Janssen, Abbvie, Takeda, Pfizer, Merck, Sandoz, Fresenius Kabi, Innomar Strategies, Iterative Health, Bristol Myers Squibb, Viatrix, Eli Lilly, and Ferring. The remaining author discloses no conflicts.

**Supplementary Table 1.** Current Advantages or Limitations of Randomized Trials and Opportunities for Improving Value of Real-World Evidence Synthesis for Inflammatory Bowel Disease to Guide Advanced Medical Therapy Integration

Limitation or advantage	Source	Impact	Solution	Value or opportunity of RWE data synthesis	RWE examples
<p>Enrollment eligibility: Limitations in number of prior treatment failures (mesalamine monotherapy failures, only 1–2 advanced therapy failures) and/or demographic characteristics of cohorts enrolled (eg, proctitis, pouch, ostomy, fistulae, and number of surgeries)</p>	<p>Industry: Attempts to minimize risk of failed trials or high AE reporting due to hard-to-treat populations with comorbid conditions or sub-groups with high anticipated placebo effect</p> <p>Academia: Lack of validity for current PRO and endoscopic measures for sub-populations (ie, ostomy, pouch, proctitis, and altered anatomy due to surgery)</p>	<p>Recruitment: Need to activate hundreds of sites, across dozens of countries, to identify eligible patients with increased reliance on certain regions (emerging markets)</p> <p>Coordination: Increased reliance on CROs with exponential impact on financial cost of trials and inefficient timelines</p>	<p>Disease measures: Develop new tools and/or modify or validate existing tools for sub-populations often excluded</p> <p>Drug discovery: Shift in drug discovery toward profiling of complex, refractory populations recognizing need for therapies specifically in biologic failure populations</p>	<p>Sub-populations: Allows for assessment of approved therapies in populations not included in trials to enhance generalizability of findings to routine practice</p> <p>Efficiency: Academic collaborations and consortiums to optimize patient identification, enrollment, and/or monitoring, minimizing cost</p>	<p>Target RWE<sup>8</sup>; US IBD Health Outcomes Consortium (VICTORY<sup>9,10</sup>, SUCCESS<sup>11</sup>); I-CARE and GETAID initiatives<sup>12</sup></p>
<p>Quality: RCTs provide high-quality assessments of demographic characteristics and measurement of PROs and endoscopy with central reading for inclusion/exclusion and measurement of treatment effect</p>	<p>Investigators: Misclassification by site investigators (consciously or subconsciously) for disease activity measures, particularly endoscopy, to increase recruitment in financially lucrative clinical trials</p> <p>Medical records: Lack of uniform data collection approaches to define patient demographic characteristics, disease activity measures, or monitoring</p> <p>PRO monitoring: Lack of tools for patient-centered data collection</p>	<p>Failed trials: Misclassification by local site investigators for endoscopic scoring has led to “negative” trial results for efficacious therapies</p> <p>Coordination: Need for centralized endoscopy scoring with added trial complexity burden and cost; feedback cycle of increased reliance on CRO</p>	<p>Routine endoscopy videos: Use of routine video recording (for adenoma detection) and allow for use of these videos for enrollment and activity assessments (centrally read or through artificial intelligence)</p> <p>Uniform data collection: Creation of EMR-based note templates or NLP approaches to harmonize data across practice locations and link to claims data sets</p> <p>Patient self-reporting: Increased utilization of PRO monitoring directly by patients</p>	<p>Central monitoring: Integrated health systems with uniform data collection and/or multisite endoscopic recording will allow for closed system completion of large observational studies to answer important questions (eg, biosimilar switch; sequencing studies) not feasible for traditional clinical trials due to cost and complexity of designs</p>	<p>Veterans Affairs assessment of treatment integration, efficacy, safety<sup>13</sup>; integrated EMR data warehouses (PCORNet)<sup>14</sup></p>

AE, adverse event; CRO, contract research organization; EMR, electronic medical record; NLP, natural language processing; PRO, patient-reported outcomes.

**Supplementary Table 2.** Strengths and Limitations of Different Nonrandomized Sources of Data for Integration of Advanced Medical Therapies for Inflammatory Bowel Diseases

Network meta-analysis		Post-hoc analysis using individual participant data		Routine practice observational cohorts	
Strengths	Limitations	Strengths	Limitations	Strengths	Limitations
Highest-quality evidence from well-controlled RCTs, level 1 evidence readily accepted by payors	Unable to account for evolutions in end-point measurement timing and method of assessment (central vs local endoscopy scoring), as well as definitions used	Use of well-characterized uniformly monitored cohorts from RCTs with an ability to partially overcome variations in timing and definitions of assessments for comparison	Not readily accepted by payors or societal guidelines, limiting impact on advanced medical therapy implementation	Readily accepted by providers (face validity of effectiveness and safety) and increasing acceptance by payors (particularly for off-label dosing) and regulatory authorities (21 <sup>st</sup> Century Cures Act)	Often not well characterized or uniformly monitored, with risks for data missingness or inaccurate end-point measurement due to natural clinical variations in follow-up and assessment
Rapidly updated as new therapies are approved	Potential limitations in transivity or comparability in trial populations	IPD-level data allows for sub-selection of intervention cohorts most closely matched	Limited by data availability and delays in IPD being uploaded to public platforms	Rapid data generation as therapies come to market for populations not studied in clinical trials	Comparability in populations limited by payor positioning, provider decisions on use, and patient acceptance
More precise than single direct or indirect estimate <sup>30,31</sup>	Covariates cannot be adjusted for at the individual participant level	Precise adjustments for covariates and confounders related to efficacy outcomes	Cohort size and event rates limit ability to fully adjust for covariates in relation to secondary outcomes of interest	Large cohort sizes with an ability to apply statistical adjustments for balancing across multiple outcomes including rare events	Missingness of data creates inability to fully adjust for key covariates and increased risk for unmeasured confounder effects in comparisons
Deal with multiple therapies and synthesize evidence for a disease	Unable to account for differences in trial design or conduct to provide clear guidance on all comparisons	Able to account for differences in trial design by “reconstructing” data sets to be comparable and better inform key comparisons	Unable to deal with multiple therapies simultaneously for comparisons	Deal with multiple therapies and comparisons, across wide-ranging populations from varying regions	Unable to account for access barriers, or other external influences that are controlled for in RCTs
Provides comparisons for interventions not previously assessed head to head	Unable to inform safety due to low event rates	Provides comparisons for interventions not previously assessed head to head	Unable to inform safety due to low event rates	Informs safety comparisons and assessment of any new safety signals	Provides comparative effectiveness estimates, but with lower confidence