



A Computational Platform Integrating a Mechanistic Model of Crohn's Disease for Predicting Temporal Progression of Mucosal Damage and Healing

Sai Phanindra Venkatapurapu · Ryuichi Iwakiri · Eri Udagawa ·
Nikhil Patidar · Zhen Qi · Ryoko Takayama · Kei Kumar ·
Yuki Sato · Marcelo Behar · Patrick Offner · Gaurav Dwivedi ·
Haruna Miyasaka · Ryohsuke Ken Suzuki · Anna Lissa Hamada ·
Paul M. D'Alessandro · Jovelle Fernandez

Received: December 14, 2021 / Accepted: March 24, 2022 / Published online: May 17, 2022
© The Author(s) 2022

ABSTRACT

Introduction: Physicians are often required to make treatment decisions for patients with Crohn's disease on the basis of limited objective information about the state of the patient's gastrointestinal tissue while aiming to achieve mucosal healing. Tools to predict changes in mucosal health with treatment are needed. We evaluated a computational approach integrating a mechanistic model of Crohn's disease with a responder classifier to predict temporal changes in mucosal health.

Methods: A hybrid mechanistic–statistical platform was developed to predict biomarker

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12325-022-02144-y>.

S. P. Venkatapurapu (✉) · N. Patidar · Z. Qi ·
M. Behar · P. Offner · G. Dwivedi ·
P. M. D'Alessandro
PricewaterhouseCoopers LLP, 301 Grant Street,
Suite 4500, Pittsburgh, PA 15219, USA
e-mail: sai.phanindra.venkatapurapu@pwc.com

R. Iwakiri · E. Udagawa · H. Miyasaka ·
R. K. Suzuki · A. L. Hamada · J. Fernandez
Japan Medical Office, Takeda Pharmaceutical
Company Limited, Tokyo, Japan

R. Takayama · K. Kumar · Y. Sato
PwC Consulting LLC, Otemachi One Tower,
1-2-1 Otemachi, Chiyoda-ku, Tokyo 100-0004,
Japan

and tissue health time courses in patients with Crohn's disease. Eligible patients from the VERSIFY study ($n = 69$) were classified into archetypical response cohorts using a decision tree based on early treatment data and baseline characteristics. A virtual patient matching algorithm assigned a digital twin to each patient from their corresponding response cohort. The digital twin was used to forecast response to treatment using the mechanistic model.

Results: The responder classifier predicted endoscopic remission and mucosal healing for treatment with vedolizumab over 26 weeks, with overall sensitivities of 80% and 75% and overall specificities of 69% and 70%, respectively. Predictions for changes in tissue damage over time in the validation set ($n = 31$), a measure of the overall performance of the platform, were considered good (at least 70% of data points matched), fair (at least 50%), and poor (less than 50%) for 71%, 23%, and 6% of patients, respectively.

Conclusion: Hybrid computational tools including mechanistic components represent a promising form of decision support that can predict outcomes and patient progress in Crohn's disease.

Keywords: Computational model; Crohn's disease; Mucosal healing

Key Summary Points

Why carry out this study?

The poor correlation in Crohn's disease (CD) between the presence of symptoms and mucosal health makes it difficult for patients to appreciate the need for continued treatment.

Patient's perception that no treatment may be required is a key driver of nonadherence.

What did the study ask?

Can a novel computational platform predict the evolution of mucosal health in patients with CD to provide treatment milestones without the need of an endoscopic evaluation and drive improved physician–patient conversations?

What were the study outcomes/conclusions?

The platform predicted endoscopic remission and mucosal healing after treatment with vedolizumab for 26 weeks with an overall sensitivity of 80% and 75% and overall specificity of 69% and 70%, respectively.

What was learned from the study?

The computational platform developed and validated in this study could be used to predict the evolution of tissue damage in response to treatment with vedolizumab.

With further clinical validation, the platform could be used to predict the progress of mucosal healing in patients with CD undergoing treatment and support shared decision-making.

DIGITAL FEATURES

This article is published with a video feature to facilitate understanding of the article. To view the video feature for this article go to <https://doi.org/10.6084/m9.figshare.19411058>.

INTRODUCTION

Crohn's disease (CD) is a chronic, relapsing inflammatory disease of the gastrointestinal (GI) tract. Current evidence points to the disease being caused by a combination of dysregulated immune response, altered microbiota (dysbiosis), genetic susceptibility, and environmental factors that contribute to the breakdown of barrier function [1–5]. Disease onset often occurs at a young age [6]. Untreated, a feedback loop between tissue damage and inflammation [7] drives progression of the disease, which can result in complications, such as fistulae (perforations in the intestinal walls) and strictures (narrowing of the lumen). Treatment for CD has historically focused on controlling disease activity to alleviate or eliminate symptoms, often defined in subjective scales. More recently, emphasis has shifted toward treatments that promote tissue healing and restore mucosal function [8–10]. Response is evaluated in terms of objective metrics, such as ulcerated area or ulceration patterns in the GI tract, which are typically included in scores such as the Crohn's Disease Endoscopic Index of Severity [11] and Simple Endoscopic Score for Crohn's Disease (SES-CD) [12]. Importantly, the progression of tissue damage often has a different trajectory to the occurrence of symptoms. Loose coupling between symptoms and the condition of the mucosa, combined with the limited frequency at which endoscopic evaluations can be conducted, creates a significant challenge for

physicians in communicating to patients with CD the need for continued treatment. Lack of perceived need for treatment by patients is one of the primary causes for nonadherence in CD, and better communication between physicians and patients could improve patient adherence [13]. Therefore, there is a critical need for tools to help physicians forecast the temporal changes in mucosal health in patients with CD, to bridge the gap between subjective and less frequent objective assessments of disease severity. Such tools would support the decision to start or continue treatment and provide critical milestone information to evaluate the progress of mucosal healing in patients undergoing treatment.

Computational tools for predicting specific outcomes (e.g., surgery) in CD exist [14, 15]. However, they typically do not provide time-resolved information on how patients will evolve toward those outcomes or allow experimentation with what-if treatment scenarios. Recently, researchers have tried to understand the dynamics of immune response and resultant patient phenotypes by using computational models of CD [16, 17]. Rogers and colleagues have developed a mechanistic model of Crohn's disease to simulate the effect of treatments on biomarkers; however, the existing models do not predict objective measures of disease activity over time [18, 19]. Here, we address this gap through a novel computational platform that integrates statistical and mechanistic approaches to predict the evolution of mucosal health in patients with CD.

Hybrid approaches, which combine elements of machine learning with domain-specific physical models, are becoming increasingly popular in many fields. In biomedicine, composite models integrating machine learning and mechanistic components reflecting known biology have been proposed [20, 21]. Because they capture the causal relationships between biology and disease, mechanistic models function as guardrails to constrain the interpretation of data during model training while at the same time providing a natural way of incorporating continuous time. This latter ability is critical in generating a time-resolved picture of how patients may respond to different treatment

scenarios [22, 23]. Anchoring machine-learning models in physiology improves the interpretability of predictions and mitigates the risks associated with the need to relearn well-established physiological relationships from potentially noisy data [24]. The platform demonstrated in this study aims to provide decision-support tools to help physicians understand how patients are reacting to a given therapeutic approach, and eventually to help them manage their treatment.

METHODS

Overview of the Computational Platform

We developed a computational platform for simulating mucosal health and the level of inflammation in patients with CD under different treatment scenarios. The platform estimates changes in mucosal health and inflammatory activity over a continuous timeline of weeks to years. The platform has the following three major components:

1. A machine learning-based responder classifier.
The purpose of this classifier is to use patient background and early response data to predict longer-term responses in a qualitative sense, e.g., complete, partial, or no response, using a simple decision tree. The responder classifier decision tree was built by analyzing treatment response data in patients from the VERSIFY study [25]. The responder classifier sets expectations for the likely long-term response and helps improve the precision of the quantitative predictions generated in subsequent steps.
2. A mechanistic model of CD pathophysiology.
The CD model allows quantitative simulation of CD progression in continuous time. The model is a mathematical representation of the salient mechanisms related to the gut immune system and CD pathophysiology, derived from the literature, in the form of a system of coupled differential equations. Model parameters may be calibrated to

represent an individual patient or a patient archetype. Once calibrated, the model may be used to forecast CD progression with various treatment scenarios.

3. A virtual library of patients with CD. A diverse library of virtual patients enables rapid mapping of a real patient to their digital twin. A library of virtual patients with CD was constructed by randomly sampling selected parameters of the mechanistic CD model, running the simulation with the sampled parameters, and recording the simulated CD progression in a database. This virtual patient library was used as a pool from which key attributes could be matched quickly to those of a real patient to identify an appropriate digital twin. For instance, if a patient is predicted by the responder classifier to achieve endoscopic remission, virtual patients showing endoscopic remission can be selected and further filtered using additional criteria.

Further details of the three components and their integration are described in subsequent sections. The integrated platform was trained and validated using patient data from the VERSIFY study. The platform was evaluated for its ability to predict the dynamics of CD progression as measured by the affected surface area of the gut for both individual patients and patient populations.

The following steps outline the methodology for predicting the response of a patient to a particular treatment using the computational platform:

1. *Determine the responder type of the patient using the responder classifier.*
Patient data (background information and early treatment response; Fig. 1, part 1) are first fed into a classification algorithm. The classifier's outputs are qualitative predictions about the likely long-term responses to treatment in terms of a predetermined set of categories.
2. *Create a digital twin of the patient using the virtual patient library.*
The responder classifier output is subse-

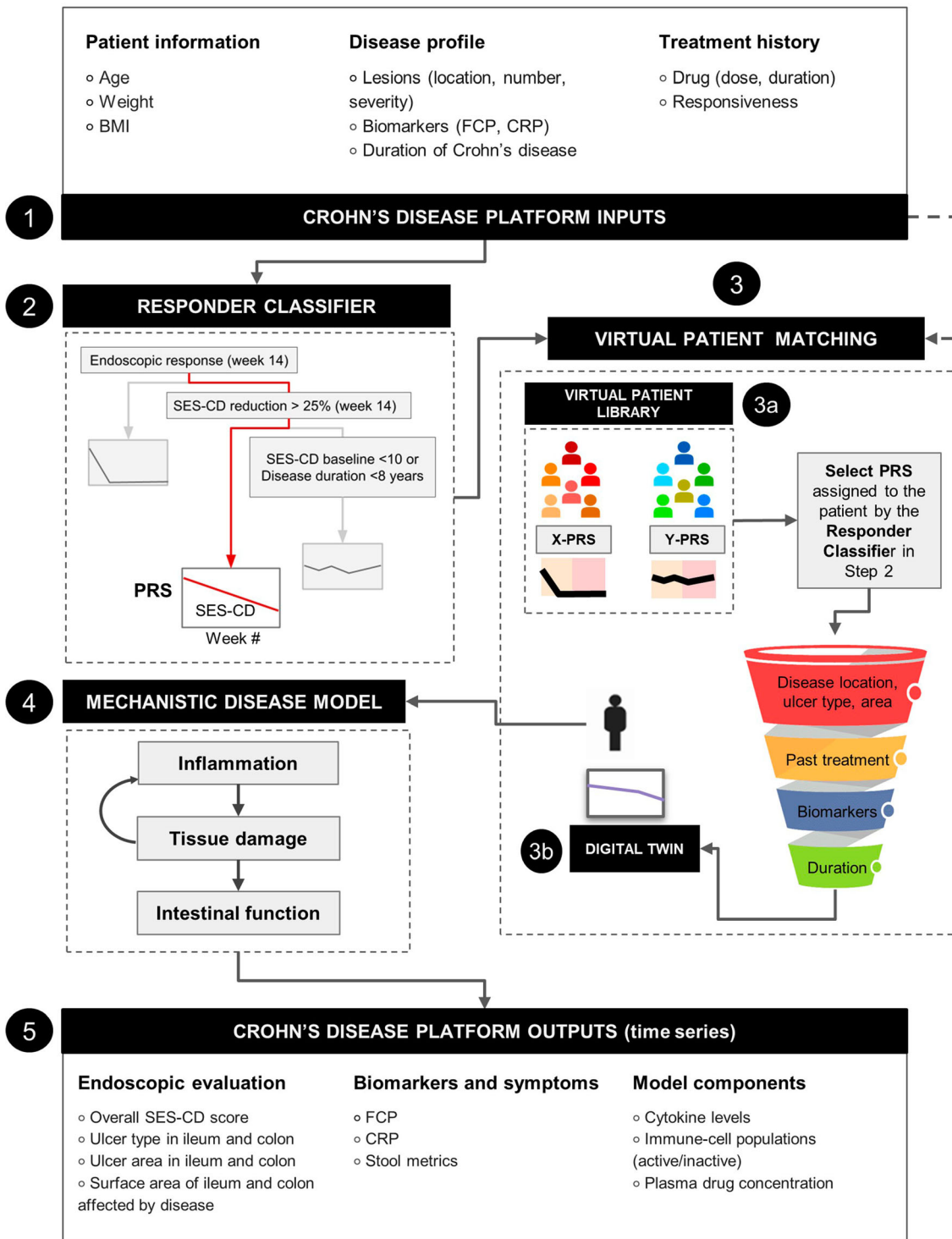
Fig. 1 Overview of the hybrid computational Crohn's disease simulation hybrid platform. Patient data are used as an input (part 1) into the classification model (responder classifier; part 2), which returns the best-matched PRS. The PRS is a semiquantitative, coarse-grained estimate of changes in mucosal health at selected times. The PRS (part 3), alongside other patient data, is used to select a virtual patient from the virtual patient library (part 3a). This virtual patient, which is an instance of the mechanistic model consistent with the patient data, represents a "digital twin" of the real patient (part 3b). The digital twin is then simulated forward in time (part 4) to produce time-series for tissue damage variables, biomarkers, and other model components (part 5). The digital twin can be used to test alternative drug regimens or treatments. *CRP* C-reactive protein, *FCP* fecal calprotectin, *PRS* progression-response scenario, *SES-CD* Simple Endoscopic Score for Crohn's Disease

quently combined with the patient's characteristics to find a virtual patient from the library that best matches the real patient's disease state, biomarkers, and clinical history. The output of this step is a "digital twin", a specific instance of the mechanistic model with a unique combination of parameters, customized to the patient (Fig. 1, part 3).

3. *Forecast the response of the digital twin using the mechanistic model of CD.*
Finally, the selected digital twin (Fig. 1, part 4) is used to simulate treatment and predict the patient's response to the treatment in terms of quantitative changes in biomarkers and clinical scores over continuous time (Fig. 1, part 5).

Data Sources

The platform was calibrated and tested against deidentified patient data from the VERSIFY study, in which patients with CD were treated with vedolizumab [25]. Data on baseline disease characteristics, treatment history, biomarkers (such as C-reactive protein [CRP] and fecal calprotectin [FCP]), and SES-CD evaluations were



used as reported in the VERSIFY study [25], with no further processing or modifications. Patients with fewer than three SES-CD evaluations were excluded from the dataset for the patient stratification analysis.

The study was conducted according to the Declaration of Helsinki and the International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices.

It was carried out retrospectively using a database of anonymized data, following Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese Ministry of Health, Labour and Welfare. The investigators could only access anonymized information from the database; therefore, in accordance with the aforementioned Japanese Ethical Guidelines, institutional ethics approval and informed consent were not required.

Progression-Response Classification

The objective of the classification model is to assign patients to one of several possible progression-response scenarios (PRSs) (Fig. 1, part 2), a construct we introduced in this study. A PRS is defined by specific patterns of changes in mucosal health and biomarkers over time. For this study, we used a set of PRSs based on the temporal dynamics of tissue damage following initiation of vedolizumab treatment. The evolution of tissue damage was tracked through the SES-CD score over a period of 1 year. To define the PRS, the 1-year period was divided into two stages: treatment initiation to 26 weeks, and 26–52 weeks. These temporal stages were selected because they correspond to timescales over which changes in mucosal health typically occur [26, 27] and match the timescales of the relevant physiological processes in the mechanistic model. The set of possible patterns of change in SES-CD from one stage to the next defines a set of PRSs. Assigning a PRS, in general terms, sets expectations for future evolution and defines an initial set of constraints on the mechanistic components of the platform. For example, if the PRS of a

patient indicates a quick resolution of tissue damage with treatment, the parameters of the mechanistic model that would lead to slow resolution can be ruled out. The platform is flexible and can use different PRS sets assuming the sets' components can be mapped to the mechanistic model and are informative for constraining that model to a patient.

Development of the Responder Classifier

To assign a patient to a PRS, we evaluated decision-tree approaches using different combinations of factors available in the data (e.g., baseline disease severity, disease duration), and considered informative a priori. To minimize the risk of overfitting and to produce algorithms potentially suitable for use in clinical settings, a design decision was made to limit the tree depth to no more than three levels and only incorporate factors with clear mechanistic links to tissue damage. To develop and test candidate classifiers, each patient in the VERSIFY study cohort was assigned manually to one PRS according to their SES-CD time course. Of the 101 patients enrolled in the VERSIFY study, patients with fewer than three endoscopy measurements ($n = 18$) were not considered for this part of the analysis, nor were patients not conforming to the general behavior of any PRS ($n = 14$). PRS cohorts were segregated randomly into training and validation sets at an approximately 2:1 ratio (Fig. 2a, b). During randomization, it was ensured that all the PRSs were represented in both training and validation sets. It must be noted that patients in the training set were also used to refine the mechanistic model. Candidate classification trees were initially assessed graphically by examining how patients in different PRSs were distributed in the multi-dimensional space defined by the factors in the tree. Cutoffs were then determined heuristically for the most promising factor combinations. Classifiers were evaluated on the ability to predict endoscopic remission and mucosal healing at 26 weeks.

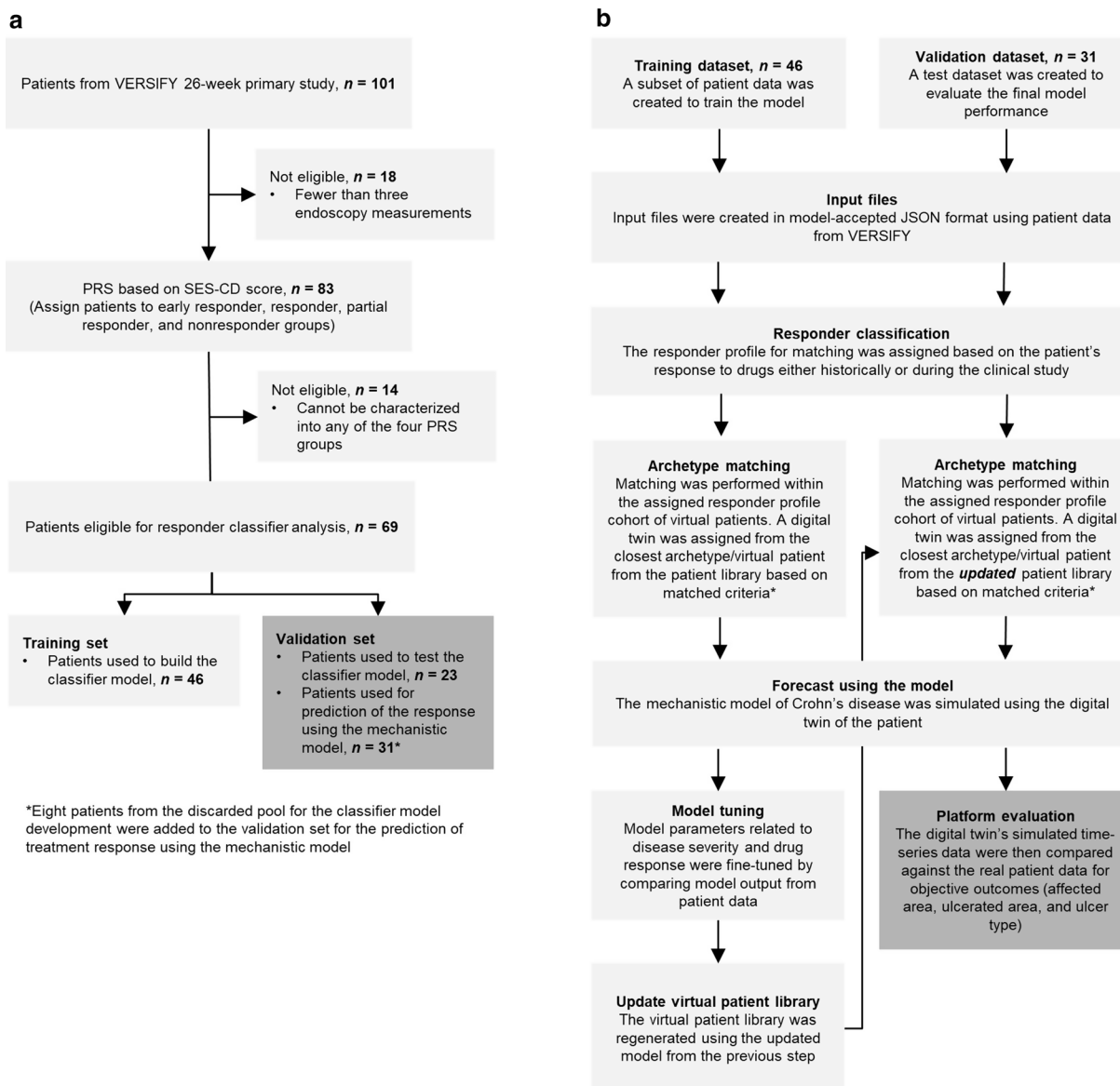


Fig. 2 Platform training and validation using VERSIFY data. **a** Patients from the VERSIFY study population were classified according to their SES-CD time course. Patients with fewer than three endoscopy measurements ($n = 18$) were not considered for this part of the analysis, nor were patients not conforming to the general behavior in any PRS ($n = 14$). PRS cohorts were segregated randomly into training and validation sets at an approximately 2:1 ratio. Patients who could not be assigned a PRS were not used to train or test the classification model but were used during evaluation of the mechanistic model. **b** The training dataset was used to tune the response to biologics in the

model and update the virtual patient library. The validation set was used to evaluate the computational platform. Each patient, in both the training and validation sets, was first assigned a responder type using the responder classifier. Subsequently, a digital twin was assigned to the patient from the virtual patient library using the archetype-matching algorithm. *Matching criteria included parameters such as demographics (age, weight, gender), biomarkers (C-reactive protein), disease parameters (ulcer type, ulcer area, disease area), and disease location. PRS progression-response scenario, SES-CD Simple Endoscopic Score for Crohn's Disease

Mechanistic Model Development and Description

A mechanistic model is a mathematical representation of biological processes, such as the inflammation–damage feedback loop in CD. Several mechanistic models of inflammatory bowel diseases exist [16, 17, 28–30]. However, to the best of our knowledge, none specifically addresses the relationship between inflammation and tissue damage in CD, a critical requirement for our goal. Therefore, we developed a new model for this study. The scope of the model was defined to incorporate relevant mechanisms for regulating changes in tissue damage and mucosal healing over weeks to years, as well as to allow representation of the mechanisms of action of drugs commonly used to treat CD. A literature review was carried out, and experts in the field were consulted to assess relevant disease pathophysiology. Physiological processes were represented as a system of coupled ordinary differential equations and algebraic expressions. Model parameters were derived from the literature or determined

through an optimization approach, as described in the Supplementary Material.

The model is organized into five main gut compartments, each corresponding to an intestinal segment, namely ileum, ascending colon, transverse colon, descending and sigmoid colon (combined into a single compartment), and rectum (Fig. S1a). Each compartment is divided into subcompartments (Fig. 3), representing the lamina propria and other subepithelial tissues (herein referred to as “lamina”), lymph nodes (“lymph”), and the epithelium (“tissue”). All the major gut compartments are connected through the circulation (“blood”) compartment, which enables movement of cell populations and signaling molecules (e.g., chemokines and cytokines) between compartments (Fig. S1a). Each main compartment has associated epithelial and mucosal layers. The epithelial layer can be in one of four possible states: healthy (EP_h), active (EP_a), damaged (EP_d), or remodeled (EP_r) (Fig. S1b), with transition rates controlled by the extent of local pro- and anti-inflammatory activity. The mucus layer has protective

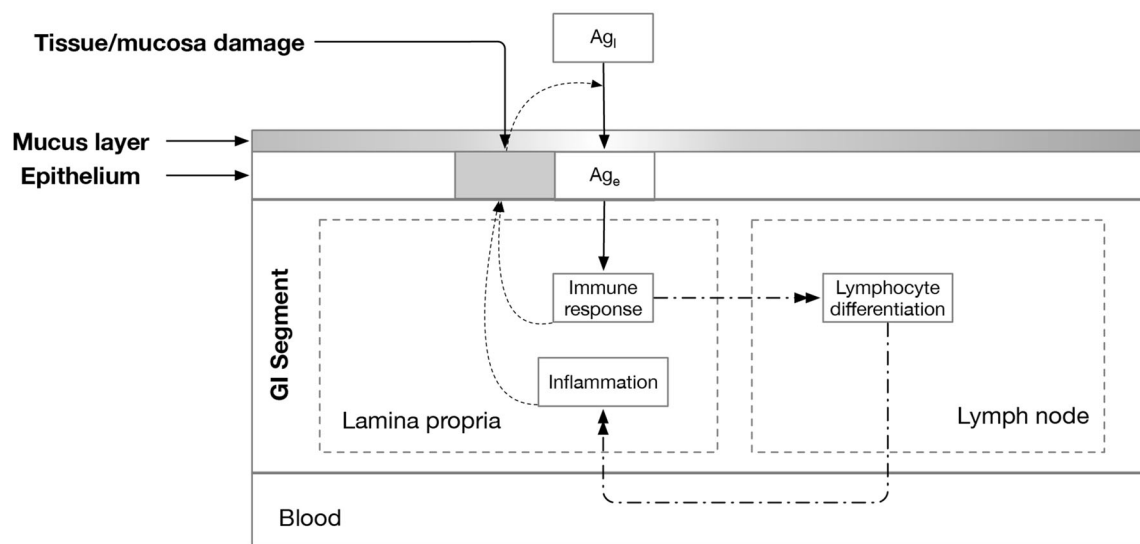


Fig. 3 High-level diagram of the model of progression of inflammation and tissue damage in Crohn’s disease. Each gut compartment in the model comprises three subcompartments, namely lamina propria, lymph node, and tissue (representing the epithelial layer). All the gut

compartments are connected by a single circulation compartment, which allows for cell migration and flow of cytokines between gut compartments. Ag_e antigen proximal to the epithelial layer, Ag_l luminal antigen, *GI* gastrointestinal

properties that prevent luminal antigens from reaching the epithelial layer. In the event of an injury to the mucus layer, possibly as a result of an infection, epithelial cells are exposed to luminal antigens and become activated [31]. The state of the mucosal layer is determined by the balance between mucus generation by specialized cells in the epithelial layer and clearance through natural processes. Disease-induced damage to mucus-generating cells results in deterioration of the mucosal layer. The layer is restored as new mucus-producing cells differentiate from stem cells upon healing. Active epithelial cells secrete chemotactic signals, which promote recruitment of immune cells to the lamina (Figs. S2 and S3) [31–33], causing further damage. Damaged tissue allows luminal antigens to reach the tissue owing to impaired barrier function, which further drives disease activity [34, 35]. Chronic inflammation and tissue damage lead to remodeling of the epithelial layer [36]. Both active and damaged tissue can revert to a healthy state through a process that is promoted by IL-22 [37, 38]. The remodeled state is irreversible, representing tissue affected by fibrosis. The fractions of active, damaged, and remodeled tissue are used to estimate SES-CD subscores for each gut compartment (Fig. S4a). Inflammatory activity is used to estimate levels of serum CRP and FCP (Fig. S4b). Intra-individual measurement variability for biomarkers was used to determine uncertainty around patient data, when comparing model predictions at an individual level (Fig. S4b) [39, 40]. Currently, treatment with steroids, immunomodulators, anti-TNF agents, and anti-integrin agents is supported (see Supplementary Material for details on pharmacological interventions), with the last of these being the focus of the present study. The model equations and parameters as well as model development are presented in detail in the Supplementary Material and Tables S1.1 to S3.3.

Virtual Patient Library

The mechanistic model refines predictions about disease progression and response to treatment beyond that implied by the PRS and

enables extrapolation to alternative treatment regimens. To facilitate calibration, reduce the need for expensive computations, and generate insights in a timely manner in a clinical setting, the platform includes a library of “virtual patients.” Virtual patients correspond to instances of the mechanistic model sharing most parameter values but differing in a selected few (Table S3.3). The parameters that changed between patients were determined through a sensitivity analysis to be the most influential for the dynamics of tissue damage and response to treatments included in the model (Tables S3.4 and S3.5). Each virtual patient was presimulated under various treatment scenarios (including no treatment) and time-series for tissue damage; inflammatory activity and biomarkers were recorded in a database. Virtual patients were preclassified according to common PRS groups (e.g., delayed responders to anti-integrin, early responders to anti-integrin, etc.; Fig. S5a). A core virtual patient library was created during model development based on literature estimates for parameter ranges, clinical cases reported in the literature, and experience from specialists treating patients with CD (see Supplementary Material for references).

To create a digital twin (an instance of the model that mimics a real patient), data from the real patient can be matched against the virtual patient library, specifically within the PRS assigned to the individual, to identify the closest match. The quality of this matching process depends on the diversity of phenotypes in the virtual patient library. The training set created from the VERSIFY study population was used to expand the library to capture the variability of responses within each PRS, as well as to verify that all the treatment regimens in the study were represented (Fig. 2b). To this end, patients in the training set were matched against existing virtual patients according to affected surface area, ulcer area, ulcer type, and biomarkers (CRP and FCP), as allowed by the available data. Patients deemed not to have a good match in the virtual patient library were used as targets to create new instances of the mechanistic model reflecting their data and augment the library (Fig. S5b). Specifically, the ranges for patient-specific parameters in Table S3.3 were calibrated

to capture the diverse phenotypes observed in the VERSIFY training set.

Digital Twin Assignment

A digital twin is an instance of the mechanistic model calibrated to match an individual patient's data. Calibration is performed in two steps. First, coarse tuning is done by selecting virtual patients from the virtual patient library consistent with the PRS assigned to the individual during the classification stage (Fig. 1, part 3a). This step effectively constrains model parameters to combinations capable of producing model outputs that match the trends for the dynamics of variables in the PRS (e.g., rapid decrease in SES-CD followed by stable levels) under the reported treatment regimen. Matching also considers the patient's body weight because of its effect on pharmacokinetics. Second, virtual patients who qualitatively match the individual time-series of tissue damage are filtered from the PRS-matched cohort, and a digital twin (Fig. 1, part 3b) is selected that recapitulates the actual patient data but with the highest fidelity. Finally, parameters controlling the biomarkers CRP and FCP are adjusted to match the range observed or expected in the individual.

Integration of the Modules into a Platform

The computational platform described in Fig. 1 was developed by integrating three primary components: (1) the responder classifier (component 1 in Fig. 1, part 2) described in the sections “[Progression-Response Classification](#)” and “[Development of the Responder Classifier](#)”; (2) digital twin creation by virtual patient matching (Fig. 1, part 3) described in the sections “[Virtual Patient Library](#)” and “[Digital Twin Assignment](#)”; and (3) the mechanistic model of CD (Fig. 1, part 4) described in the section “[Mechanistic Model Development and Description](#).” As described in the section “[Overview of the Computational Platform](#),” patient data provided as an input to the platform (Fig. 1, part 1) are first used by the responder classifier to predict likely qualitative

Table 1 Objective endoscopic assessment

Size of ulcers	0: None
	1: Aphthous ulcers (diameter 0.1–0.5 cm)
	2: Large ulcers (diameter 0.5–2 cm)
Ulcerated surface	3: Very large ulcers (diameter > 2 cm)
	0: None
	1: < 10%
Affected surface	2: 10–30%
	0: None
	1: < 50%
	2: 50–75%
	3: > 75%

response to a treatment based on patient's background information and early treatment response data. Subsequently, a digital twin of the patient is created by matching patient data to a subset of the virtual patients with the same responder profile as the patient. Finally, the digital twin, which is in fact a specific instance of the mechanistic model of Crohn's pathophysiology, is simulated to predict the response of the patient to treatment over time (Fig. 1, part 5).

Outcomes and Definitions

Ulcer area, affected surface area, and ulcer type are defined as in the SES-CD [12] score (Table 1). Endoscopic remission is defined as an SES-CD score of 4 or less. Endoscopic response is defined as at least a 50% reduction in SES-CD score.

Platform Evaluation Approach

The integrated platform was evaluated using the validation set ($n = 31$), created from the VERSIFY patient population (Fig. 2b). Demographic and treatment regimen information of the

patients in the validation set was unblinded from the beginning to ensure that we generated a virtual patient library that captured the diversity of the entire study population. Patient time-course biomarker and treatment response data were blinded except for factors used explicitly in the classification tree. Patients were matched to a PRS using the classification tree approach described above. A digital twin was selected from among the virtual patients consistent with the PRS based on the best match for the area affected by the disease at baseline. The digital twin’s presimulated time-series for tissue damage was then compared against the full unblinded data. Specifically, the predicted response for each patient was evaluated in terms of affected surface area. Model performance was categorized as “good,” “fair,” or “poor” according to the following criteria: (1) if the prediction for a metric matched at least 70% of the data points, the patient was labeled as “good” under that metric; (2) patients were labeled as “fair” if the prediction for the metric matched at least 50% of the data points; and (3) patients in the “poor” category had predictions that matched less than 50% of the data points. For patients with disease reported in more than one GI segment, this evaluation was done only on the section that most closely resembled the global SES-CD to reflect the predominant role of this score in the PRS. The integrated platform was also evaluated at a population level by comparing the percentage of patients in different categories of disease severity, as measured by the affected surface area and ulcerated area. This comparison was done at both 26 weeks and 52 weeks after treatment initiation. The initial matching (i.e., virtual patient assignment) was evaluated by comparing the distribution of patients in different disease-severity categories at week 0. Observed and predicted percentages of patients showing a reduction in biomarkers, namely CRP and FCP, at 26 weeks and 52 weeks were compared for patients in the validation set. A summary of the metrics used to evaluate the integrated platform, including specific components, is provided in Table 2. All simulations were performed in Python 3.7 using the

Table 2 Summary of different metrics used for evaluating the performance of the integrated Crohn’s disease platform

Component	Evaluation method
Responder classifier	<ol style="list-style-type: none"> 1. Sensitivity and specificity of predicting endoscopic remission at 26 weeks post-treatment 2. Sensitivity and specificity of predicting mucosal healing at 26 weeks post-treatment
Virtual patient matching	Matching of affected surface area and ulcerated area at baseline (population-level distributions)
Overall platform (responder classifier + virtual patient matching + mechanistic model)	<ol style="list-style-type: none"> 1. Individual-level comparison of time-series predictions of affected surface area against data. Each individual prediction was labeled as “good,” “fair,” or “poor” based on the criteria described in the section “Platform Evaluation Approach” 2. Population-level distributions of disease severity in patients, as indicated by affected surface area and ulcerated area at 26 weeks and 52 weeks post-treatment

function `solve_ivp` with the integration method set to “LSODA” in the package SciPy (version 1.3). All other computational work was also done in Python 3.7. The simulation analysis plots were generated using Plotly Python Graphing Library.

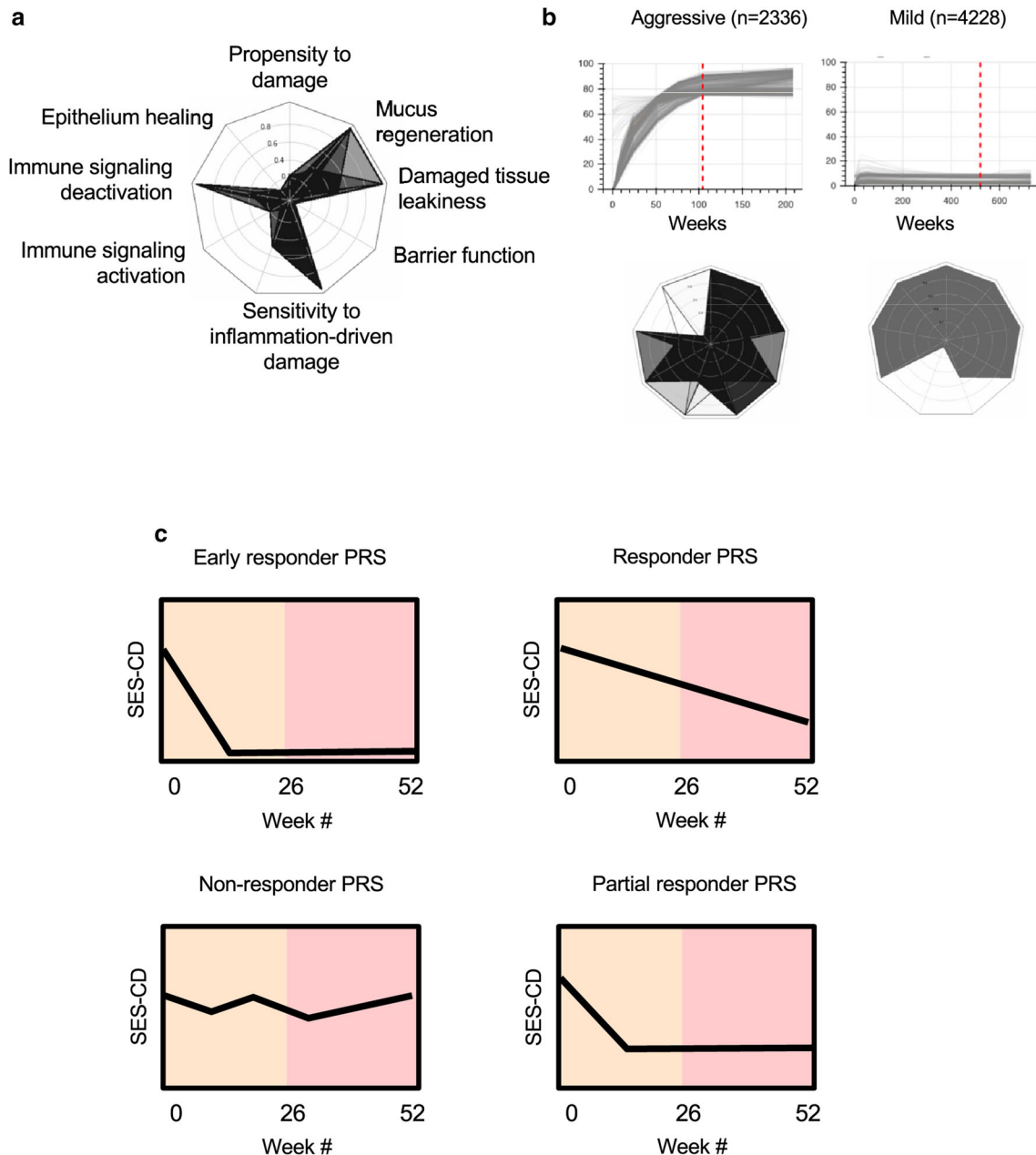


Fig. 4 Archetypical disease progression and responses to therapy. **a** Virtual patients were classified according to parameters corresponding to nine physiological processes (Sensitivity to inflammation-driven damage is represented by two spokes, one each for $\text{IFN}\gamma$ - and $\text{TNF}\alpha$ -dependent pathways). The extensions along the spokes in the spider diagram represent the values of the corresponding parameters, with outer layers representing larger values. Values for each parameter are normalized across the population so the scale of all spokes is 0–1. **b** Top panels: progression of damage for the “aggressive” and “mild” cohorts. Each line

represents a virtual patient (some removed for clarity). The red dotted line indicates when stratification was performed (2 and 10 years for aggressive and mild cohorts, respectively). Lower panel: superposition of footprints of all virtual patients in the aggressive (left) and mild (right) cohorts. Darker shades indicate more virtual patients’ footprints overlapping. Spokes should be interpreted as in **a**. **c** PRSs for this study. *PRS* progression-response scenario, *SES-CD* Simple Endoscopic Score for Crohn’s Disease

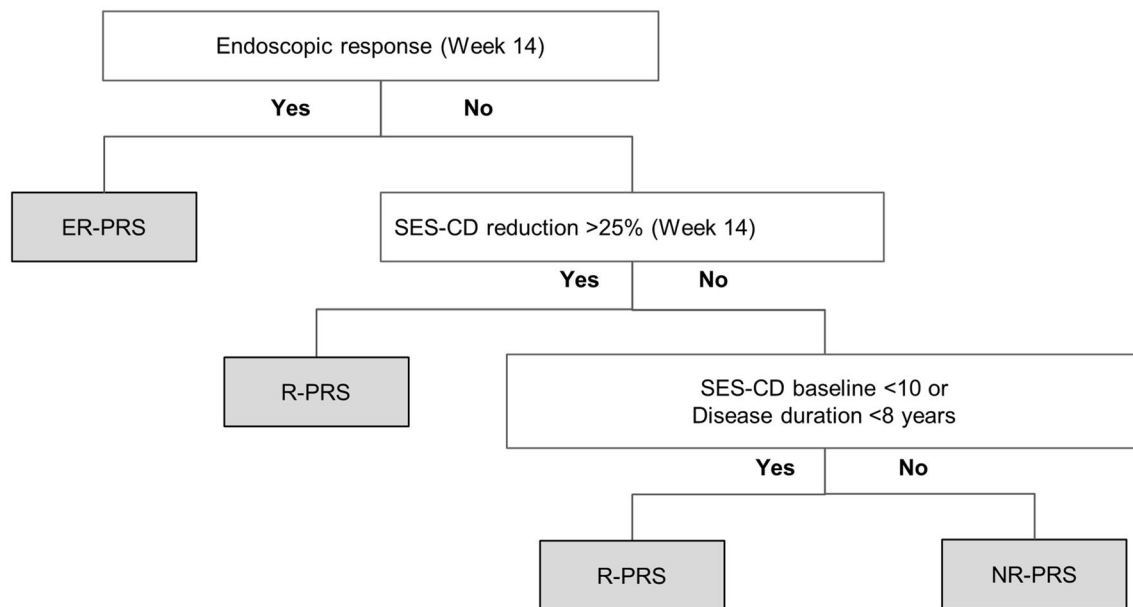


Fig. 5 Classification tree. The classification tree used for the classification stage assigns a PRS based on baseline and early response patient information. The PRS sets in this study were designed for prediction of mucosal health at 26 and 52 weeks after initiation of biologic treatment.

ER-PRS early responder progression-response scenario, *NR-PRS* nonresponder progression-response scenario, *PRS* progression-response scenario, *R-PRS* responder progression-response scenario, *SES-CD* Simple Endoscopic Score for Crohn’s Disease

Table 3 Confusion matrix for evaluating the performance of the responder classifier in predicting endoscopic remission at 26 weeks

Patient achieved endoscopic remission in 26 weeks	Patient assigned ER-PRS by the responder classifier	Correctness of classification
Yes	Yes	True positive
Yes	No	False negative
No	Yes	False positive
No	No	True negative

ER-PRS early responder progression-response scenario

Table 4 Evaluation of the responder classifier

Outcome predicted	Dataset (n)	Sensitivity (%)	Specificity (%)
Endoscopic remission at 26 weeks	Training (46)	75	76
	Validation (23)	100	57
	Overall (69)	80	69
Mucosal healing at 26 weeks	Training (46)	75	76
	Validation (23)	75	58
	Overall (69)	75	70

RESULTS

Virtual Library of Patients with CD

The library of virtual patients generated in this study contained approximately 35,000 patients. A period of 10 years after disease onset was simulated for each virtual patient under a variety of treatment scenarios. A measure of the quality of the virtual patient library is diversity of disease behavior. The library for this study was found to contain a diverse set of endoscopic response patterns to common treatments. Different disease-progression patterns in the absence of treatment are also present, such as patients with “aggressive” disease (corresponding to 75% or more of the GI tract impacted by disease in 2 years, but not sooner) or “mild” disease (10% or less GI tract damage in 10 years). The aggressive and mild disease cohorts contained 2336 and 4228 virtual patients, respectively. Further analysis of these two cohorts showed intra- and inter-cohort differences in the distribution of values for the sensitive mechanistic parameters (Fig. 4a, b, lower panels). The aggressive disease cohort is characterized by reduced healing capacity, whereas low sensitivity to inflammation-driven damage is strongly represented in the mild disease cohort. The between-cohort differences in the distribution of mechanistic parameters demonstrate that the temporal dynamics of tissue damage do introduce constraints in the model parameters as expected. This analysis resulted in a better characterization of model behavior and its strengths and limitations.

PRs

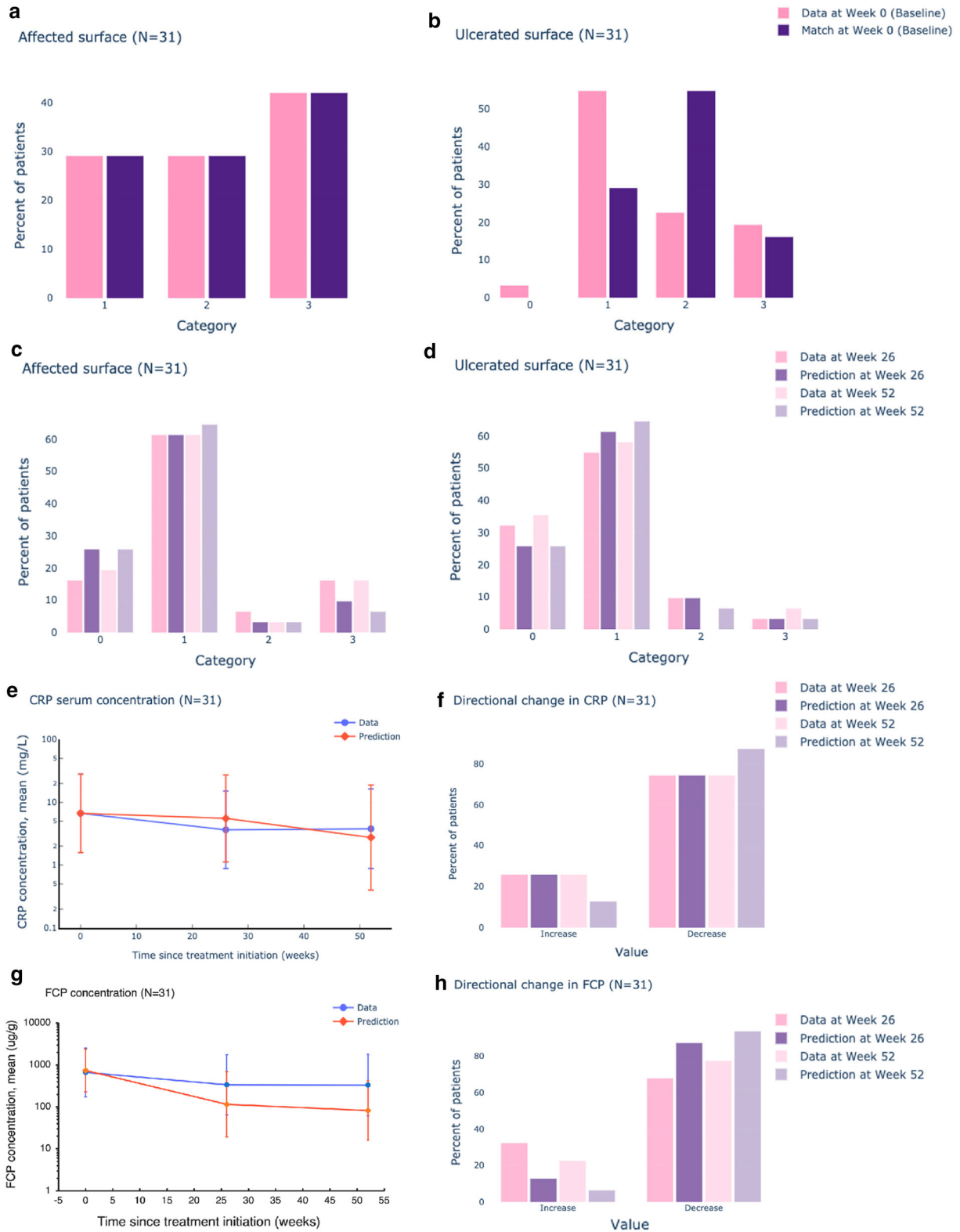
Four PRs were defined for this study. The early responder PR (ER-PR) is characterized by endoscopic remission ($SES-CD \leq 4$) in less than 26 weeks (Fig. 4c). The responder PR (R-PR) displays slower dynamics with a gradual drop in $SES-CD$ that continues over the 52 weeks of the study. The nonresponder PR (NR-PR) corresponds to $SES-CD$ fluctuations around baseline levels without a marked improvement or deterioration in $SES-CD$ score over time. The partial

Fig. 6 Performance of the integrated platform. **a, b** Comparison of affected surface area (**a**) and ulcerated surface (**b**) at baseline for digital twin (purple bars) and patient data (pink bars) in terms of $SES-CD$ categories. **c, d** Comparison of predicted (purple bars) and actual (pink bars) time evolution of mucosal healing at weeks 26 and 52 using the $SES-CD$ score categories for affected surface area (**c**) and ulcerated area (**d**). **e** Predicted (red) and observed (blue) geometric mean of serum CRP concentration in the validation cohort at baseline, 26 weeks, and 52 weeks. Error bars represent one geometric standard deviation factor. **f** Predicted (purple) and observed (pink) directional changes in levels of serum CRP for the validation cohort. **g** Predicted (red) and observed (blue) geometric mean of FCP concentration in the validation cohort at baseline, 26 weeks, and 52 weeks. Error bars represent one geometric standard deviation factor. **h** Predicted (purple) and observed (pink) directional changes in levels of FCP for the validation cohort. *CRP* C-reactive protein, *FCP* fecal calprotectin, *SES-CD* Simple Endoscopic Score for Crohn’s Disease

responder PR (PR-PR) is characterized by an initial reduction in $SES-CD$ in the first 26 weeks that either stalls or rebounds later. All patients in the study were manually assigned one of the PRs.

Classification into PRs

An algorithm to automate the classification of patients into PRs was developed. Of the alternatives evaluated, a classification tree based on combinations of early $SES-CD$ observations and historical data on disease duration produced the best results overall for the test cohort. The first test was done on evidence of endoscopic response (at least 50% reduction in $SES-CD$) by week 14 based on the $SES-CD$ score. Positives were mapped to the ER-PR group. Of the remaining patients, those showing a greater than 25% reduction in $SES-CD$ by week 14 were mapped to the R-PR. Patients failing to show a $SES-CD$ reduction of more than 25% by week 14 were also assigned to R-PR unless baseline $SES-CD$ was less than 10 or disease duration was less than 8 years, in which case they were assigned to NR-PR (Fig. 5). Including previous treatment (bio-naive or biologic) as a factor did not



improve the classification. Because of the limited number of patients with a long enough SES-CD history, classification into the PR-PRS was not attempted. To account for this limitation, patients assigned to R-PRS by the classification scheme were scored as “correct” if the original category was R-PRS or PR-PRS and “incorrect” otherwise. Similarly, patients assigned to ER-PRS by the classification scheme were scored as “correct” if the original category was ER-PRS or PR-PRS, and “incorrect” otherwise.

To evaluate the performance of the responder classifier, we used it to predict whether a patient would achieve endoscopic remission and mucosal healing at 26 weeks based on their PRS. A patient was labeled as true positive if they achieved endoscopic remission in 26 weeks, and the classifier mapped the patient to the ER-PRS group. Any other mapping was labeled as a false negative. If the patient did not achieve endoscopic remission in 26 weeks, but the classifier mapped the patient to the ER-PRS group, the patient would be labeled as a false positive. Any other mapping in the previous case was labeled as a true negative (Table 3). Sensitivity and specificity were calculated from the confusion matrix so created. Sensitivity and specificity for mucosal healing predictions at 26 weeks were determined in a similar manner. With the training set ($n = 46$), the classification scheme resulted in a sensitivity of 75% for ER-PRS, with 76% specificity for predicting endoscopic remission at 26 weeks. With the validation set ($n = 23$), the classification scheme produced sensitivity and specificity of 100% and 57%, respectively, for predicting endoscopic remission at 26 weeks (Table 4).

Prediction of Temporal Changes in Mucosal Health Due to Treatment

To evaluate the performance of the platform as a whole, we assigned digital twins to patients in the validation set and scored the predictions for tissue damage and biomarkers against their observed counterparts, as described in the section “[Integration of the Modules into a Platform](#).” Baseline surface area was similar between patients and their assigned digital twins,

whereas some deviations were observed for the ulcerated surface variable (Fig. 6a, b). This difference is due to the matching algorithm’s internal prioritization criteria, which gives preference to an exact match for an affected area, as predictions of mucosal healing dynamics are very sensitive to this specific variable, and mismatches can introduce significant errors given the size of the corresponding SES-CD categories (Table 1).

The comparison of predicted and observed disease severity at weeks 26 and 52 using the SES-CD score categories for affected surface area and ulcerated area is shown in Fig. 6c, d. There was no substantial difference between disease-severity distributions in data and predictions, indicating the response of the virtual patients created for the clinical study validation set was no different from the patient data at a population level. The integrated platform also predicted the overall directional change in biomarkers such as CRP and FCP (Fig. 6e–h). The geometric mean and geometric standard deviation factor of serum CRP and FCP were comparable between data and predictions. At an individual patient level, of 31 patients in the validation set, 22 (71%), seven (23%), and two (6%) scored good, fair, and poor, respectively, in predicting the effect of treatment on tissue damage, reported as affected surface area in the model (Fig. 7).

Incorrect PRS matches were not modified for this evaluation. As the assignment of PRS was the first step in the digital twin-based simulations, the predictive accuracy of tissue damage evolution was indicative of the overall performance of the hybrid platform. Eight patients who did not conform to any of the PRS motifs were assigned digital twins by matching against the whole virtual patient library.

DISCUSSION

A hybrid computational platform has been demonstrated that is capable of predicting the time course of mucosal health in patients with CD undergoing treatment with vedolizumab. Because of the causal relationships built into its mechanistic components, the platform provides

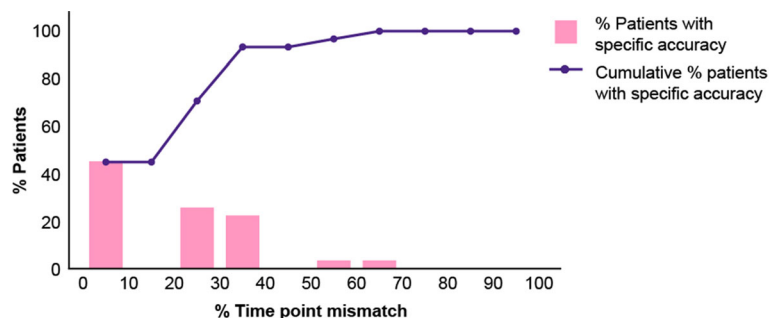


Fig. 7 Performance of the integrated platform measured by accuracy of affected surface area predictions at an individual patient level. The histogram shows the percentage of patients with specific accuracy in the predictions of affected surface area category. The *x*-axis represents percentage of time points with mismatch and the *y*-axis represents percentage of patients. The first bin (0–10) indicates the percentage of patients with mismatch in

affected surface area category predictions in up to 10% of the time points. The black line indicates cumulative frequency histogram. For example, the leftmost bar indicates that the platform correctly matched between 90% and 100% of the measurements (i.e., 0–10% mismatch) of affected surface area in 45% of the patients in the validation cohort as assessed across all available time points

time-resolved estimates of a range of variables related to inflammatory activity and biomarkers. The platform presented in this study is designed for flexibility. Both the PRSs used by the responder classifier and the virtual patient library can be modified or extended as needed.

The performance achieved by the classification stage in this study is comparable to related predictive tools for CD [14]. The first stage of the computational platform (i.e., responder classifier) was developed using the PRS assignments in the training dataset of the study cohort. The classifier was evaluated on its ability to predict endoscopic remission and mucosal healing in patients at 26 weeks by classifying them as ER-PRS. With an overall sensitivity of 75% and specificity of 70% in predicting mucosal healing, the performance of the classifier is comparable to the best possible outcomes for the prognostic clinical decision-support tool (CDST) [14] developed for vedolizumab (sensitivity 98% and specificity 30% for a 13-point cutoff in the CDST; sensitivity 40%, specificity 80% for a 19-point cutoff in the CDST for mucosal healing at 26 weeks). Whereas the CDST was developed to predict both clinical and endoscopic remission after 26 weeks of treatment, among other endpoints, the objective of the computational platform in

this study is to predict temporal evolution of mucosal health.

The performance of the classifier was sufficient for constraining the mechanistic part of the platform. Indeed, tracking of the time-series for tissue damage over 52 weeks, a more stringent test than endpoint matching, was considered good or fair for 94% of the patients in the test group. It is important to point out that the “resolution” of the matching is limited by the fact that affected area measurements are available only in terms of SES-CD categories, some of which are broad (Table 1). For example, the categorical nature of the data does introduce artifacts as changes in affected surface area (e.g., from at least 50% to less than 50%) translated into a unit category change. The platform also predicted overall directional change in biomarkers such as CRP and FCP. The means and standard deviations of serum CRP and FCP were comparable between data and predictions. However, the predicted mean FCP concentration was lower than the data at weeks 26 and 52. The scope of the CD model was limited to mechanisms related to inflammation-driven tissue damage and mucosal healing. Additional mechanisms may need to be incorporated to predict fecal biomarker trajectories with higher fidelity. Another limitation of the study was the coverage of the virtual patient library generated

in the study, which was not exhaustive. For example, the differences in disease progression due to other potential risk factors such as smoking are not captured by the current version of the platform. Future work with additional patient datasets could help identify gaps in the coverage of the virtual patient library and guide enhancements to the platform. In conclusion, despite the aforementioned limitations, the results in this study present digital twin-based simulations as a viable approach for predicting likely responses to treatments based on limited early treatment response data.

The inclusion of the mechanistic component brings several strengths to the platform but also challenges. Among the strengths is the ability to generalize the “learning” from a patient’s response to one treatment and extrapolate likely responsiveness to different regimens or drugs. This is possible because the response to a treatment on the rate of damage progression contains information about the pathophysiology underlying the disease in a particular patient. The mechanistic model introduces flexibility, as it can be interrogated to answer a variety of questions with time resolution and potentially informative variables not typically available to a clinician. As with most modeling approaches, validation remains a challenge that requires additional efforts, especially to assess the performance beyond the treatments tested in the present study. As further validation with other drugs occurs, this platform will enable the simulation of what-if experiments for estimating a patient’s response to different treatment alternatives, not just in terms of endpoint outcomes, but also incorporating the progression leading to those outcomes.

The hybrid approach is well suited to addressing some weaknesses that, despite their undeniable power, machine-learning approaches tend to share. Incorporating a mechanistic component eliminates the need for a model to “relearn” known physiological relationships during training, which may be difficult in scenarios with limited, incomplete, or noisy data. In contrast to machine-learning models trained to answer very specific sets of predetermined questions that cannot be easily repurposed, the approach demonstrated here is flexible, as the

model can be interrogated to answer different questions or investigate what-if scenarios. Finally, the approach is more transparent compared with “black box”-type machine-learning approaches, thus, at least in principle, allowing clinical users to leverage their expertise to assess confidence in the model predictions for specific patients. However, taking full advantage of the transparency afforded by the mechanistic component will likely require the platform to be wrapped into an informative patient- and physician-centric user experience.

While the mechanistic model offers a powerful scientific tool to dissect the underlying pathophysiology of disease, the computational complexity can make real-time use of the model challenging. Numerically solving a large system of differential equations can be time-consuming and, depending on the combination of software and hardware in use, simulating a few years of a patient’s life may require several seconds to minutes of clock time. Furthermore, calibration of model parameters to create an instance of the model tailored to a specific patient requires extensive exploration of the parameter space of the model. Any optimization of the algorithm would require running hundreds if not thousands of iterations of the model to find a near-optimal solution in a high-dimensional parameter space [41]. This process could require several minutes to hours on a typical personal computer. For this type of a platform to be useful for real-time applications (i.e., for physician–patient interactions), solutions must be found that eliminate or reduce the lag introduced by the calibration and simulation processes. Our approach of using a virtual library of patients alleviates this problem greatly because the computationally expensive step of exploring the parameter space is front-loaded in the process. The process of parameter calibration is replaced by a simple database search to find a pre-generated virtual patient that is sufficiently similar to the real patient, and the search can be executed on a subsecond-to-second timescale. A limitation of this approach is that the best match found in the library may not be good enough in some instances. The library must be continually updated and augmented as new

patients with inadequately matched digital twins are discovered.

Managing patients with CD often necessitates decisions to be made with limited visibility of the actual state of the patient's GI tract, in terms of both tissue quality and frequency of observation. Therefore, to alleviate pain and manage CD symptoms, patients are often prescribed medication on a "trial and error" approach. To improve the quality of life for patients with CD, there is a critical need to better understand how individual patients will respond to a new or adjusted treatment, not just in terms of outcomes but also in terms of the path to those outcomes. Computational models can help answer such questions by supplementing clinical data with insights based on system behaviors emerging from the known pathophysiological mechanisms of a disease [42, 43]. In particular, if the digital twin was available to provide physicians with the foresight of disease outcomes for a patient, then doctors could move from "trial and error" to a personalized approach with presumably fewer errors. For example, a physician could use the digital twin to evaluate if an early intervention (i.e., a top-down or accelerated step-up approach, among other strategies) is required to reduce bowel damage in a patient. The computational platform developed and validated in this study could be used to predict the evolution of tissue damage in response to treatment with vedolizumab. Similar training and validation for other biologics will enable simulation of an increasingly rich set of treatment options. A clinical tool based on this platform could enable a physician to forecast the effect of treatment on important clinical measures, e.g., ulcerated area, size of ulcers, and overall SES-CD score and biomarkers. Importantly, in this framework, such forecasts could be customized using the individual patient's historical data, rather than relying on a population average. These individualized predictions open possibilities for rationally comparing alternative treatments, based on model-informed prognosis. A gastroenterologist could use the digital twin to engage patients in conversations about treatment goals with objective outcomes as a focus, and jointly decide upon a treatment plan.

Engaging patients in shared decision-making could improve their adherence to the treatment plan and overall prognosis [13]. Consequently, we expect that the platform developed in this study will eventually enable model-based clinical decision-support tools, to aid physicians in managing treatment plans and facilitate shared decision-making with patients. However, to enable such clinical use, our predictive platform needs to undergo more rigorous testing, validation, and refinement. While we have retrospectively tested our predictions against a limited patient sample, larger-scale qualification, and virtual population augmentation against a more comprehensive database of patients with CD, preferably a real-world cohort, is the next logical step. This will build confidence in the tool, prior to adoption by practicing gastroenterologists.

Model development to date has focused on "virtualizing" the research process to accelerate drug development and optimize clinical trial design. The concept of model-based drug development, or model-informed drug development, usually involves pharmacokinetic and/or pharmacodynamic data and other laboratory measures [44]. The term "in silico trial" refers to the use of computer modeling and simulation for the testing and evaluation of a drug or medical device, with the aim of reducing, refining, or replacing the number of in vivo tests in animals and humans [45]. Computational models must be verified and validated to be classified as a "credible" methodology to be included in regulatory submissions [46]. The computational platform described in this study extends beyond the context of clinical development to simulate possible clinical implications of managing patients with chronic disease. The inclusion of both evidence-based mechanisms of disease and the validation of the digital twin using real patient data support the credibility of this computational model to simulate the disease course [46]. Thus, in addition to the real-world application of forecasting treatment outcomes, the inclusion of the digital twin technology within in silico trials could "refine" human trials [45]. Furthermore, data accumulated during clinical development and from real-world experience of a drug could

inform the model in a way that is not possible during development.

We believe that predictive approaches based on sophisticated computational models, like the one presented in this work, will play an important role in clinical practice in the future. An important aspect in the clinician's adoption of advanced analytical tools is appreciation and understanding of the value these tools could add to their clinical practice. Knowing the strengths and pitfalls of these tools could help clinicians critically synthesize predictive model outputs into their own knowledge in treating patients. We hope that this work, aside from its technical value, will also contribute to awareness about model-based analytical tools and their promise to improve patients' lives in the real world.

ACKNOWLEDGEMENTS

Funding. This work was supported by Takeda Pharmaceutical Company Limited, Tokyo, Japan, and by PwC Consulting LLC, Tokyo, Japan. PwC Consulting LLC received funding from Takeda Pharmaceutical Company Limited for some aspects of this study. The journal's Rapid Service and Open Access Fees were funded by Takeda Pharmaceutical Company Limited.

Medical Writing and Editorial Assistance. The authors would like to thank Philippe Pinton, MD, PhD, for comments that greatly improved the manuscript. Medical writing assistance was provided by MIMS, and editing of the Discussion was conducted by Prudence Stanford, PhD, of ProScribe–Envision Pharma Group. MIMS and ProScribe were funded by Takeda Pharmaceutical Company Limited. ProScribe's services complied with international guidelines for Good Publication Practice (GPP3).

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of

the work as a whole, and have given their approval for this version to be published.

Author Contributions. SPV, RI, EU, ZQ, NP, MB, and JF contributed to the conception of the work. SPV, ZQ, NP, MB, and JF contributed to the design of the work. SPV contributed to the acquisition of data. SPV, ZQ, and NP contributed to the analysis of data. SPV, RI, EU, RT, KK, YS, MB, PO, GD, HM, RKS, ALH, PMD, and JF contributed to the interpretation of data. All authors were involved in critical review, editing, revision, and approval of the final manuscript.

Prior Presentation. The results of this study were presented at the United European Gastroenterology Week (UEGW) congress between the 3–5 of October 2021.

Disclosures. At the time of this research, Sai Phanindra Venkatapurapu, Nikhil Patidar, Zhen Qi, Ryoko Takayama, Kei Kumar, Yuki Sato, Marcelo Behar, Patrick Offner, Gaurav Dwivedi, and Paul M. D'Alessandro were employees of PricewaterhouseCoopers. Ryuichi Iwakiri, Eri Udagawa, Haruna Miyasaka, Ryohsuke Ken Suzuki, and Anna Lissa Hamada are employees of Takeda Pharmaceutical Company Limited. Jovelle Fernandez was an employee of Takeda Pharmaceutical Company Limited at the time the work was conducted, has restricted stock of Takeda Pharmaceutical Company Limited and GSK, and is President of Jovelle Fernandez LLC.

Compliance with Ethics Guidelines. The study was conducted according to the Declaration of Helsinki and the International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices. It was carried out retrospectively using a database of anonymized data, following Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese Ministry of Health, Labour and Welfare. The investigators could only access anonymized information from the database; therefore, in accordance with the Japanese Ethical Guidelines mentioned above, institutional ethics approval and informed consent were not required.

Data Availability. The data underlying the development of the platform are available in the article, references and in the online Supplementary Material. To validate the platform, the authors obtained access to patient data from the sponsors of the VERSIFY study. Any requests regarding the data should be made to the sponsors of the VERSIFY study. The datasets, including the redacted study protocol, redacted statistical analysis plan, and individual participants' data supporting the results reported in this article, will be made available within 3 months from initial request, to researchers who provide a methodologically sound proposal. The data will be provided after its de-identification, in compliance with applicable privacy laws, data protection and requirements for consent and anonymization.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Baumgart DC, Sandborn WJ. Crohn's disease. *Lancet*. 2012;380:1590–605.
- Boyapati R, Satsangi J, Ho GT. Pathogenesis of Crohn's disease. *F1000 Prime Rep*. 2015;7:44.
- Hibi T, Ogata H. Novel pathophysiological concepts of inflammatory bowel disease. *J Gastroenterol*. 2006;41:10–6.
- Okumura R, Takeda K. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med*. 2017;49: e338.
- Sands BE. Inflammatory bowel disease: past, present, and future. *J Gastroenterol*. 2007;42:16–25.
- Loftus EV Jr, Sandborn WJ. Epidemiology of inflammatory bowel disease. *Gastroenterol Clin N Am*. 2002;31:1–20.
- Gersemann M, Stange EF, Wehkamp J. From intestinal stem cells to inflammatory bowel diseases. *World J Gastroenterol*. 2011;17:3198–203.
- Colombel J-F. Introduction: where are we in 2017? *EMJ Gastroenterology*. 2017;6:8.
- Peyrin-Biroulet L, Panes J, Sandborn WJ, et al. Defining disease severity in inflammatory bowel diseases: current and future directions. *Clin Gastroenterol Hepatol*. 2016;14:348-54 e17.
- Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389:1741–55.
- Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. *Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID)*. *Gut*. 1989;30: 983–9.
- Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest Endosc*. 2004;60:505–12.
- Panes J, O'Connor M, Peyrin-Biroulet L, Irving P, Petersson J, Colombel JF. Improving quality of care in inflammatory bowel disease: what changes can be made today? *J Crohns Colitis*. 2014;8:919–26.
- Dulai PS, Boland BS, Singh S, et al. Development and validation of a scoring system to predict outcomes of vedolizumab treatment in patients with Crohn's disease. *Gastroenterology*. 2018;155:687-95 e10.
- Siegel CA, Horton H, Siegel LS, et al. A validated web-based tool to display individualised Crohn's disease predicted outcomes based on clinical, serologic and genetic variables. *Aliment Pharmacol Ther*. 2016;43:262–71.
- Park A, Kim S, Jung IH, Byun JH. An immune therapy model for effective treatment on

- inflammatory bowel disease. *PLoS One*. 2020;15: e0238918.
17. Lo WC, Arsenescu V, Arsenescu RI, Friedman A. Inflammatory bowel disease: how effective is TNF-alpha suppression? *PLoS One*. 2016;11: e0165782.
 18. Rogers KV, Martin SW, Bhattacharya I, Singh RSP, Nayak S. A dynamic quantitative systems pharmacology model of inflammatory bowel disease: Part 1—model framework. *Clin Transl Sci*. 2021;14: 239–48.
 19. Rogers KV, Martin SW, Bhattacharya I, Singh RSP, Nayak S. A dynamic quantitative systems pharmacology model of inflammatory bowel disease: Part 2—application to current therapies in Crohn's disease. *Clin Transl Sci*. 2021;14:249–59.
 20. Hernandez-Lemus E, Siqueiros-Garcia JM. Mechanistic-enriched models: integrating transcription factor networks and metabolic deregulation in cancer. *Theor Biol Med Model*. 2015;12:16.
 21. Wang S, Fan K, Luo N, et al. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. *Nat Commun*. 2019;10:4354.
 22. Sarkar J, Dwivedi G, Chen Q, et al. A long-term mechanistic computational model of physiological factors driving the onset of type 2 diabetes in an individual. *PLoS One*. 2018;13: e0192472.
 23. Roberts P, Spiros A, Geerts H. A humanized clinically calibrated quantitative systems pharmacology model for hypokinetic motor symptoms in Parkinson's disease. *Front Pharmacol*. 2016;7:6.
 24. Pena-Chilet M, Esteban-Medina M, Falco MM, et al. Using mechanistic models for the clinical interpretation of complex genomic variation. *Sci Rep*. 2019;9:18937.
 25. Danese S, Sandborn WJ, Colombel JF, et al. Endoscopic, radiologic, and histologic healing with vedolizumab in patients with active Crohn's disease. *Gastroenterology*. 2019;157:1007–18 e7.
 26. Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol*. 2015;110:1324–38.
 27. Catt H, Hughes D, Kirkham JJ, Bodger K. Systematic review: outcomes and adverse events from randomised trials in Crohn's disease. *Aliment Pharmacol Ther*. 2019;49:978–96.
 28. Balbas-Martinez V, Ruiz-Cerda L, Irurzun-Arana I, et al. A systems pharmacology model for inflammatory bowel disease. *PLoS One*. 2018;13: e0192949.
 29. Dwivedi G, Fitz L, Hegen M, et al. A multiscale model of interleukin-6-mediated immune regulation in Crohn's disease and its application in drug discovery and development. *CPT Pharmacometr Syst Pharmacol*. 2014;3: e89.
 30. Wendelsdorf K, Bassaganya-Riera J, Hontecillas R, Eubank S. Model of colonic inflammation: immune modulatory mechanisms in inflammatory bowel disease. *J Theor Biol*. 2010;264:1225–39.
 31. Eckmann L, Kagnoff MF, Fierer J. Epithelial cells secrete the chemokine interleukin-8 in response to bacterial entry. *Infect Immun*. 1993;61:4569–74.
 32. Kucharzik T, Hudson JT 3rd, Lugering A, et al. Acute induction of human IL-8 production by intestinal epithelium triggers neutrophil infiltration without mucosal injury. *Gut*. 2005;54:1565–72.
 33. Mazzucchelli L, Hauser C, Zraggen K, et al. Expression of interleukin-8 gene in inflammatory bowel disease is related to the histological grade of active inflammation. *Am J Pathol*. 1994;144: 997–1007.
 34. Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med*. 2009;361:2066–78.
 35. Turner JR. Molecular basis of epithelial barrier regulation: from basic mechanisms to clinical application. *Am J Pathol*. 2006;169:1901–9.
 36. Pastorelli L, De Salvo C, Mercado JR, Vecchi M, Pizarro TT. Central role of the gut epithelial barrier in the pathogenesis of chronic intestinal inflammation: lessons learned from animal models and human genetics. *Front Immunol*. 2013;4:280.
 37. Mizoguchi A, Yano A, Himuro H, Ezaki Y, Sadanaga T, Mizoguchi E. Clinical importance of IL-22 cascade in IBD. *J Gastroenterol*. 2018;53:465–74.
 38. Neurath MF. Cytokines in inflammatory bowel disease. *Nat Rev Immunol*. 2014;14:329–42.
 39. Bower JK, Lazo M, Juraschek SP, Selvin E. Within-person variability in high-sensitivity C-reactive protein. *Arch Intern Med*. 2012;172:1519–21.
 40. Cremer A, Ku J, Amininejad L, et al. Variability of faecal calprotectin in inflammatory bowel disease patients: an observational case-control study. *J Crohns Colitis*. 2019;13:1372–9.
 41. Wang QJ. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resour Res*. 1991;27:2467–71.

-
42. Tawhai MH, Clark AR, Wilsher ML, Milne DG, Subramaniam K, Burrowes KS. Spatial redistribution of perfusion and gas exchange in patient-specific models of pulmonary embolism. 9th IEEE Int Sympos Biomed Imaging (ISBI). 2012;1365–1368. <https://doi.org/10.1109/ISBI.2012.6235821>.
 43. Winslow RL, Trayanova N, Geman D, Miller MI. Computational medicine: translating models to clinical care. *Sci Transl Med*. 2012;4:158rv11.
 44. Kimko H, Pinheiro J. Model-based clinical drug development in the past, present and future: a commentary. *Br J Clin Pharmacol*. 2015;79:108–16.
 45. Viceconti M, Emili L, Afshari P, et al. Possible contexts of use for in silico trials methodologies: a consensus-based review. *IEEE J Biomed Health Inform*. 2021;25(10):3977–82.
 46. Viceconti M, Pappalardo F, Rodriguez B, Horner M, Bischoff J, Musuamba Tshinanu F. In silico trials: verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods*. 2021;185:120–7.