




Utility of routinely collected electronic health records data to support effectiveness evaluations in inflammatory bowel disease: a pilot study of tofacitinib

Vivek Ashok Rudrapatna ^{1,2}, Benjamin Scott Glicksberg ^{1,3},
Atul Janardhan Butte ^{1,4}

To cite: Rudrapatna VA, Glicksberg BS, Butte AJ. Utility of routinely collected electronic health records data to support effectiveness evaluations in inflammatory bowel disease: a pilot study of tofacitinib. *BMJ Health Care Inform* 2021;**28**:e100337. doi:10.1136/bmjhci-2021-100337

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2021-100337>).

Received 05 February 2021
Revised 07 April 2021
Accepted 26 April 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, California, USA

²Division of Gastroenterology, Department of Medicine, University of California, San Francisco, CA, USA

³Mount Sinai Medical Center, New York, New York, USA

⁴Center for Data-Driven Insights and Innovation, University of California Health System, Oakland, California, USA

Correspondence to
Dr Atul Janardhan Butte;
atul.butte@ucsf.edu

ABSTRACT

Objectives Electronic health records (EHR) are receiving growing attention from regulators, biopharmaceuticals and payors as a potential source of real-world evidence. However, their suitability for the study of diseases with complex activity measures is unclear. We sought to evaluate the use of EHR data for estimating treatment effectiveness in inflammatory bowel disease (IBD), using tofacitinib as a use case.

Methods Records from the University of California, San Francisco (6/2012 to 4/2019) were queried to identify tofacitinib-treated IBD patients. Disease activity variables at baseline and follow-up were manually abstracted according to a preregistered protocol. The proportion of patients meeting the endpoints of recent randomised trials in ulcerative colitis (UC) and Crohn's disease (CD) was assessed.

Results 86 patients initiated tofacitinib. Baseline characteristics of the real-world and trial cohorts were similar, except for universal failure of tumour necrosis factor inhibitors in the former. 54% (UC) and 62% (CD) of patients had complete capture of disease activity at baseline (month -6 to 0), while only 32% (UC) and 69% (CD) of patients had complete follow-up data (month 2 to 8). Using data imputation, we estimated the proportion achieving the trial primary endpoints as being similar to the published estimates for both UC (16%, *p* value=0.5) and CD (38%, *p*-value=0.8).

Discussion/Conclusion This pilot study reproduced trial-based estimates of tofacitinib efficacy despite its use in a different cohort but revealed substantial missingness in routinely collected data. Future work is needed to strengthen EHR data and enable real-world evidence in complex diseases like IBD.

INTRODUCTION

Real-world evidence (RWE) refers to the use of observational data to support inference on diseases and treatments. This area has been growing for a variety of reasons,¹⁻⁴ including (1) rising costs and other challenges to the feasibility of randomised trials,⁵ (2) concerns

Summary

What is already known?

- Real-world data (RWD) are receiving increasing attention from regulators, payors and biopharmaceuticals as an emerging source of evidence on treatment effects.
- Although electronic health records (EHR) data are an important and granular source of RWD, their suitability for real-world evidence remains unknown in part due to their complexity.
- Tofacitinib was recently approved for the inflammatory bowel disease (IBD-subtype ulcerative colitis), but its effectiveness and safety in real-world cohorts who may not meet trial eligibility criteria is unclear.

What does this paper add?

- Although EHR data contain much of the data needed to assess treatment effectiveness in IBD, we found these data to be less accessible (primarily found within free text) and associated with significant missing values at baseline and follow-up.
- We propose an approach for estimating real-world treatment effects from these data using data abstraction protocols and methods for stochastic imputation of missing data.
- Although a majority of the patients treated at our centre did not meet the eligibility criteria corresponding to randomised trials of tofacitinib in IBD, its effectiveness appeared to be the same as that measured in the trials.

that trial cohorts may be unrepresentative of real-world patients^{6,7} and (3) the emergence of new datasets and methods for assessing treatment in routine clinical contexts.

Of the sources of real-world data (RWD) that are being explored for this purpose, electronic health records (EHR) are receiving particular attention. They have served as the primary ledger for clinical encounters over two decades and capture rich data on

exposures and outcomes. However, this optimism has been tempered by several challenges.¹ Beyond limitations common to observational settings (eg, confounding, mismeasurement), EHR data is commonly captured in free text rather than a tabular format. This creates a challenge for the study of diseases whose assessments may be captured in narratives (eg, patient-reported outcomes). Such data typically require the use of text processing, methods that can achieve high accuracy but remain laborious. However, the utility of pursuing these approaches remains unclear because the availability of the underlying data (ie, disease activity scores) in free text is understudied.

An example of a disease currently assessed by complex measures is inflammatory bowel disease (IBD). IBD is comprised of two subtypes, ulcerative colitis (UC) and Crohn's disease (CD). Treatment involves immunosuppression that is usually continued until treatment failure (eg, inadequate efficacy, adverse events). In trials, effectiveness is measured according to the Mayo Score and Crohn's Disease Activity Index (CDAI) for UC and CD respectively.

The first small molecule approved for IBD is tofacitinib. Tofacitinib induced clinical remission in 18.5% of the 476 participants with UC who were treated for 8 weeks (OCTAVE 1) and maintained remission in 34.3% of the 197 induction responders assigned to 52 weeks of continued treatment.⁸ Tofacitinib was also evaluated in phase 2b randomised controlled trials (RCTs) of CD.⁹ In these trials, 43% of the 86 patients randomised to the 10 mg arm achieved clinical remission following induction (week 8) and 39.5% of the 60 induction responders assigned to the 5 mg arm achieved response or remission at week 26. However, unlike for UC, tofacitinib did not show statistical superiority to placebo for CD and thus was never approved for that indication. Nonetheless, it has sometimes been used off-label for CD.

In this pilot study, we assessed the utility of EHR data for treatment effectiveness evaluations in a cohort of patients with IBD treated with tofacitinib. Our primary objective was to assess disease activity data at timepoints roughly corresponding to the trial endpoints. An exploratory objective was to estimate tofacitinib's effectiveness using EHR data and compare it with the trials. Other exploratory objectives included characterising differences in patient cohorts, time-to-treatment-failure and the reasons for treatment failure.

METHODS

This retrospective cohort study of patients with IBD treated with tofacitinib was performed according to the STROBE and RECORD guidelines (online supplemental file 1).

Patient identification

We directly queried an existing database derived from all EHR records at the University of California, San Francisco

(UCSF). This previously described database¹⁰ contains records from 6/2012 (instantiation of the Epic EHR) through 4/2019 (query date) and includes diagnoses, procedures, demographics and medications. Eligible records met these criteria: (1) age over 18 years, (2) a tofacitinib order and (3) a gastroenterologist-assigned IBD diagnosis code (ICD-10-CM K50*/K51*) (table 1). Records meeting the above informatics criteria were all manually reviewed to identify a cohort of all adult patients at UCSF who had (1) ever been prescribed tofacitinib for the treatment of IBD and (2) initiated treatment.

Study endpoints

The primary endpoint was the proportion of patients with complete measurements of the Mayo Score and CDAI at baseline and follow-up. For this pilot study, baseline was defined as month -6 to 0 relative to the start date of tofacitinib, and follow-up was defined as month 2 to 8. These time-windows were chosen to reflect typical patterns of data collection in clinical practice while also allowing for rough comparisons to the timepoints assessed in trials.

An exploratory endpoint was the proportion of patients meeting the endpoints as defined by the OCTAVE trials⁸ in UC and the CD trials by Panés *et al*⁹ (see 'Comparison to trial endpoints' below). Other exploratory endpoints included characterising differences in patient cohorts, time-to-treatment-failure, and the reasons for treatment failure.

Disease activity scores

The Mayo score is scored on a 0-12 scale corresponding to the sum of four equally weighted subscores. The CDAI ranges from 0 to over 600; it incorporates three patient-reported outcomes, comorbidities, weight, haematocrit and medication use. In the gastroenterology clinic at UCSF, elements of these scores are individually captured in clinical narratives as relevant to the provision of routine care; these are not captured as structured data (eg, 'smartforms').

Data quality, completeness, and handling of missing data

We assessed the quality of the data in detail prior to proceeding with downstream analysis. We annotated missing data and characterised its distribution (figures 1 and 2). The proportion of patients with complete capture of the Mayo score and CDAI at baseline and follow-up were computed (primary endpoint). We also computed the proportion of non-missing data elements taken as a whole.

We handled missing data using a model-based approach, which relies on the data meeting the missing at random assumption. This was deemed plausible because (1) the clinical decision to pursue additional testing is typically dictated by the results of other correlated data and the risks/benefits of additional studies, and (2) we collected a wide range of auxiliary variables that inform clinical decision making (see 'Covariate abstraction').

Table 1 Baseline demographics

	OCTAVE induction 1 (n=475)	Sample of UC cohort (n=28)
Male sex, n (%)	277 (58.2)	16 (57)
Age, years	41.3±14.1	43.2±14.4
Duration of disease, years		
Median	6.5	10.2
Range	0.3–42.5	2.2–51.4
Extent of disease, n/total n (%)		
Proctosigmoiditis	64/475 (13.7)	3/28 (10.7)
Left-sided colitis	158/475 (33.3)	6/28 (21.4)
Extensive colitis/pancolitis	252/475 (53.1)	19/28 (67.9)
Total Mayo score	9.0±1.4	8.5±1.8
Partial Mayo score	6.3±1.2	6±1.6
CRP, mg/L		
Median	4.4	5.8
Range	0.1–208.4	0.8–70.6
Glucocorticoid use at baseline*	214 (45.0)	17 (60.7)
Previous treatment with TNF inhibitor, n (%)	254 (53.4)	28 (100)
Previous treatment failure, n (%)		
TNF inhibitor	243 (51.1)	28 (100)
Glucocorticoid	350 (73.5)	24 (85.7)
Immunosuppressant	360 (75.6)	21 (75)
	Panés <i>et al</i> ⁹ (n=86)	Sample of CD cohort (n=13)
Female, n (%)	47 (54.7)	9 (69.2)
Age, years		
Mean (SD)	39.3 (13.7)	39.7 (19.5)
Weight, kg		
Mean (SD)	71.6 (18.8)	69.9 (16.3)
Race, n (%)		
White	72 (83.7)	9 (69.2)
Black	2 (2.3)	0 (0)
Asian	11 (12.8)	1 (7.7)
Others	1 (1.2)	0 (0)
Duration since CD diagnosis, years		
Mean (SD)	11.3 (9.7)	14.4 (8.2)
Extent of disease, n (%)		
L1 (Ileal)	7 (8.1)	1 (7.7)
L1/4 (Ileal + Upper GI)	2 (2.3)	2 (15.4)
L2 (Colonic)	5 (5.8)	0 (0)
L2/4 (Colonic + Upper GI)	16 (18.6)	1 (7.7)

Continued

Table 1 Continued

	OCTAVE induction 1 (n=475)	Sample of UC cohort (n=28)
L3 (Ileocolonic)	15 (17.4)	4 (30.8)
L3/4 (Ileocolonic)	39 (45.3)	5 (38.5)
Prior use of TNF inhibitor, n (%)	66 (76.7)	13 (100)
Use of corticosteroids at study entry, n (%)	28 (32.6)	7 (53.8)
Baseline CDAI score		
Mean (SD)	320 (61.66)	374 (183.73)
Baseline CRP, mg/L		
Median (min-max)	5.5 (0.2–126)	28.7 (3.5–107)

Within each pair of columns, the left column corresponds to the patient demographics of the tofacitinib-assigned arm in corresponding RCTs (eg, the OCTAVE trials reported by Sandborn *et al*, the CD trials reported by Panés *et al*). The right column reports the corresponding demographics of a sample of tofacitinib-treated patients at UCSF. Treatment failure is defined as an inadequate response to any treatment (eg, steroids, TNF inhibitor) as defined and documented by the treating clinician. CDAI, Crohn's Disease Activity Index; CRP, C-reactive protein; RCTs, randomised controlled trials; TNF, tumour necrosis factor; UCSF, University of California, San Francisco.

We performed multiple imputation by chained equations using random forest models (online supplemental file 1). These methods have a lower false discovery rate than last-observation-carried-forward,¹¹ a method commonly used in IBD trials.

Covariate abstraction

Patient records were reviewed via the clinician-facing interface, which contains all clinical data, including notes, patient-provider messaging, procedure reports and laboratory results (online supplemental eTable 1). The EHR contains all clinical data generated within UCSF as well as that shared from other health systems during clinical care.

All patients were assessed by the time-to-treatment-failure, defined as either a lack of efficacy or a significant adverse event recognised by both the clinician and the patient (figure 2). This variable was distinguished from treatment non-compliance defined as a patient-initiated discontinuation against medical advice. This was separately measured during abstraction and was found to be available for all patients (online supplemental file 1). Patients who had not failed treatment at the time of data abstraction were treated as having had non-informatively censored events. Treatment discontinuation due to loss of insurance coverage as well as relocation or other lost-to-follow-up events were rare and were treated as non-informatively censored.

A random sample of the patient records in this study was selected for abstraction of the remaining variables. This

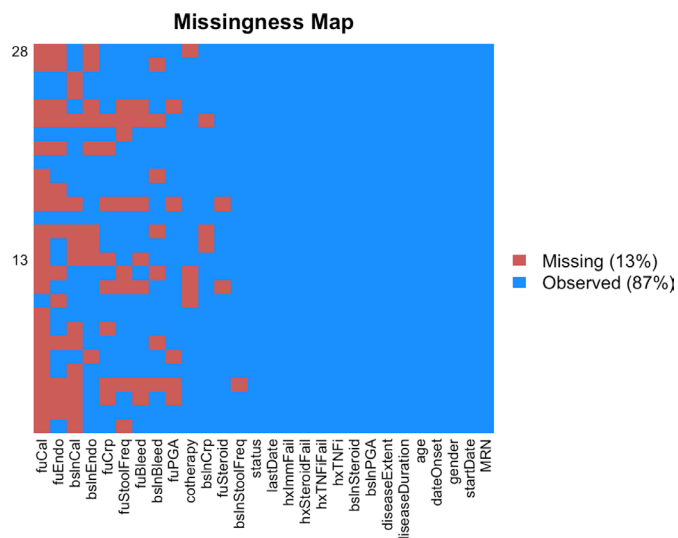


Figure 1 Distribution of missing data in the ulcerative colitis dataset. Variables are listed on the x-axis in order of decreasing missingness. Each row in the y-axis corresponds to a different patient. Variable abbreviations correspond to the following: MRN=medical record number; dateOnset=date of disease onset; diseaseDuration=length of disease; diseaseLocation=location of disease by Montreal classification; startDate=date of treatment initiation; lastDate=date of last known use of treatment; status=0 if still using tofacitinib at last date, 1 if no longer using tofacitinib at last date; cotherapy=use of other concomitant therapies (eg, mesalamine, curcumin, simple carbohydrate diet); bslnCrp=baseline C reactive protein; fuCrp=follow up C reactive protein; bslnCal=baseline faecal calprotectin; fuCal=follow-up faecal calprotectin; bslnSteroid=baseline corticosteroid use; fuSteroid=follow-up corticosteroid use; bslnStoolFreq=baseline Mayo stool frequency subscore; fuStoolFreq=follow-up Mayo Stool Frequency subscore; bslnBleed=baseline Mayo Rectal Bleeding subscore; fuBleed=follow-up Mayo Rectal Bleeding subscore; bslnPGA=baseline Mayo Physician Global Assessment subscore; fuPGA=follow-up Mayo Physician Global Assessment subscore; bslnEndo=baseline Mayo Endoscopic subscore; fuEndo=follow-up Mayo Endoscopic subscore.

was done to strike a balance between estimating parameters with reasonable precision and the effort required for this manual review process (32 and 47 variables per record for UC and CD). The full list and definition of these variables is available in the protocol (online supplemental file 1).

CDAI elements incorporating an average daily rating over 7 days were calculated by extrapolating from a single day's mention within the time windows mentioned above. This decision was made based on accepted practices of the handling missing CDAI diary data in registrational trials (eg, UNITI-1 Statistical Analysis Plan section 5.2.1¹²) and the methods used to derive the CDAI.¹³

Comparison to trial outcomes

An exploratory endpoint of this study involved estimating the proportion of patients meeting the endpoint of the trials. As mentioned, a follow-up window of months 2–8

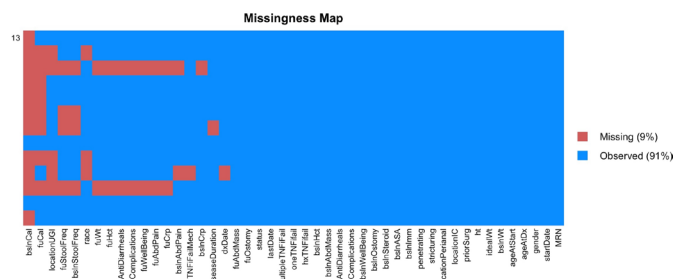


Figure 2 Distribution of missing data in the Crohn's disease dataset. Variables are listed on the x-axis in order of decreasing missingness. Each row in the y-axis corresponds to a different patient. Covariate abbreviations are as follows: 'bsln' and 'Fu' prefixes refer to variable at baseline or at follow-up; MRN=medical record number; dxDate=diagnosis date; startDate=date of treatment initiation; lastDate=date of last known use of treatment; status=0 if still using tofacitinib at last date, 1 if no longer using tofacitinib at last date; ageAtDx=age at diagnosis; ageAtStart=age at treatment start; Wt=weight; idealWt=ideal wt; ht=height; priorSurg=history of prior gastrointestinal surgery; locationLC=disease location in the lower gastrointestinal tract; locationPerianal=presence of disease in the perianal region; locationUGI=disease location in the upper gastrointestinal tract; Ostomy=presence of an ostomy; Imm=use of immunomodulators; ASA=use of aminosaliclates; steroid=use of corticosteroids; complications=complications CDAI subscore; wellbeing=wellbeing CDAI subscore; AbdPain=abdominal pain CDAI subscore; penetrating=penetrating disease behaviour; structuring=structuring disease behaviour; Hct=haematocrit; hxTNFiFail=history of TNF inhibitor failure; oneTNFiFail=history of only one prior TNF inhibitor failure; multipleTNFiFail=history of multiple TNF inhibitor failures; TNFiFailMech=classification of TNF inhibitor failure. CDAI, Crohn's Disease Activity Index; TNF, tumour necrosis factor.

after treatment was used to assess disease activity after initiating treatment. This window was chosen to account for the typical follow-up time in practice, but does not precisely match either the induction or maintenance endpoint times for either OCTAVE (weeks 8 and 52) or the corresponding CD trials⁹ (weeks 8 and 26).

Because our timepoint more closely matched that of maintenance than of induction, and because each trial only assessed remission among those achieving treatment response following induction (ie, others were assumed to be maintenance-phase non-responders), we recomputed the maintenance endpoint probability as the probability of induction patients being eligible for the maintenance trial by the probability of maintenance response among those enrolled. This probability was statistically compared with the endpoint probabilities in the UCSF-cohort.

These binary endpoints were computed using the same definitions as those in the corresponding trials. For UC, this was the proportion with a total Mayo score ≤ 2 , no individual subscore greater than 1 and a rectal bleeding subscore of 0. For CD, this corresponded to the probability that a patient had either achieved a 100-point

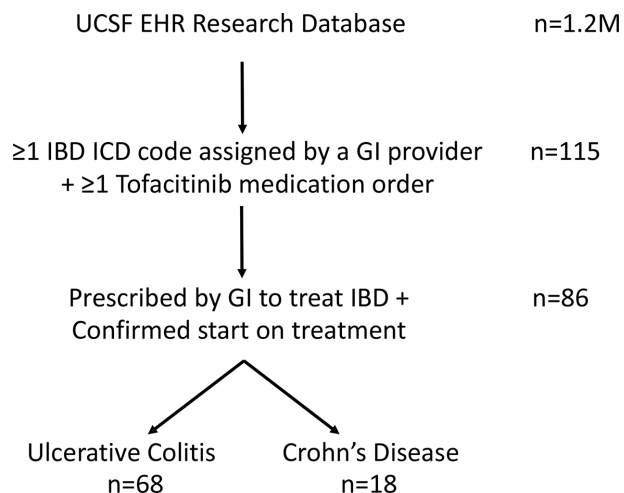


Figure 3 Cohort selection schematic. EHR, electronic health records; GI, gastrointestinal; IBD, inflammatory bowel disease; UCSF, University of California, San Francisco.

reduction in the CDAI from baseline or an absolute CDAI less than 150 at follow-up.

Statistics/computing

We computed point estimates and performed hypothesis testing using Wald test statistics with pooled standard errors.¹⁴ For analyses comparing the probability of remission in the real-world cohort with that of the RCTs, the prespecified null hypothesis was these two probabilities were equal. We estimated the time-to-treatment-failure survival distributions using the product-limit estimator. No competing events were observed. Code written in R was independently reviewed by a co-author. Data and analysis files were version-controlled using Docker.

RESULTS

Cohort identification

We identified 115 patient records following a query of our EHR database. Manual review confirmed that 86 patients—68 with UC and 18 with CD—had initiated tofacitinib specifically to treat IBD (figure 3). The other 29 patients were excluded during this process for multiple reasons, including failure to start treatment due to payor denial, the decision to forgo the ordered medical treatment in favour of surgery and treatment initiated by a non-gastroenterologist for another autoimmune condition. Non-compliance, defined as patient-initiated discontinuation of tofacitinib against medical advice, was rare (4%) in this cohort.

Data completeness

Out of 28 patients with UC randomly sampled for full assessment of the Mayo score and all other auxiliary variables at baseline and follow-up, 15 (54%) had a complete capture of the Mayo score at baseline and 9 (32%) had a complete capture at follow-up. The least available subscore was endoscopy (figure 1). With respect to the partial Mayo score, 21 (75%) and 17 (61%) were complete

at these timepoints. Out of 13 patients with CD sampled, 8 (62%) had complete capture of the CDAI at baseline and 9 (69%) had this available at follow-up.

Taken as a proportion of the total number of collected variables, 13% of the UC-related data and 9% of the CD-related were missing (figures 1 and 2). These missing data were handled by multiple imputation.

Cohort characterization

The baseline demographics of the subjects under study in the UCSF and RCT cohorts were similar (table 1). Notable differences include the universal failure of TNF inhibitors in the UCSF cohort, as well as a longer duration of disease in the patients with UC. Patient groups had similar baseline Mayo scores, C-reactive protein levels and prevalent corticosteroid use. Sixty-one per cent of the cohort had been using corticosteroids at baseline. Thirty-nine per cent of the cohort used at least one form of additional IBD treatment: these included mesalamine, curcumin and dietary changes.

Zero per cent of the patients with UC initiated on tofacitinib met the eligibility criteria of the corresponding phase 3 RCT.⁸ The reasons for this were multifactorial (online supplemental eTable 2) but include use of vedolizumab within the previous year, high-dose steroids at the time of treatment initiation and the possibility of requiring surgery during the treatment period.

We separately explored what proportion of patients met the specific RCT entry criteria defined by the Mayo score and CDAI for UC and CD, respectively. Ninety-three per cent (73–98) of the patients with UC had an eligible baseline Mayo score (6–12), whereas 50% (19–82) of the patients with CD had a baseline CDAI within the eligibility range of the corresponding RCT (220–450).

Effectiveness and safety

Time-to-treatment-failure analysis on the full cohort revealed similar survival distributions irrespective of IBD disease subtype (online supplemental eFigure 1). The overall probability of incident users continuing tofacitinib long-term was 68% (58%–80%). All failure events occurred within the first 7 months; among continued responders by month 6, the probability of sustained absence of treatment failure was 94%. Of note, the first use of the tofacitinib occurred in 2013, and the longest duration of effectiveness data relevant to treatment maintenance was 3.7 years.

We assessed the reasons for treatment failure (online supplemental eFigure 2). In the UC cohort, there were 17 treatment failure events: 12 with insufficient treatment efficacy, 4 with adverse events/intolerances and 1 due to patient preference. Of the 12 efficacy failures, 8 patients (67%) contained evidence of ongoing inflammation on the basis of biomarkers, imaging or lower endoscopy performed within the 2-month period prior to treatment failure. All patients who did not undergo objective confirmation of inflammation during this timeframe did have objective evidence of inflammation prior to treatment

Table 2 Potential approaches to strengthen routinely collected electronic health records data and better support real-world evidence studies

Problem	Example	Potential solutions
Complex and cumbersome disease activity scores limit practical use	The CDAI incorporates a comprehensive list of elements but only some apply to any given patient (eg, abdominal pain predominant, fistula predominant). Elements that are not relevant to a given patient are unlikely to be captured during routine clinical care	Develop and validate novel scores that accurately capture disease activity, are easy to administer and capture in real-world contexts and are relevant to different patient subgroups
Data capture by free text rather than structured data capture (eg, EHR smartforms)	<ul style="list-style-type: none"> ▶ Typing out clinical narratives is faster and more natural to clinicians than point-and-click interfaces ▶ These narratives are relatively inaccessible for RWE studies (requires natural language processing) and are subject to ambiguous documentation ▶ Unclear if current, documentation-oriented reimbursement schemes are compatible with smartform-entered data 	<ul style="list-style-type: none"> ▶ More streamlined and relevant scores as above ▶ Partnership between clinical, quality, operations, IT, user experience and research teams to optimise data capture ▶ Payers may be able to incentivise better data capture in support of outcome-based and risk-adjusted reimbursement
Patient-oriented and decision-oriented data capture rather than cohort-oriented data capture	<p>Patient 1 has a colonoscopy showing severe endoscopic disease. A precise characterisation and documentation of current patient symptoms is irrelevant to treatment decision making.</p> <p>Patient 2 has worsening symptoms and a rise in biochemical markers consistent with prior flares. The decision is made to change treatment without additional testing (eg, enterography, colonoscopy)</p> <p>Patients 1 and 2 individually have sufficient data to support personalised decision making, but collectively have inadequate data to support cross-cutting RWE studies of treatment outcomes</p>	<ul style="list-style-type: none"> ▶ Partnership between clinical, quality, operations, IT, user experience and research teams to optimise data capture ▶ Payers may be able to incentivise a shift towards and improved measurement of healthcare quality and disease-oriented population health management
Encounters are not well-timed relative to important clinical events	<p>Week -5: Patient is seen in clinic and agrees to switch therapy. Symptoms and disease activity captured in the EHR.</p> <p>Week 0: Patient fills prescription and begins treatment as an outpatient.</p> <p>Week 7: Patient returns for follow-up.</p> <p>Results: (1) No symptom capture at the time of treatment initiation, (2) Week 7 follow-up might not align with data capture of other patients</p>	<ul style="list-style-type: none"> ▶ Clinic-level harmonisation of practices concerning the timing of patient encounters and follow-up ▶ Use of interactive remote technologies (mobile apps, chatbots) to generically increase the frequency of data capture or time data capture ▶ Payers incentivise patients to participate in disease tracking (lower premiums), potentially in collaboration with pharmacies or infusion centres (optimally timed capture)
Encounter presence/absence correlated with clinical outcomes	<p>Patient 1 is feeling well 8 weeks after starting tofacitinib and is on a high-deductible plan. She does not want to take time off from work to go to clinic or pay the copay when she has no current clinical needs.</p> <p>Patient 2 is not feeling well 8 weeks after starting tofacitinib. He stops taking the medication and does not follow-up because he does not think the clinicians can help him.</p>	<ul style="list-style-type: none"> ▶ Use of interactive remote technologies (mobile apps, chatbots) to increase touchpoints with the clinic, improve trust and avoid the time and monetary expenses of a clinic visit ▶ Supplementing EHR data and supporting the function of the clinic with staff-initiated outreach (eg, an EHR/RWD-augmented registry)

CDAI, Crohn's Disease Activity Index; EHR, electronic health records; RWE, real-world evidence.

initiation. All but one patient with inadequate response completed a minimum of 7 weeks of treatment induction (11 weeks on average) prior to the adjudication of treatment failure.

In the CD cohort, there were five treatment failure events: one due to an adverse event (zoster) and one due to insufficient efficacy (all with concomitant objective evidence of ongoing inflammation). These patients completed 13.2 weeks of treatment on average.

Twenty-two per cent of all subjects participating in the induction phase of the UC RCT⁸ met the primary maintenance endpoint of week 52 clinical remission. We observed a similar response (16%) in the corresponding UCSF cohort (6%–37%, *p* value=0.5). Similarly, the proportion achieving the primary endpoint in the CD RCT⁹ (34%) was similar to the point estimate of the real-world cohort (38%, *p*-value=0.8).

We explored the extent to which steroid use may account for some of these results. In the UC cohort, 33% of patients had been using steroids at the time of follow-up. Among the patients who had been using steroids at baseline, 56% were steroid-free at the time of follow-up.

DISCUSSION

We assessed the completeness of routinely collected EHR data to support RWE studies of diseases with complex activity measures. Taking a use case of tofacitinib as used to treat IBD (both on-label and off-label), we found that the capture of the total Mayo score and the CDAI is currently modest at best, even at a tertiary-care medical centre.

On exploratory analyses, the real-world effectiveness of this drug appeared to be consistent with its published effectiveness from randomised trials despite its use in a substantially different cohort. We found that patients with IBD using tofacitinib appear to generally tolerate it well and that unlike biologics commonly used for IBD, secondary loss of response events for this small molecule was uncommon.

RWD has been receiving growing interest from a variety of parties including the FDA³ and EMA,¹⁵ biopharmaceuticals and payors. Despite this interest, it must be recognised that not all RWD are created equal. Unlike prospectively planned disease and treatment registries, the EHR data capture mechanism has historically been designed with other objectives in mind: healthcare coordination and delivery, revenue generation and medico-legal documentation among others.

Our pilot study highlights the substantial work that will be needed to close the quality gap between retrospective EHR data and prospective data and realise the promise of RWE. We outline the root causes of this quality gap as well as outline potential solutions in [table 2](#). Many of these solutions will ultimately require a close partnership between the many stakeholders in real-world clinical care: clinicians, patients, health IT, operations and especially

payors. Undoubtedly, this may require a significant investment in both time and money by these participants. However, we are of the opinion that the eventual rewards are worth the investment. These include the ability to better measure the quality of care, discover practice-changing evidence and enable continuous-improving learning health systems.

Strengths of this study include the use of a preregistered protocol and analysis plan, the use of rigorous methods for handling missing data,¹¹ as well as openly available code accompanied by deidentified raw EHR data in order to maximise the reproducibility and reusability of this work. The primary limitation of this work lies in its inability to draw inferences related to the real-world effectiveness of tofacitinib.

CONCLUSION

Routinely collected EHR data currently has uneven capture of the data needed to optimally assess IBD treatment effectiveness at baseline and follow-up. This work provides several insights into real-world practice, including typical patterns of data collection and the real-world effectiveness and safety of tofacitinib for IBD. It also offers an analytical approach to the analysis of missing real-world data. Future efforts are needed to improve inference from these data, such better data capture mechanisms and novel measures more suitable to routine care.

Twitter Vivek Ashok Rudrapatna @vivicality, Benjamin Scott Glicksberg @BenGlicksberg and Atul Janardhan Butte @atulbutte

Acknowledgements The authors thank the UCSF Academic Research Services and Clinical Data Research Consultation services for clinical informatics support. The authors would like to acknowledge Dana Ludwig for his help in deidentifying and interpreting the UCSF EHR.

Contributors VAR and AB conceived the project. VAR designed the chart review protocol, performed chart extraction, conducted statistical analyses and drafted this manuscript. BG performed code review and critically edited this manuscript. AB supervised the project and critically edited this manuscript.

Funding Research reported in this publication was supported by funding from the UCSF Bakar Computational Health Sciences Institute and the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1 TR001872. VAR was supported by the National Institute of Diabetes and Digestive and Kidney Disease of the National Institutes of Health grant under award T32 DK007007.

Disclaimer The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Competing interests Atul Butte is a co-founder and consultant to Personalis and NuMedii; consultant to Samsung, Mango Tree Corporation, and in the recent past, 10x Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Facebook, Alphabet (Google), Microsoft, Amazon, Snap, Snowflake, 10x Genomics, Illumina, Nuna Health, Assay Depot (Scientist.com), Vet24seven, Regeneron, Sanofi, Royalty Pharma, Pfizer, BioNTech, AstraZeneca, Moderna, Biogen, Twist Bioscience, Pacific Biosciences, Editas Medicine, Invitae, and Sutro, and several other non-health related companies and mutual funds; and has received honoraria and travel reimbursement for invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, several investment and venture capital firms, and many academic institutions, medical or disease specific foundations and associations, and health systems. Atul Butte receives royalty payments through

Stanford University, for several patents and other disclosures licensed to NuMedii and Personalis. Atul Butte's research has been funded by NIH, Northrup Grumman (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervallien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity.

Patient consent for publication Not required.

Ethics approval We obtained Institutional Review Board approval to obtain patient data and abstract all covariates.

Provenance and peer review Commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The analytic code in the form of a R markdown file as well as the accompanying data set needed to reproduce the analysis in this work are available in a Docker container to all investigators without restriction (<https://doi.org/10.7272/Q6PZ5715>). These individual participant data were de-identified to comply with the US Department of Health and Human Services 'Safe Harbor' guidance and applicable laws and regulations concerning privacy and/or security of personal information. The data dictionary is documented within the study protocol section of Supplemental Content.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Vivek Ashok Rudrapatna <http://orcid.org/0000-0003-1789-3004>

Benjamin Scott Glicksberg <http://orcid.org/0000-0003-4515-8090>

Atul Janardhan Butte <http://orcid.org/0000-0002-7433-2740>

REFERENCES

- 1 Rudrapatna VA, Butte AJ. Opportunities and challenges in using real-world data for health care. *J Clin Invest* 2020;130:565–74.
- 2 Chatterjee A, Chilukuri S, Fleming E. Real-World evidence: driving a new drug development paradigm in oncology, 2018. Available: https://www.mckinsey.com/~/media/mckinsey/industries/pharmaceuticals_and_medical_products/our_insights/real_world_evidence_driving_a_new_drug_development_paradigm_in_oncology/real-world-evidence-driving-a-new-drug-development-p [Accessed 2 Jul 2019].
- 3 Framework for FDA's Real-World Evidence Program, 2018. Available: www.fda.gov [Accessed 26 Jun 2019].
- 4 Medicines Agency E. An agency of the European Union Regulatory Perspective on Real World Evidence (RWE) in scientific advice EMA Human Scientific Committees' Working Parties with Patients' and Consumers' Organisations (PCWP) and Healthcare Professionals' Organisations (HCPWP).
- 5 Moore TJ, Zhang H, Anderson G, *et al*. Estimated costs of pivotal trials for novel therapeutic agents Approved by the US food and drug administration, 2015–2016. *JAMA Intern Med* 2018;178:1451–7.
- 6 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;10:1002–7.
- 7 Zoccali C, Blankestijn PJ, Bruchfeld A, *et al*. Children of a lesser God: exclusion of chronic kidney disease patients from clinical trials. *Nephrol Dial Transplant* 2019;34:1112–4.
- 8 Sandborn WJ, Su C, Sands BE, *et al*. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2017;376:1723–36.
- 9 Panés J, Sandborn WJ, Schreiber S, *et al*. Tofacitinib for induction and maintenance therapy of Crohn's disease: results of two phase IIb randomised placebo-controlled trials. *Gut* 2017;66:1049–59.
- 10 Norgeot B, Glicksberg BS, Trupin L, *et al*. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. *JAMA Netw Open* 2019;2:e190606.
- 11 Buuren S van. flexible imputation of missing data.
- 12 Feagan BG, Sandborn WJ, Gasink C, *et al*. Ustekinumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med* 2016;375:1946–60.
- 13 Best WR, Beckett JM, Singleton JW, *et al*. Development of a Crohn's disease activity index. National cooperative Crohn's disease study. *Gastroenterology* 1976;70:439–44.
- 14 Rubin DB, Wiley J, York N. Multiple imputation for nonresponse in surveys, 1987. Available: <https://www.onlinelibrary.wiley.com/doi/pdf> [Accessed 20 Jun 2019].
- 15 Cave A, Cerreta F. Use of real world data in development programmes, 2017. Available: https://www.ema.europa.eu/en/documents/presentation/presentation-use-real-world-data-development-programmes-dr-alison-cave-dr-francesca-cerreta_en.pdf [Accessed 26 Jun 2019].