Validation and update of a prediction model for risk of relapse after cessation of anti-TNF treatment in Crohn's disease

Huinink, Sebastiaan ten Bokkel; de Jong, Djuna C.; Nieboer, Daan; Thomassen, Doranne; Steyerberg, Ewout W.; Dijkgraaf, Marcel G. W.; Bodelier, Alexander G. L.; West, Rachel L.; Romkens, Tessa E. H.; Hoentjen, Frank

Published in:
European journal of gastroenterology & hepatology

DOI:
10.1097/MEG.0000000000002403

Validation and update of a prediction model for risk of relapse after cessation of anti-TNF treatment in Crohn's disease

Sebastiaan ten Bokkel Huinink[a], Djuna C. de Jong[b], Daan Nieboer[c], Doranne Thomassen[d], Ewout W. Steyerberg[c,d], Marcel G.W. Dijkgraaf[e], Alexander G.L. Bodelier[f], Rachel L. West[g], Tessa E.H. Römkens[h], Frank Hoentjen[i], Rosalie C. Mallant[j], Bas A.C. van Tuyl[k], Wout G.N. Mares[l], Frank H.J. Wolfhagen[m], Gerard Dijkstra[n], Jurriën G.P. Reijnders[o], Nanne K. de Boer[p], Adriaan C.I.T.L. Tan[q], Petra G.A. van Boeckel[r], Greetje J. Tack[s], Dirk P. van Asseldonk[t], Geert R.A.M. D'Haens[b], C Janneke van der Woude[a], Marjolijn Duijvestein[b,p] and Annemarie C de Vries[a]

Background Anti-tumor necrosis factor (TNF) therapy is effective for the treatment of Crohn's disease. Cessation may be considered in patients with a low risk of relapse. We aimed to externally validate and update our previously developed prediction model to estimate the risk of relapse after cessation of anti-TNF therapy.

Methods We performed a retrospective cohort study in 17 Dutch hospitals. Crohn's disease patients in clinical, biochemical or endoscopic remission were included after anti-TNF cessation. Primary outcome was a relapse necessitating treatment. Discrimination and calibration of the previously developed model were assessed. After external validation, the model was updated. The performance of the updated prediction model was assessed in internal-external validation and by using decision curve analysis.

Results 486 patients were included with a median follow-up of 1.7 years. Relapse rates were 35 and 54% after 1 and 2 years. At external validation, the discriminative ability of the prediction model was equal to that found at the development of the model [c-statistic 0.58 (95% confidence interval (CI) 0.54–0.62)], though the model was not well-calibrated on our cohort [calibration slope: 0.52 (0.28–0.76)]. After an update, a c-statistic of 0.60 (0.58–0.63) and calibration slope of 0.89 (0.69–1.09) were reported in internal-external validation.

Conclusion Our previously developed and updated prediction model for the risk of relapse after cessation of anti-TNF in Crohn's disease shows reasonable performance. The use of the model may support clinical decision-making to optimize patient selection in whom anti-TNF can be withdrawn. Clinical validation is ongoing in a prospective randomized trial. Eur J Gastroenterol Hepatol 34: 983–992

Introduction

Anti-tumor necrosis factor (anti-TNF) therapy is frequently prescribed as induction and maintenance treatment in moderate to severe Crohn's disease (CD) [1–3]. A majority of CD patients receive long-term anti-TNF therapy to maintain remission. However, exposure to anti-TNF therapy is associated with significant disadvantages, including side effects, such as infections, an increased risk of malignancy [4–6], chronic fatigue [7–9], work-productivity loss [8,9] and significant healthcare costs [10,11].

In daily practice, cessation of anti-TNF therapy in CD patients in remission is still debated. Anti-TNF therapy is infrequently withdrawn mainly due to the uncertainty of

European Journal of Gastroenterology & Hepatology 2022, 34:983–992

Keywords: anti-TNF therapy, cessation, Crohn's disease, prediction

[a]Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, [b]Department of Gastroenterology and Hepatology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, [c]Department of Public Health, Erasmus Medical Center, Rotterdam, [d]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, [e]Department of Epidemiology and Data Science, Amsterdam UMC, University of Amsterdam, Amsterdam, [f]Department of Gastroenterology and Hepatology, Amphia Hospital, Breda, [g]Department of Gastroenterology and Hepatology, Franciscus Gasthuis & Vlietland, Rotterdam, [h]Department of Gastroenterology and Hepatology, Jeroen Bosch Hospital, Den Bosch, [i]Department of Gastroenterology and Hepatology, Radboud University Medical Center, Nijmegen, [j]Department of Gastroenterology and Hepatology, Flevo Hospital, Almere, [k]Department of Gastroenterology and Hepatology, Diakonessenhuis Utrecht, Utrecht, [l]Department of Gastroenterology and Hepatology, Gelderse Vallei Hospital, Ede, [m]Department of Gastroenterology and Hepatology, Albert Schweitzer Hospital, Dordrecht, [n]Department of Gastroenterology and Hepatology, University Medical Center Groningen, Groningen, [o]Department of Gastroenterology and Hepatology, Haga Hospital, Den Haag, [p]Department of Gastroenterology and Hepatology, AG&M Research Institute, Amsterdam University Medical Centre, Vrije Universiteit Amsterdam, Amsterdam, [q]Department of Gastroenterology and Hepatology, Canisius Wilhelmina Hospital, Nijmegen, [r]Department of Gastroenterology and Hepatology, St. Antonius Hospital, Nieuwegein, [s]Department of Gastroenterology and Hepatology, Medical Centre Leeuwarden, Leeuwarden and [t]Department of Gastroenterology and Hepatology, NWZ Alkmaar, Alkmaar, The Netherlands

Correspondence to Annemarie de Vries, MD, PhD, Department of Gastroenterology and Hepatology, Erasmus Medical Center, PO Box 2040, 3000 CA, Rotterdam, The Netherlands

Tel: +31 10 7040704; e-mail: a.c.devries@erasmusmc.nl

Received 25 November 2021 **Accepted** 16 May 2022

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.eurojgh.com.

the risk of relapse in the individual CD patient [12,13]. A more personalized treatment approach, including a prediction model for anti-TNF cessation, will benefit the individual CD patient and the healthcare system at large. Therefore, a stratification tool to identify patients who can safely cease anti-TNF therapy can be clinically useful.

Until recently, the model developed in the infliximab discontinuation in Crohn's disease patients in stable remission on combined therapy withimmunosuppressors (STORI) trial by GETAID has been the only available prediction model with a reported predictive power (concordance statistic and c-statistic) of 0.71 in the original article [14]. However, external validation of the infliximab discontinuation in Crohn's disease patients in stable remission on combined therapy with immunosuppressors (STORI) model in an individual patient data meta-analysis (IPD-MA) by our group (CEASE phase 0) on 14 cohorts ($n = 1317$), showed a less robust discriminative ability, with a c-statistic of 0.51. Based on this IPD-MA, a prediction model was developed to safely cease anti-TNF therapy with a reported c-statistic of 0.58 in internal-external validation. After an update of the prediction model with fecal calprotectin, an improved c-statistic of 0.63 was reported in a subgroup analysis.

In the current study (CEASE phase I), we aimed to validate and update the previously developed prediction models in a large independent Dutch CD cohort.

Material and methods

Study design

We performed a multicentre, retrospective cohort study (CEASE phase 1 cohort), in 17 hospitals in the Netherlands (five academic and 12 general teaching hospitals). CD patients who discontinued anti-TNF therapy between August 2019 and January 2000 were included in this study. CD patients were identified either from medical records through a search in the electronic patient database or the available medical lists from the hospital pharmacy using the keywords 'Crohn's disease', 'anti-TNF therapy', 'infliximab' and 'adalimumab'.

Patients were included between July 2019 and January 2020. Included patients received anti-TNF therapy (adalimumab or infliximab) ≥6 months for the primary indication of luminal CD. Included patients had to be in remission at the moment of discontinuation of anti-TNF, and concomitant treatment with immunomodulatory therapy was allowed. Remission was defined as either clinical, biochemical or endoscopic remission. Due to the infrequent availability of standardized tools to quantify disease activity (i.e. Harvey-Bradshaw Index or Crohn's Disease Activity Index), clinical remission was defined as the absence of symptoms based on the global assessment and documentation of the treating physician. Biochemical remission was defined as the absence of biochemical markers of inflammation [C-reactive protein (CRP < 5 mg/l and fecal calprotectin <250 µg/g)]. Endoscopic remission was defined as the absence of macroscopic inflammation (erosions or ulcerations), based on the findings in the endoscopy report. Patients were excluded if they ceased anti-TNF therapy primarily due to other reasons (e.g. infections or side effects), or if a top-down strategy was applied where patients received anti-TNF therapy less than 6 months.

Sample size

For external validation, at least 100 events are required to reliably estimate the performance of a prediction model [15]. We assumed that a minimum of 20% of the included patients would relapse within the follow-up time. Therefore, to include at least 100 events (relapses) the estimated sample size was 500 patients for the full cohort. Based on the model performance in phase 0, the required sample size for a full re-estimation of the phase 0 fecal calprotectin model was calculated as well [15]. To obtain a shrinkage factor of 0.85, an estimated sample of 487 patients was needed. This sample size also satisfies the second and third criterion outlined by Riley *et al.* [15]: a small difference between apparent and adjusted Nagelkerke R^2 (<0.05); and a precise estimate of overall risk [95% confidence interval (CI) width <0.1]). Hence, a sample size of 500 patients was expected to provide sufficient statistical power for external validation and a model update.

Endpoints

The primary endpoint was the proportion of documented relapses, defined as a relapse of luminal disease activity or the occurrence of (new) CD complications [i.e. extra-intestinal manifestations (EIM), (perianal) fistula with or without an abscess] that necessitated the introduction of additional treatment, including biologicals, corticosteroids, immune-suppressants or surgery. Clinical relapse was defined as the presence of symptoms, such as abdominal pain, diarrhea, perianal fistulas or the presence of EIM (e.g. arthritis, uveitis, erythema nodosum, pyoderma gangrenosum). A biochemical relapse was defined as CRP ≥5 mg/l and/or fecal calprotectin ≥250 µg/g. An endoscopic relapse was defined as the presence of macroscopic inflammation at endoscopy (i.e. erosions and/or ulcerations), as interpreted by the endoscopist. The secondary endpoint was the sustained effect of retreatment with the same or other anti-TNF agent. The sustained clinical benefit of retreatment with anti-TNF therapy was considered successful if patients were still treated with this agent 12 months after their relapse.

Data collection

Electronic patient records were retrospectively reviewed. Information on patient characteristics, disease-specific information, biochemical markers and endoscopic data were obtained, at the moment of stopping anti-TNF therapy (baseline) and at relapse. At baseline, biochemical markers and endoscopic appearance were recorded if they were obtained 1 year before, or 2 months after discontinuation of anti-TNF therapy.

Previously collected data

In addition to the data collected in CEASE phase 1, we used data from phase 0 of the CEASE project in some analyses. We refer to the CEASE phase 0 cohorts as development cohorts, because the model was initially developed within these cohorts. A detailed report of the study and patient characteristics of the CEASE phase 0 cohorts was provided [16], in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis statement [17]. The cohort of phase 0 was based on several international cohorts,

including one Dutch cohort [18]. There were minor differences in the inclusion and exclusion criteria between phase 0 and phase 1, including the duration of anti-TNF exposure (12 months vs. 6 months, respectively). We included the same IPD that Pauwels *et al.*[15] used in the development of the phase 0 model, with the addition of patients between 13 and 15 years old, resulting in a total number of 1330 IPD from the phase 0 cohorts in our analyses.

The validation cohort refers to the cohort described in this article, phase 1 of the CEASE project, based on patients from Dutch hospitals. We externally validated the previously developed model on the validation (phase 1) cohort only. After external validation, the development (phase 0) and validation (phase 1) cohorts were combined to perform the model update.

Previously developed model

The previously developed model (phase 0 model) is a Cox regression model that includes the following predictors: younger age at diagnosis (HR = 1.5 for A1 vs. A2), age at cessation (HR = 1.2 per 10 years younger), upper gastrointestinal tract involvement (HR = 1.3 for L4 vs. non-L4), clinical symptoms at cessation (HR = 2.2), smoking (HR = 1.4), longer disease duration (HR = 1.07 per 5 years), no concomitant immunosuppressant's (HR = 1.4), second-line anti-TNF (HR = 1.3), adalimumab (HR = 1.22 for adalimumab vs. infliximab) and CRP (HR = 1.04 per doubling) (Table 1).

In a subgroup analysis, fecal calprotectin was added to the model as a predictor (phase 0 fecal calprotectin model). This improved the discriminative ability (c-statistic 0.63) (Table 1). Low fecal calprotectin levels were associated with a favorable outcome after anti-TNF therapy cessation, which is in line with available literature [14,19].

Statistical analysis

Reporting on this study was performed according to the TRIPOD statement [20]. Statistical analyses were performed using IBM SPSS Statistics version 26.0 and R version 4.0.3 [21]. Descriptive statistics were provided with frequencies and percentages for qualitative variables and medians and interquartile ranges (IQRs) for quantitative variables. We assumed missing values were missing at random and imputed missing values using the mice algorithm [22]. Kaplan–Meier analysis was used to quantify the crude risk of relapse after cessation of anti-TNF therapy.

External validation of the previously developed model

To evaluate differences in case-mix between the development (CEASE phase 0) and validation (CEASE phase 1) samples, we compared the distributions of predicted 1-year relapse risk in the respective samples [23]. All predictions were calculated using the exact formulae of the previously developed models. The previous models were developed in a meta-analysis stratified by cohort, meaning that each development (CEASE phase 0) cohort had its own baseline hazard estimate. One phase 0 cohort consisted of Dutch patients, so we assumed the baseline hazard of this cohort in our predictions for the validation cohort.

The discrimination and calibration of the previously developed models with and without fecal calprotectin were assessed in the validation cohort. The discriminative ability of the models was quantified using Harrell's c-statistic [24]. The c-statistic ranges from 0 to 1, where 0.5 indicates that the prediction model is equivalent to a coin flip, whereas a value of 1 indicates perfect discrimination. Calibration was assessed graphically using calibration plots, which were characterized by the calibration slope and calibration-in-the-large. In our calibration plots, the validation cohort is divided into five groups defined by predicted event rate quintiles. The observed event rates in these groups are plotted against the predicted event rates, which ideally should lie on the 45-degree line. The calibration-in-the-large compares the average predicted risk to the average observed risk and is equal to zero in case of perfect agreement. The calibration slope measures whether predictor effects are on average correct and should ideally be equal to one.

Model update

After the external validation of the previously developed models, a model update was performed. In this update,

Table 1. Hazard ratios and associated 95% confidence intervals of the CEASE prognostic models

0Predictor	Previously developed model (phase 0) HR (95% CI)	Previously developed model with fecal calprotectin (phase 0) HR (95% CI)	Updated model (phase 1) HR (95% CI)
Age, per 10 years	0.86 (0.75–1.00)	0.90 (0.71–1.14)	0.94 (0.86–1.03)
Smoking, yes vs. no	1.39 (1.15–1.67)	1.52 (1.10–2.08)	1.31 (1.11–1.53)
Age at diagnosis, A2 vs. A1	0.69 (0.53–0.90)	0.46 (0.30–0.72)	
Age at diagnosis, A3 vs. A1	0.71 (0.40–1.25)	0.74 (0.29–1.92)	
Age at diagnosis, per 5 years			0.94 (0.90–1.00)
L4 Upper GI, yes vs. no	1.32 (0.96–1.79)	1.62 (0.98–2.70)	1.15 (0.89–1.49)
Disease duration, every 5 years	1.07 (0.98–1.17)	1.02 (0.90–1.16)	
Immunosuppressant, yes vs. no	0.70 (0.58–0.85)	0.87 (0.61–1.23)	0.68 (0.58–0.79)
Type of anti-TNF used, IFX vs. ADA	0.82 (0.67–1.01)	0.96 (0.66–1.41)	0.87 (0.74–1.03)
Second-line anti-TNF, yes vs. no	1.32 (1.01–1.69)	1.72 (1.09–2.70)	1.13 (0.92–1.39)
Clinical remission, yes vs. no	0.45 (0.25–0.83)	0.31 (0.16–0.58)	
CRP, per doubling, *mg/L*	1.04 (1.00–1.08)	1.00 (0.94–1.08)	1.02 (0.98–1.07)
FC, per doubling, *μg/g*		1.13 (1.02–1.27)	1.10 (1.05–1.16)
Apparent c-statistic	0.59	0.63	0.61
Internal-external validation pooled c-statistic	0.58 (0.55–0.61)[a]	0.63 (0.59–0.67)[b]	0.60 (0.58–0.63)[c]

ADA, adalimumab; CRP, C-reactive protein; FCP, Fecal Calprotectin; GI, gastrointestinal; HR, hazard ratio; IFX, infliximab
[a]On the phase 0 data.
[b]On a subset of the phase 0 data.
[c]On the combined phase 0 and phase 1 data.

fecal calprotectin was included in the model, the continuous version of age at diagnosis replaced the Montreal A classification, and disease duration and clinical remission were removed (due to linear dependence on age and age at diagnosis and an extremely small number of patients without clinical remission in the phase 1 cohort, respectively). The model was then refitted in an IPD-MA on the combined cohorts of CEASE phase 0 and CEASE phase 1. To check the validity of combining the data, we statistically tested for differences in effects between phase 0 and phase 1 cohorts using a model with interaction terms. In addition, we performed cross-validation, where the updated model was fitted on the phase 1 cohort and validated on the phase 0 cohorts, and vice versa.

Validation of the updated model

The resulting updated prediction model (phase 1 model) was validated using internal-external validation. Internal-external validation is a procedure where every cohort is left out once so that a model can be developed on the remaining cohorts and validated on the cohort that was left out [25]. In these validations, the discriminative ability was assessed using the c-statistic, and calibration was quantified using calibration-in-the-large and calibration slope. A pooled c-statistic, calibration-in-the-large and calibration slope were estimated with a random-effects model. Heterogeneity across studies in performance measures was quantified by the I^2 statistic [26]. 95% CI of the pooled performance measures were calculated.

Decision curve analysis was used to assess the clinical usefulness of the updated prediction model [27,28]. In a decision curve analysis, the ability of a prediction model to select patients for ceasing treatment is compared to the default strategies of continuing or stopping anti-TNF treatment in all patients. The net benefit of using a prediction model for patient selection is calculated by summing the benefits (correctly identifying patients who would relapse within 1 year) and subtracting the harms (continuing treatment in patients who would not relapse), using a weighting factor related to the corresponding threshold probability. The weighting factor expresses the number of patients one is willing to continue anti-TNF treatment, to correctly identify one patient who would relapse within 1 year. The net benefit of using the CEASE phase 1 model was investigated for a range of clinically relevant threshold probabilities.

Implementation of the updated model

Finally, we constructed a prognostic tool as a web interface to present a user-friendly version of the updated CEASE model to predict the 1-year risk of relapse after cessation of anti-TNF treatment, which will become available upon publication of this article.

Ethical approval and patient consent

The ethical committee of the Amsterdam University Medical Centers approved this study (reference number W19_100#19.130). The Institutional Review Board waived the need for an informed consent procedure. Instead, patients were actively informed about the study and were given the right to 'opt-out'. At the Medical Centre Leeuwarden a written informed consent procedure was performed.

Results

Baseline characteristics

In total, 7226 CD patients who met the search criteria were screened for eligibility. Of these patients, 486 were included in the final analysis (Fig. 1). Median follow-up time after cessation of anti-TNF therapy was 1.7 years (IQR 0.8–3.1). Baseline characteristics are shown in Table 2. A total of 129 patients (27%) were previously exposed to adalimumab or infliximab. Totally 132 patients (27%) underwent prior intestinal resection, of which the majority were ileocolonic resections ($n = 99$; 75%). After cessation of anti-TNF therapy, concomitant therapy was maintained in 176 patients (36%; thiopurines, $n = 153$, methotrexate, $n = 23$).

A colonoscopy report was available in 192 patients (39%), with procedures performed at a median time of 0.6 months before the cessation of anti-TNF therapy (IQR 0.2–2.4 months). The absence of any endoscopic inflammation was documented in 90% ($n = 172$). In the other 20 patients (10%), the mild disease was observed with signs of erosions in 10 (50%) and/or ulcerations in five patients (25%), which was contained to the ileum in 45% ($n = 9$). Despite mild disease activity on endoscopy, these patients were included as they were in either clinical and/or biochemical remission.

Relapse

In total, 277 patients (57%) experienced a relapse after cessation of anti-TNF therapy [including clinical relapse $n = 220$ (79%), biochemical relapse $n = 130$ (47%), endoscopic relapse $n = 118$ (43%), relapse confirmed by imaging other than colonoscopy $n = 31$ (11%)], with median time to relapse of 0.8 years (IQR 0.4–1.7, Table 3). Relapse rates were 35% (31–39%) and 54% (49–59%) after 1 and 2 years, respectively (Fig. 2).

In 20 patients (7%), recurrence of perianal disease was the indication to restart treatment. However, we did not observe the development of a new perianal fistula after the cessation of anti-TNF. Seven patients (3%) required reintroduction of anti-TNF therapy due to EIM.

In 31 patients (11%), the diagnosis of relapse and reintroduction of therapy was not supported by objective measures of inflammation (either biochemical analysis, endoscopy or imaging) and was solely based on patients' symptoms.

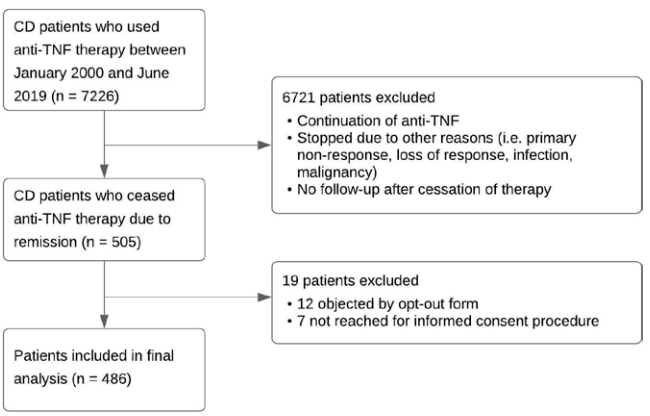

Fig. 1. Flow chart of screening and inclusion.

Table 2. Baseline characteristics ($n = 486$)

Factor	*N* (%) or median (IQR)
Follow-up time, years	1.7 (0.8–3.1)
Age, years	37.9 (29.1–50.3)
Female	309 (64)
Active smoker ($n = 326$)	94 (29)
Disease duration, years	9.1 (4.7–17.1)
Montreal classification ($n = 374$)	
Age at diagnosis	
A1 ≤16 years	45 (12)
A2 17–40 years	273 (73)
A3 >40 years	56 (12)
Location	
L1 Terminal ileum	96 (26)
L2 Colon	90 (24)
L3 Ileocolonic	185 (9)
+ L4 Upper GI	28 (7)
Behavior	
B1 Non-stricturing, non-penetrating	242 (5)
B2 Stricturing	70 (19)
B3 Penetrating	71 (19)
Perianal disease	102 (27)
Prior intestinal resection	132 (27)
Previously treated with anti-TNF	129 (27)
Anti-TNF used	
Adalimumab	237 (49)
Infliximab	249 (51)
Schedule adalimumab	
Every other week	203 (86)
Every week	14 (6)
Interval lengthened	20 (8)
Schedule infliximab	
Standard (8 weeks)	219 (88)
Intensified (6 weeks)	13 (5)
Other	17 (7)
Duration anti-TNF therapy, years	4.1 (2.0–6.6)
Concomitant medication continued after cessation of anti-TNF	
Thiopurine	153 (31)
Methotrexate	23 (5)
Biochemical (within 1 year before, or 1 month after stop anti-TNF)	
Hemoglobin, mmol/L ($n = 431$)	8.6 (8.0–9.1)
Leukocytes, * 10^9/L ($n = 422$)	7.3 (6.0–9.1)
Thrombocytes, * 10^3/mm^3($n = 414$)	268 (227–320)
Albumin, g/L ($n = 213$)	40 (37–44)
CRP, mg/L ($n = 408$)	3 (1–5)
Calprotectin, mg/kg ($n = 249$)	43 (16–136)
Trough level, µg/ml ($n = 147$)	3.2 (1.0-6.4)
Colonoscopy ($n = 192$)	
No signs of inflammation	172 (90)
Signs of inflammation	20 (10)

CRP, C-reactive protein; TNF, tumor necrosis factor.

Treatment after relapse

In total 174 patients (63%) were treated with a biologic agent after they experienced a relapse, most of whom restarted anti-TNF therapy ($n = 162$, 93%). However, 98 patients (56%) started adalimumab and 64 patients (37%) infliximab, of whom 15% received combination therapy with azathioprine ($n = 20$) or methotrexate ($n = 4$). Also, 133 patients (82.1%) restarted the same anti-TNF treatment that was ceased before (infliximab; $n = 60$ and adalimumab; $n = 73$). In the remaining 12 patients, vedolizumab ($n = 9$, 5%) and ustekinumab ($n = 3$, 2%) were started.

Median follow-up after relapse was 1.8 years (IQR, 0.8–3.3). Seventeen patients (10%) experienced primary nonresponse or loss of response after reintroduction of anti-TNF, and eight patients (5%) ceased their anti-TNF due to side effects. In total 104 patients (81%) achieved either clinical response or remission after anti-TNF was restarted. Retreatment with the same anti-TNF agent was effective (exposure >12 months) in 85 and 82% for infliximab and adalimumab, respectively. However, 33 patients did not have a sufficient follow-up period to conclude whether anti-TNF was effective after reintroduction. However, these patients were still receiving this treatment at the end of the follow-up period (median follow-up 0.5 years, IQR, 0.2–0.8), suggesting they responded sufficiently.

External validation of the previously developed model (CEASE phase 0 model)

Comparability of the development and validation cohorts

We refer to Pauwels *et al.* [16] for details on the baseline characteristics in the development (CEASE phase 0) cohorts. For the previously developed model without fecal calprotectin, the median predicted 1-year risk of relapse was 0.31 (IQR, 0.23–0.41) in the development sample compared to 0.30 (IQR, 0.24–0.37) in the validation sample. The model with fecal calprotectin generated a median predicted 1-year risk of 0.35 (IQR, 0.26–0.48) in its development sample and 0.27 (IQR 0.21–0.39) in the validation sample. The distributions of predicted 1-year relapse risk were reasonably similar across development and validation samples (see Appendix 1).

Performance of the previously developed model in the validation cohort

At external validation, the c-statistic of the previously developed model without fecal calprotectin was 0.58 (95% CI, 0.54–0.62) (Fig. 3a). The previously developed model with fecal calprotectin had similar performance [c = 0.57 (95% CI, 0.53–0.61)] (Fig. 3b). There was reasonable agreement between the predicted and observed relapse rates, though both phase 0 models under-predicted risk for low-risk patients and over-predicted risk for high-risk patients in our data. On average, the predictor effects were too strong, as indicated by calibration slopes below one [calibration slopes were 0.52 (95% CI, 0.28–0.76) and 0.31 (95% CI, 0.13–0.49) for the phase 0 model and the phase 0 fecal calprotectin model, respectively]. The calibration-in-the-large [0.14 (95% CI, –0.01 to 0.30) and 0.19 (95% CI, –0.19 to 0.57)] shows that the average predicted risk of both models was below the average observed risk.

Model update

For our model update, data from the 486 patients in the validation (phase 1) cohort was combined with the data from the development (phase 0) cohorts, amounting to a total of 1816 patients. The prognostic model that resulted from our model update is shown in Table 1. The model formula is available as Appendix 2. Statistical interaction-by-phase tests revealed no statistically significant differences in predictor effects between phase 0 and phase 1 (Table 1). Cross-validation showed a comparable performance of the models in both datasets (Appendix 3). Both findings suggest that it was reasonable to combine the datasets for a model update.

Validation of the updated model

At internal-external validation (Fig. 4a), we obtained a pooled c-statistic of 0.60 (0.58, 0.63), with I^2 estimated

Table 3. Relapse characteristics ($n = 277$)

Factor	*N* (%) or median (IQR)
Time until relapse, years	0.80 (0.41–1.69)
Relapse within 1 year	160 (58)
Relapse within 2 years	228 (82)
Type of relapse	
Clinical	220 (79)
Biochemical	130 (47)
Endoscopic	118 (43)
Perianal disease	20 (7)
Extra-intestinal manifestations	7 (3)
Type of reintroduced therapy	
Antibiotics	6 (2)
Aminosalicylates	11 (4)
Thiopurines	62 (22)
Methotrexate	9 (3)
Steroids	96 (35)
Biological	174 (63)
Surgery	12 (4)
Need for hospitalization	42 (15)
Type of reintroduced biological after relapse ($n = 174$)	
Adalimumab	98 (56)
Infliximab	64 (37)
Vedolizumab	9 (5)
Ustekinumab	3 (2)
Effect reintroduction anti-TNF ($n = 129$)	
Response/remission	104 (81)
Stopped due to nonresponse/side-effects	25 (19)
Retreatment successful with the same anti-TNF	
Adalimumab ($n = 61$)	50 (82)
Infliximab ($n = 47$)	40 (85)

TNF, tumor necrosis factor.

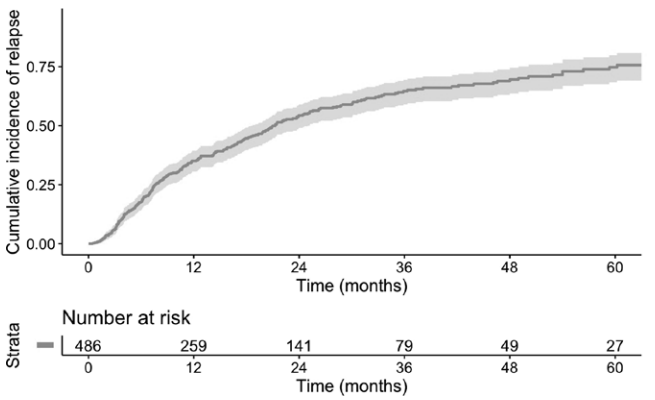

Fig. 2. Kaplan Meier survival curve.

at 0%, suggesting no between-cohort heterogeneity in the discriminative ability of the model. The calibration slope was estimated at 0.89 [(0.69–1.09), $I^2 = 0$], only slightly below 1 (Fig. 4b). Calibration-in-the large was estimated at 0.02 (–0.22 to 0.27) with an I^2 of 80%, reflecting substantial differences in baseline risk (Figure 4c). Decision curve analysis (Fig. 5) showed that the use of the updated model as a decision tool yields a higher net benefit than default strategies for threshold probabilities over 20%.

Discussion

A valid tool for patient identification to safely cease anti-TNF therapy is highly desirable. Recently, we published a prediction model with modest discriminative ability, based on an IPD-MA of 14 studies [16]. In this cohort, our previously developed phase 0 prediction model was externally validated and updated to estimate the risk of relapse in individual patients after cessation of anti-TNF therapy with reasonable discriminative ability.

The initial phase 0 model was not well-calibrated on the validation cohort: the predictions were too low for low-risk patients and too high for high-risk patients. Overall, the predictor effects were too strong for the phase 1 data. This could possibly be explained by, but may not be limited to, the difference in the distribution of baseline characteristics between the phase 0 cohorts and the phase 1 cohort, for example, concomitant immunotherapy, 63 vs. 36%, and endoscopic remission, 44 vs. 82%. Second, the collected data of the initial model enclosed a different period compared to the data of the validation cohort, leading to potential differences in treatment strategies. Third, in the phase 0 model with fecal calprotectin, 'clinical remission' was significantly associated with sustained remission after cessation of anti-TNF (HR, 0.30; 95% CI, 0.16–0.57). In the validation cohort, only 12 patients (2%) were not in clinical remission (this subgroup of patients demonstrated baseline remission as indicated by either biochemical or endoscopic remission before cessation of anti-TNF therapy). Furthermore, 'clinical remission' was based on the medical records instead of validated questionnaires which were mostly used in the cohorts of phase 0. Therefore, this variable could not be used to accurately predict the risk of relapse in phase 1.

Although the updated model still showed moderate discriminative performance, the performance was stable across cohorts. In other fields, including oncology and fertility research, prediction models with similar c-statistics

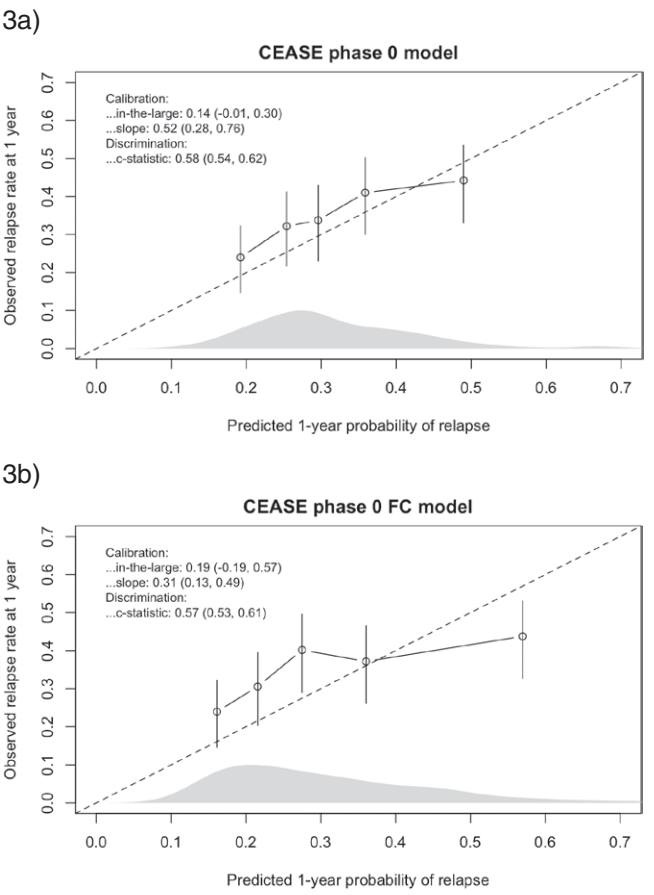

Fig. 3. (a) Calibration plot Phase 0 model. (b) Calibration plot Phase 0 model with fecal calprotectin.

varying between 0.58 and 0.64 are frequently used as a guide for making decisions [29–41]. Despite this seemingly moderate c-statistic, these models may still have added value for decision-making in daily practice. In addition, we demonstrated that our model may be useful as a prognostic tool for individualized decision-making in clinical practice across a wide range of thresholds (0.2–0.7). This threshold expresses how the benefit of treatment, that is, the prevention of relapse, is weighed against the harm of treatment, that is treating nonrelapsing patients unnecessarily. If a clinician saw no harm at all in treating patients unnecessarily, it would be most beneficial to keep all patients on treatment. If, however, there is harm in treating patients unnecessarily, such that the clinician is willing to treat at most five patients to prevent relapse in one of them, then the use of our updated model showed increased net benefit compared to keeping all patients on treatment. The model is most useful if the decision threshold is around 33%, implying a benefit-to-harm ratio of around two (67/33) and a willingness to treat around two patients to prevent one relapse. In addition to the willingness to cease anti-TNF therapy by the treating physician, the patient's decision is equally important. The prediction model might aid patients as well in the process of shared decision-making.

Our updated model may support a better cessation strategy compared to current international Inflammatory Bowel Disease guidelines which state that anti-TNF cessation is recommended only in patients in long-standing stable deep remission (clinical, biochemical and endoscopic) [42]. Another important indication for using the prediction model could be that it not only supports the decision of anti-TNF cessation in low-risk patients, but it will also avoid unjustifiable anti-TNF cessation in a subgroup of patients with a high risk of relapse. In addition, a stimulating thought for using this prediction model is the knowledge regarding the successful retreatment rates with anti-TNF therapy after relapse. Our cohort reported a high success rate of 81% which is in line with available literature [14,43,44].

This external cohort reported a 1-year relapse rate of 32%, which is in line with available literature (26–44%) [14,16,43–46]. The differences between cohorts could be partly explained by heterogeneity in the definitions of 'remission' and 'relapse'. Louis *et al*. [14] included patients who were in steroid-free remission [14] and in others, discontinuation was attempted in patients who were in clinical remission [43,45]. In the study by Bots *et al*. [46], as well as in our study, patients who were in either clinical, biochemical or endoscopic remission were included. Moreover, relapse was defined as clinical symptoms [14] or as disease activity leading to a therapeutic intervention [16,17,22,23]. This implicates the endpoint to be largely subjective, as it is only based on the judgement of the treating gastroenterologist (risking noninflamed patients to be designated as having a relapse) and not based on objective evidence such as biomarkers or endoscopy to confirm active inflammation.

Although the prediction models have been reasonably validated, some limitations need to be discussed. As we collected patient data retrospectively from electronic patient databases, the assessment of 'clinical remission' could have been interpreted differently by the treating physicians. In addition, due to its retrospective character, missing data on biochemical markers and endoscopic procedures were inevitable. Moreover, anti-TNF serum concentrations were particularly difficult to obtain, as this was not routinely measured in many patients. Due to this,

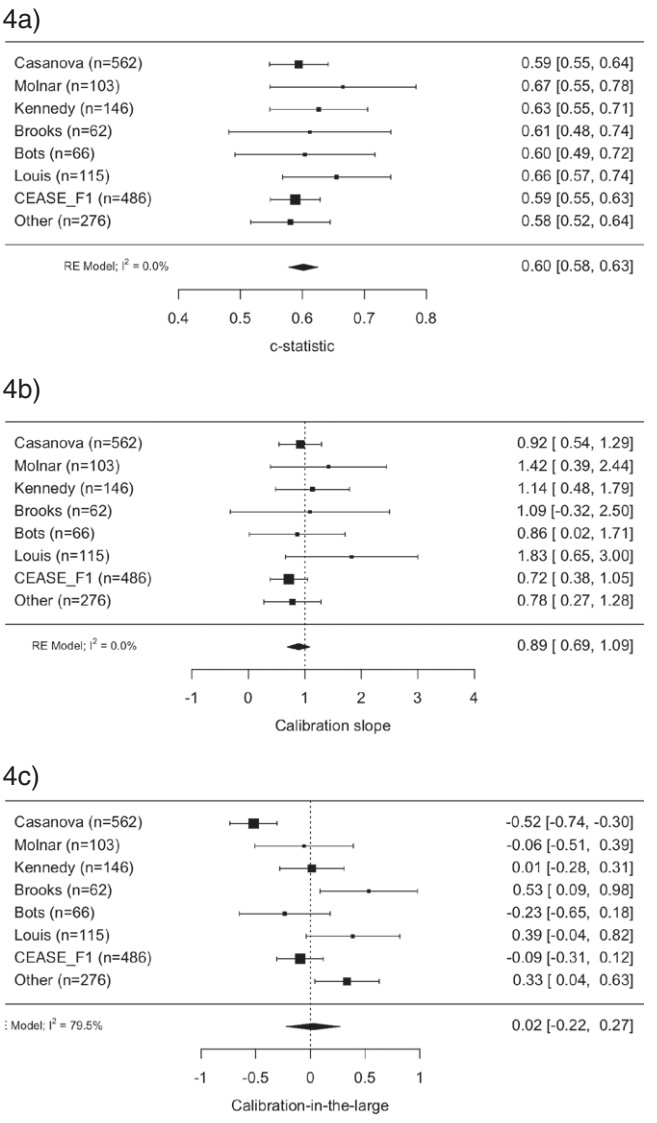

Fig. 4. (a) Internal-external validation CEASE phase 1 model (predictive performance). (b) Internal-external validation CEASE phase 1 model (calibration-in-the-large). (c) 4c Internal-external validation CEASE phase 1 model (calibration slope).

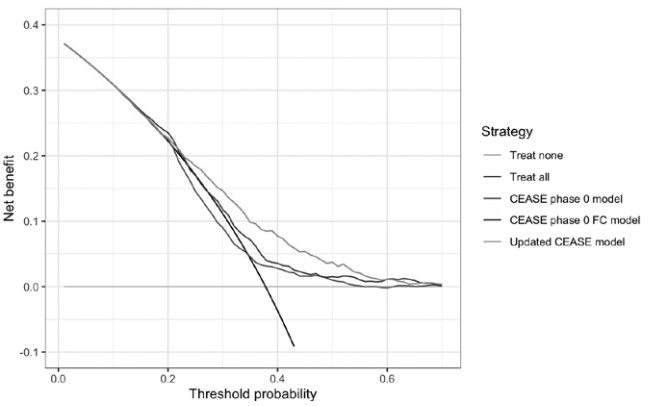

Fig. 5. Decision curve analysis.

the level of anti-TNF serum concentration could not be identified as a significant predictor for relapse, which has previously been reported in the literature [19,46]. In our prospective follow-up study, anti-TNF serum concentration will be measured at baseline.

Our study accentuates the difficulty of predicting the risk of relapse in CD patients who cease anti-TNF therapy in daily clinical practice as the underlying pathophysiology of relapse is poorly understood. Identification of new biomarkers for better discrimination between high- and low-risk patients is necessary. Further research is warranted to update the prediction model, including biochemical, serological, histologic and/or genetic markers. Previous studies reported on mucosal cytokines and serological markers which might be associated with the risk of relapse, as normalization of IL-17a expression and mucosal TNF predicts long-term remission after anti-TNF discontinuation [47]. In addition, a recently published study reported on protein biomarkers and metabolomics markers which were associated with relapse [48], while other studies discovered potential biomarker candidates associated with the risk of short- and long-term relapse after discontinuation of infliximab [48,49]. More quantified research into such predictors is warranted to further update and strengthen our prediction model.

We have used the updated prediction model to create a prognostic tool which will be publicly available as a user-friendly web interface on Evidencio. However, further evaluation of the prognostic performance of the model is necessary before it can be used in daily practice. To do this, we have initiated a multicentre (200 patients from 19 centers), center-specific stepped-wedge randomized controlled trial (RCT) (Netherlands Trial Register: NL8891). In addition, this RCT will provide prospective data for further updating the prediction model with biomarkers, histological and endoscopic data, as well as insight in the cost-effectiveness of the new strategy of stopping anti-TNF therapy based on the prediction model.

In conclusion, our previously developed prediction model to safely cease anti-TNF therapy has been validated and updated in this external Dutch multicentre CD cohort. After validation and update, the model showed reasonable discriminative performance and improved calibration. A further update of the model with biochemical and histological data is necessary to improve our ability to adequately select patients for cessation of anti-TNF therapy. We will further improve this prediction model through a large national RCT to assess whether this updated prediction model leads to a better selection of patients for anti-TNF cessation as compared to daily practice.

Acknowledgements

The authors would like to thank R.W.M. Pauwels, M.J. Casanova, J.P. Gisbert, N.A. Kennedy, C.W. Lees, E. Louis, T. Molnár, K. Szántó, E. Leo, S. Bots, R. Downey, M. Lukas, W.C. Lin, A. Amiot, C. Lu, X. Roblin, K. Farkas, J.B. Seidelin.

This work was supported by ZonMW GoedGebruik-Geneesmiddelen (project number: 848101009).

The formula of the model is available as supplementary information. The updated CEASE prediction model will be available as a user-friendly web-calculator with the publication of this article, to be found at the web-based platform Evidencio.

Poster presentation at the 14th congress of ECCO – Inflammatory Bowel Diseases 2020, Copenhagen, Denmark), "P388 Prediction model to safely CEASE anti-TNF therapy in Crohn's disease: validation of a predictive diagnostic tool for the cessation of anti-TNF treatment in CD in a Dutch population".

Oral presentation at the Dutch Digestive Days (Veldhoven) 2020 "53 Prediction model to safely CEASE anti-TNF therapy in Crohn's disease: validation of a predictive diagnostic tool for the cessation of anti-TNF treatment in CD in a Dutch population".

Study design: A.C.V., C.J.W., E.W,S., M.G.D., M.D., G.D. Data collection: S.T.B.H., D.J. Offering data: A.B., R.W., T.R., F.H., R.M., S.T., W.M., F.H., G.D., J.R., N.B., A.T., P.B. G.T., D.A. Statistical analysis and interpretation of data: D.T., D.J., S.T.B.H., E.W.S., D.N. Drafting of the manuscript: S.T.B.H., D.J., D.T. All coauthors revised and approved the manuscript for important intellectual content.

Conflicts of interest

R.W. has participated in advisory boards or as a speaker or consultant for AbbVie, and Janssen, Takeda and Galapagos. T.R. has participated in advisory board and/or received financial compensation from the following company: Takeda. F.H. has served on advisory boards or as speaker for Abbvie, Janssen-Cilag, MSD, Takeda, Celltrion, Teva, Sandoz and Dr Falk. Funding (Grants/Honoraria): Dr Falk, Janssen-Cilag, Abbvie, Takeda. Consulting Fees: Celgene. N.B. has served as a speaker for AbbVie and MSD and has served as consultant and principal investigator for TEVA Pharma BV and Takeda. Received a (unrestricted) research grant from Dr. Falk, TEVA Pharma BV, MLDS and Takeda, all outside the submitted work. D.A. advisor for galapagos, Takeda, Ferring. Received educational grants from Janssen, DrFalk and Ferring. G.R.D. has served as advisor for Abbvie, Ablynx, Alimentiv, Amgen, AM Pharma, Biogen, Bristol Meiers Squibb, Boerhinger Ingelheim, Celgene/Receptos, Celltrion, Cosmo, Covidien/Medtronics, Ferring, DrFALK Pharma, Eli Lilly, Engene, Galapagos, Genentech/Roche, Gilead, Glaxo Smith Kline, Immunic, Johnson and Johnson, Lamepro, Lument, Mitsubishi Pharma, Merck Sharp Dome, Mundipharma, Nextbiotics, Novonordisk, Otsuka, Pfizer, Polpharm, Prometheus laboratories/Nestle, Procise diagnostics, Protagonist, Salix, Samsung Bioepis, Sandoz, Setpoint, Shire, Takeda, Tigenix, Tillotts, Topivert, Versant and Vifor; received speaker fees from Abbvie, Biogen, Ferring, Johnson and Johnson, Merck Sharp Dome, Mundipharma, Norgine, Pfizer, Samsung Bioepis, Shire, Takeda, Tillotts and Vifor. C.W.received grant support from Falk Benelux and Pfizer; received speaker fees from AbbVie, Takeda, Ferring, Dr. Falk Pharma, Hospira, Pfizer; and served as a consultant for AbbVie, MSD, Takeda, Celgene, Mundipharma and Janssen. M.D.has served as advisor for Echo pharma and Robarts Clinical Trials, reports nonfinancial support from Dr. Falk Pharma, and received speaker fees from Janssen, Merck & Co., Pfizer, Takeda and Tillotts Pharma. A.V.Advisory board

Jansen, Takeda and Abbvie. For the remaining authors, there are no conflicts of interest.

References&&

1 Peyrin-Biroulet L, Deltenre P, de Suray N, Branche J, Sandborn WJ, Colombel JF. Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: meta-analysis of placebo-controlled trials. *Clin Gastroenterol Hepatol* 2008; 6:644–653.

2 Dassopoulos T, Sultan S, Falck-Ytter YT, Inadomi JM, Hanauer SB. American gastroenterological association institute technical review on the use of thiopurines, methotrexate, and anti-TNF-α biologic drugs for the induction and maintenance of remission in inflammatory Crohn's disease. *Gastroenterology* 2013; 145:1464–78.e1.

3 Stidham RW, Lee TC, Higgins PD, Deshpande AR, Sussman DA, Singal AG, *et al*. Systematic review with network meta-analysis: the efficacy of anti-TNF agents for the treatment of Crohn's disease. *Aliment Pharmacol Ther* 2014; 39:1349–1362.

4 D'Haens G, Reinisch W, Colombel JF, Panes J, Ghosh S, Prantera C, *et al*; ENCORE investigators. Five-year safety data from ENCORE, a European observational safety registry for adults with Crohn's disease treated with infliximab [Remicade®] or conventional therapy. *J Crohns Colitis* 2017; 11:680–689.

5 Lichtenstein GR, Feagan BG, Cohen RD, Salzberg BA, Diamond RH, Price S, *et al*. Serious infection and mortality in patients with Crohn's disease: more than 5 years of follow-up in the TREAT™ registry. *Am J Gastroenterol* 2012; 107:1409–1422.

6 Papamichael K, Mantzaris GJ, Peyrin-Biroulet L. A safety assessment of anti-tumor necrosis factor alpha therapy for treatment of Crohn's disease. *Expert Opin Drug Saf* 2016; 15:493–501.

7 Villoria A, García V, Dosal A, Moreno L, Montserrat A, Figuerola A, *et al*. Fatigue in out-patients with inflammatory bowel disease: prevalence and predictive factors. *PLoS One* 2017; 12:e0181435.

8 Mandel MD, Michael MD, Bálint A, Lovász BD, Gulácsi L, Strbák B, *et al*. Work disability and productivity loss in patients with inflammatory bowel diseases in Hungary in the era of biologics. *Eur J Health Econ* 2014; 15 (Suppl 1):S121–S128.

9 van Gennep S, Evers SW, Rietdijk ST, Gielen ME, de Boer NKH, Gecse KB, *et al*. High disease burden drives indirect costs in employed inflammatory bowel disease patients: the WORK-IBD study. *Inflamm Bowel Dis* 2021; 27:352–363.

10 Severs M, Oldenburg B, van Bodegraven AA, Siersema PD, Mangen MJ; initiative of Crohn's and Colitis. The economic impact of the introduction of biosimilars in inflammatory bowel disease. *J Crohns Colitis* 2017; 11:289–296.

11 Lawton J, Achit H, Pouillon L, Boschetti E, Demore B, Matton T, *et al*. Cost-of-illness of inflammatory bowel disease patients treated with anti-tumour necrosis factor: a French large single-centre experience. *United European Gastroenterol J* 2019; 7:908–913.

12 Waljee AK, Chaisidhivej N, Saini SD, Higgins PDR. De-escalation of IBD therapy: when, who, and how? *Crohns Colitis* 2019; 360:1.

13 van Maag-Darm-Leverartsen NV. Kennisagenda NVMDL 2016. https://www.demedischspecialist.nl/sites/default/files/kennisagenda_NVMDL.pdf. [Accessed 24 July 2020]

14 Louis E, Mary JY, Vernier-Massouille G, Grimaud JC, Bouhnik Y, Laharie D, *et al*. Maintenance of remission among patients with Crohn's disease on antimetabolite therapy after infliximab therapy is stopped. *Gastroenterology* 2012; 142:63–70 e5; quiz e31.

15 Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019; 38:1276–1296.

16 Pauwels RWM, van der Woude CJ, Nieboer D, Steyerberg EW, Casanova MJ, Gisbert JP, *et al*. P138 prediction model to safely cease anti-TNF therapy in Crohn's disease: individual patient data meta-analysis (IPD-MA). *J Crohns Colitis* 2019; 13 (Suppl 1):S158–S159.

17 Moher D, Altman DG, Liberati A, Tetzlaff J. PRISMA statement. *Epidemiology* 2011; 22:128; author reply 128.

18 Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; 368:m441.

19 Ben-Horin S, Chowers Y, Ungar B, Kopylov U, Loebstein R, Weiss B, *et al*. Undetectable anti-TNF drug levels in patients with long-term remission predict successful drug withdrawal. *Aliment Pharmacol Ther* 2015; 42:356–364.

20 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Eur J Clin Invest* 2015; 45:204–214.

21 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; 2020. http://www.r-project.org/index.html.

22 van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011; 45:1–67.

23 Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015; 68:279–289.

24 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361–387.

25 Debray TP, Moons KG, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med* 2013; 32:3158–3180.

26 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327:557–560.

27 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26:565–574.

28 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019; 3:18.

29 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29–36.

30 Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail *et al*. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001; 93:358–366.

31 Berry DA, Parmigiani G, Sanchez J, Schildkraut J, Winer E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J Natl Cancer Inst* 1997; 89:227–238.

32 Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989; 81:1879–1886.

33 Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, *et al*. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst* 2007; 99:1782–1792.

34 Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, Gail MH. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J Natl Cancer Inst* 2011; 103:951–961.

35 Banegas MP, John EM, Slattery ML, Gomez SL, Yu M, LaCroix AZ, *et al*. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst* 2017; 109.

36 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004; 23:1111–1130.

37 Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet* 1991; 48:232–242.

38 Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer* 1994; 73:643–651.

39 Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast Cancer Res Treat* 1993; 28:115–120.

40 McCarthy AM, Guan Z, Welch M, Griffin ME, Sippo DA, Deng Z, *et al*. Performance of breast cancer risk assessment models in a large mammography cohort. *J Natl Cancer Inst* 2019.

41 Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, *et al*. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009; 15:537–552.

42 Doherty G, Katsanos KH, Burisch J, Allez M, Papamichael K, Stallmach A, *et al*. European Crohn's and colitis organisation topical review on treatment withdrawal ['Exit Strategies'] in inflammatory bowel disease. *J Crohns Colitis* 2018; 12:17–31.

43 Brooks AJ, Sebastian S, Cross SS, Robinson K, Warren L, Wright A, *et al*. Outcome of elective withdrawal of anti-tumour necrosis factor-α therapy in patients with Crohn's disease in established remission. *J Crohns Colitis* 2017; 11:1456–1462.

44 Kennedy NA, Warner B, Johnston EL, Flanders L, Hendy P, Ding NS, *et al*; UK Anti-TNF withdrawal study group. Relapse after withdrawal from anti-TNF therapy for inflammatory bowel disease: an observational study, plus systematic review and meta-analysis. *Aliment Pharmacol Ther* 2016; 43:910–923.
45 Casanova MJ, Chaparro M, García-Sánchez V, Nantes O, Leo E, Rojas-Feria M, *et al*. Evolution after anti-TNF discontinuation in patients with inflammatory bowel disease: a multicenter long-term follow-up study. *Am J Gastroenterol* 2017; 112:120–131.
46 Bots SJ, Kuin S, Ponsioen CY, Gecse KB, Duijvestein M, D'Haens GR, Löwenberg M. Relapse rates and predictors for relapse in a real-life cohort of IBD patients after discontinuation of anti-TNF therapy. *Scand J Gastroenterol* 2019; 54:281–288.
47 Rismo R, Olsen T, Cui G, Paulssen EJ, Christiansen I, Johnsen K, *et al*. Normalization of mucosal cytokine gene expression levels predicts long-term remission after discontinuation of anti-TNF therapy in Crohn's disease. *Scand J Gastroenterol* 2013; 48:311–319.
48 Borren NZ, Plichta D, Joshi AD, Bonilla G, Sadreyev R, Vlamakis H, *et al*. Multi-"-Omics" profiling in patients with quiescent inflammatory bowel disease identifies biomarkers predicting relapse. *Inflamm Bowel Dis* 2020; 26:1524–1532.
49 Pierre N, Baiwir D, Huynh-Thu VA, Mazzucchelli G, Smargiasso N, De Pauw E, *et al*. Discovery of biomarker candidates associated with the risk of short-term and mid/long-term relapse after infliximab withdrawal in Crohn's patients: a proteomics-based study. *Gut* 2021; 70:1450–1457.