Open

Constipation Predominant Irritable Bowel Syndrome and Functional Constipation Are Not Discrete Disorders: A Machine Learning Approach

James K. Ruffle, MBBS, BSc[1,2,3], Linda Tinkler, MSc[4], Christopher Emmett, MD, FRCS[4], Alexander C. Ford, MBChB, FRCP[5], Parashkev Nachev, FRCP, PhD[3], Qasim Aziz, FRCP, PhD[1], Adam D. Farmer, FRCP, PhD[1,6] and Yan Yiannakou, MBChB, MRCP, MD[4]

INTRODUCTION: Chronic constipation is classified into 2 main syndromes, irritable bowel syndrome with constipation (IBS-C) and functional constipation (FC), on the assumption that they differ along multiple clinical characteristics and are plausibly of distinct pathophysiology. Our aim was to test this assumption by applying machine learning to a large prospective cohort of comprehensively phenotyped patients with constipation.

METHODS: Demographics, validated symptom and quality of life questionnaires, clinical examination findings, stool transit, and diagnosis were collected in 768 patients with chronic constipation from a tertiary center. We used machine learning to compare the accuracy of diagnostic models for IBS-C and FC based on single differentiating features such as abdominal pain (a "unisymptomatic" model) vs multiple features encompassing a range of symptoms, examination findings and investigations (a "syndromic" model) to assess the grounds for the syndromic segregation of IBS-C and FC in a statistically formalized way.

RESULTS: Unisymptomatic models of abdominal pain distinguished between IBS-C and FC cohorts near perfectly (area under the curve 0.97). Syndromic models did not significantly increase diagnostic accuracy ($P >$ 0.15). Furthermore, syndromic models from which abdominal pain was omitted performed at chance-level (area under the curve 0.56). Statistical clustering of clinical characteristics showed no structure relatable to diagnosis, but a syndromic segregation of 18 features differentiating patients by impact of constipation on daily life.

DISCUSSION: IBS-C and FC differ only about the presence of abdominal pain, arguably a self-fulfilling difference given that abdominal pain inherently distinguishes the 2 in current diagnostic criteria. This suggests that they are not distinct syndromes but a single syndrome varying along one clinical dimension. An alternative syndromic segregation is identified, which needs evaluation in community-based cohorts. These results have implications for patient recruitment into clinical trials, future disease classifications, and management guidelines.

INTRODUCTION

Chronic constipation is a common symptom, with a pooled prevalence of 14% (1). Although there are multiple causes of chronic constipation, among the most prevalent are irritable bowel syndrome with constipation (IBS-C) and functional constipation (FC). These functional bowel disorders are delineated by the Rome classification, now in its fourth iteration (2–5), based on the presence or absence of recurrent abdominal pain and its relation to defecation, stool frequency, or appearance (Table 1 and Table 2) (2,3).

A syndrome, by definition, is "a group of signs and symptoms that occur together and characterize a particular abnormality or condition" (6). Segregating syndromes with very similar symptom profiles

Table 1. Rome III and IV criteria for functional constipation

Rome III l Functional Constipation
For at least the last 3 months with symptom onset ≥ 6 months ago; ≥ 2 of the following (for > 25% of defecations):
Straining
Lumpy or hard stools
Sensation of incomplete evacuation
Sensation of anorectal obstruction/blockage
Manual maneuvers to facilitate defecation
<3 defecations per week
Must have both of the following:
Loose stools rarely present without laxative use
Does not meet Rome III criteria for IBS
Rome IV l Functional Constipation
For at least the last 3 months with symptom onset ≥ 6 months ago; ≥ 2 of the following (for > 25% of defecations):
Straining
Lumpy or hard stools
Sensation of incomplete evacuation
Sensation of anorectal obstruction/blockage
Manual maneuvers to facilitate defecation
<3 spontaneous bowel movements per week
Must have both of the following:
Loose stools rarely present without laxative use
Does not meet Rome IV criteria for IBS

Table 2. Rome III and IV criteria for irritable bowel syndrome

Rome III l Irritable Bowel Syndrome
Recurrent abdominal pain or discomfort, 3 days/month in the last 3 months, associated with ≥ 2 of the following:
Improvement with defecation
Onset associated with change in stool frequency
Onset associated with change in stool form (appearance)
Rome IV l Irritable Bowel Syndrome
Recurrent abdominal pain at least one day per week in the last 3 months on average, associated with ≥ 2 of the following:
Related to defecation (increasing or improving pain)
Associated with change in stool frequency
Associated with change in stool form (appearance)

value is uncertain. Previous studies, including those from primary care, epidemiological and smaller-scale investigatory studies, have questioned the mutual exclusivity of IBS-C and FC as disease entities, suggesting instead that the 2 lie on a single spectrum (7–12). Should this be the case, the differentiation of the 2 disorders as discrete is arguably confusing for both the patient and healthcare provider. An example of this would be a previous study illustrating that patients with FC and IBS-C will frequently switch diagnostic label over time, from IBS-C to FC, and *vice versa* (8). Management algorithms and treatment trials are often constrained by these categories, being limited to either IBS-C or FC, so establishing the validity of this differentiation is crucial to clinical practice.

Previous studies, although informative in this regard, have generally focused on one specific aspect of clinical data only, such as reporting demographic differences between IBS-C and FC. The advent of data science and machine learning represents an opportunity to model the interpretation of complex, multidimensional data, such as the clinical history, patient-reported outcome measures, examination, and investigation findings (13–17). In addition, machine learning or graph-based network analyses provide an innovative ability for models to learn complex interaction between large numbers of patient factors, for instance, how a given patient's demographic, clinical history, psychophysiology, examination findings, and investigations all interact to provide an overall disease phenotype (13). Given these disorders are inherently complex, involving multiple disease facets, it is argued the use of such an approach is wise.

Therefore, the primary aim of this study was to assess the value of Rome criteria-delineated diagnostic classification labels in patients with symptoms of chronic constipation using a machine learning approach. Our hypothesis was that the diagnostic groups would differ along visorosensory measures, especially pain, but otherwise would be equivalent, therefore evidencing that IBS-C and FC are together a single syndrome varying along one dimension, as opposed to 2 distinct syndromes. We tested this hypothesis by comparing the diagnostic accuracy of "unisymptomatic" vs "syndromic" statistical models, relying on machine learning to make comparisons between complex patterns of clinical features. If the former (unisymptomatic models) perform equivalent or superiorly to syndromic models, a single syndrome would be a better description of this patient population. If the latter (syndromic models) perform better, the presence of 2 separate disorders would be supported. We further used statistical clustering to find any natural syndromic segregation within the patient cohort.

into single entities implies the presence of a number of differences, plausibly these would include differences in causal mechanisms and therefore would be worth preserving. If 2 disorders differ only by a single feature, varying "unisymptomatically," the basis for expecting dissimilarity of causal mechanisms is much weaker. Moreover, if such segregation is predetermined by diagnostic criteria only, as opposed to pathophysiological or treatment response differences, then its

METHODS
Study design
Study data were collected prospectively at the Durham Bowel Dysfunction Service, a single tertiary referral center in Northern England. Approximately 50% of patients are referred directly from their primary care physician (catchment area of 250,000 individuals) and the other 50% as tertiary referrals (catchment area of 3 million individuals). The study is reported in accordance with the STROBE cohort study checklist (see Supplementary Material, Supplementary Digital Content 9, http://links.lww.com/AJG/B632). Ethical approval was obtained from the local Research Ethics Committee (Ref: 09/H0906/86).

Study participants and data collection
Seven hundred sixty-eight consecutive adult patients with chronic constipation attending the Durham Bowel Dysfunction Service

over 7 years were invited to give their consent to have their data recorded on a prospectively maintained database. All bar 2 patients consented for study inclusion. This accrued a large and comprehensively phenotyped group of patients, albeit selective given that all had symptoms severe enough to warrant tertiary center referral.

Data collection was prospective, protocol based and supplemented by careful validation from source data. Patients were classified using the Rome III criteria (2), which were the gold standard at the time of data collection, into those with IBS-C or FC. Classification was assessed by proforma-based interview questions undertaken by experienced clinicians, which included the Rome III criteria for IBS-C and FC (2). All 6 coding clinicians worked in the team for a minimum of 2 years and attended regular team meetings. Patients with secondary causes of constipation such as underlying neurological conditions or drug-induced (including opioid-induced) constipation were excluded. Other exclusion criteria included patients who did not fulfill the diagnostic criteria for either IBS-C or FC, those with frequent diarrhea (except when due to laxative use) suggestive of mixed stool pattern IBS, or those who were unable to provide informed consent. Patients with coexistent pelvic floor dysfunction (PFD) were included (as per the Rome III classification), with symptoms of PFD assessed by symptoms and defecating proctogram and included in the statistical analysis. The rationale of this was that although PFD can frequently coexist in this patient group, its presence does not differentiate the disorders in current diagnostic criteria. After exclusion criteria and removal of subjects with missing data, this accrued a cohort of 661 patients, 365 with IBS-C and 296 with FC (see Supplementary Figure 1, Supplementary Digital Content 5, http://links.lww.com/AJG/B621).

Demographic and clinical data were collected using a standardized proforma, adapted from the Cleveland Clinic Score (18), with patients asked to provide information on duration of symptoms, age at onset of symptoms, symptom characteristics including evacuatory dysfunction, abdominal pain, abdominal bloating, tenesmus, stool frequency, and stool consistency, as well as the relationship of these symptoms to laxative use. These symptoms were rated by the patient on a 5-point Likert scale (from 0 to 4). Stool frequency and consistency were converted to the same categorical scale according to predetermined groupings. All Rome III-based symptom domains were also assessed by the clinician according to standardized questions, allowing the classification of patients into IBS-C or FC. As per the Rome criteria, where IBS-C was diagnosed, FC was excluded (n = 24) (2). Patients additionally reported the presence of extracolonic symptoms, including genitourinary symptoms (urinary frequency, nocturia, and stress incontinence), and the presence of nausea and vomiting. All patients completed the validated Patient Assessment of Constipation symptoms and quality of life (QOL) questionnaires at their first clinic visit (19,20). We chose not to use stool diary-based data measures. The reasons for this were that first, the use of stool diaries in routine clinical practice is not particularly high. Second, we wished to emulate data variables that could be obtained from a general gastroenterology clinic. Finally, we were aware of the argument that stool diaries can confound data analysis by factors such as poor completion or under-reporting of measures (21).

All patients underwent a radio-opaque marker transit study, wherein transit time was determined according to a modified Metcalf protocol (22). All laxatives were stopped for this assessment. Both segmental and total transit were assessed, with capsules of 24 identical markers administered on each of the first 3 days.

After the cessation of data collection, a period of data cleaning and validity testing was performed, with retrospective searching of medical records and investigation results to verify the accuracy of data entry and to account for missing values. When this process was completed, the database was locked for analysis to take place.

Statistical analysis

We provide a comprehensive description of all data measures quantified and used in the analysis within the Supplementary Data (see Supplementary Digital Content 4, http://links.lww.com/AJG/B620). All statistical tests performed were corrected for multiple comparisons by the means of the Benjamini & Hochberg/Yekutieli false-discovery rate to reduce the risk of type one error (23,24). All P values depicted in the manuscript are corrected, and our false-discovery rate-*corrected* statistical criterion was $P < 0.05$.

Principal component analysis

Given the large amount of data collected from each patient, it was likely that there would be a number of redundant features. To exclude any such redundancy, we used principal component (PC) analysis to generate unique features, which would combine redundant overlapping data. This approach is further described in the Supplementary Material (see Supplementary Digital Content 3, http://links.lww.com/AJG/B619).

Supervised machine learning to differentiate IBS-C or FC

We used machine learning, an area of computer science whereby a system can develop the ability to "learn" with data without explicit programming. This advanced statistical modeling has gained significant traction in recent years, both in the commercial sector, such as with driverless cars or robotics, and academia/health care, such as in automated lesion detection in endoscopy (25). A review primer of machine learning with a focus to gastroenterology is provided here for further reading (13). Specifically, we used machine learning to ascertain whether IBS-C or FC could be accurately distinguished by modeling (13). Machine learning affords the opportunity to build models, wherein a complex, multidimensional set of features, such as patient demographic, symptoms, examination findings, and investigation results, are coalesced to predict an outcome—in this case, the patient's diagnosis of IBS-C or FC. Our rationale for this was that if a machine could accurately distinguish and classify one diagnosis over another only when using many features, it would support a distinction between the 2 disease groups as individual syndromes. However, should a model be equally accurate when provided with a single feature only, i.e. unisymptomatic, then it would support that the 2 disorders are not distinct syndromes but rather are on a spectrum of that specific feature. This approach is further discussed in the Supplementary Material (see Supplementary Digital Content 3, http://links.lww.com/AJG/B619).

Unsupervised machine learning to identify clustering patterns in patients with chronic constipation

We considered that supervised machine learning models might not support IBS-C and FC as distinct, multifaceted syndromes. Should this be the case, we sought to determine if there was instead a robust clustering structure of patients into disease phenotypes, i.e., a strictly syndromic model using a large array of unique features. First, this was used to ascertain if, when provided with clinical data, a machine would identify the presence of multiple disease entities as per current diagnostic criteria (i.e., IBS-C and FC). Second, we also considered

that this unsupervised approach may in fact yield multiple subgroups within disease phenotypes, or the converse, that in fact there would be no distinguishable groups at all. To investigate this, we undertook 2-step cluster analysis, in addition to dimension reduction with uniform manifold approximation and projection (UMAP), which we describe in the Supplementary Material (see Supplementary Digital Content 3, http://links.lww.com/AJG/B619) (26).

RESULTS

Study population

After exclusion of 34 patients with secondary causes of constipation, 25 who did not fulfill the Rome III diagnostic criteria for either IBS-C or FC, and 48 patients with incomplete data, we included 661 individuals (597 women, mean age ± SD 41.76 ± 15.38) in the final analysis (see Supplementary Figure 1, Supplementary Digital Content 5, http://links.lww.com/AJG/B621). This cohort consisted of 365 patients with IBS-C (322 [91.0%] women, mean age 39.93 ± 14.91) and 296 with FC (257 [86.8%] women, mean age 44.02 ± 15.68). Stool transit study results did not significantly differ between the IBS-C and FC groups (see Supplementary Material and Supplementary Figure 2, Supplementary Digital Contents 3 and 6, http://links.lww.com/AJG/B619 and http://links.lww.com/AJG/B622).

Principal component analysis identifies unique components of chronic constipation

First, a pairwise correlation matrix of all clinical data was generated (see Supplementary Figure 3, Supplementary Digital Content 7, http://links.lww.com/AJG/B623). Kaiser-Meyer-Olkin measure of sampling (0.72) and the Bartlett test of sphericity ($P < 0.0001$) indicated that the data matrix was suitable for dimension reduction techniques and therefore appropriate for machine learning later. Oblimin with Kaiser normalization PC analysis (eigenvalue threshold > 1) identified 23 PCs that accounted for 69% of the total variance and converged in 44 iterations. The PCs with the largest contributing variance and eigenvalues were as follows: (i) PC1: QOL (which included self-confidence, condition obsession, dietary impact, impact on daily routine, anxiety over dietary choices, stress, decreased appetite, and embarrassment and upset regarding both their condition and stool frequency), 20% of total variance, eigenvalue 17; (ii) PC2: viscerosensory (which included stomach discomfort and cramping, frequency and severity of abdominal pain, abdominal bloating, nausea and vomiting, Supplementary Data (see Supplementary Digital Content 4, http://links.lww.com/AJG/B620), 5% of total variance, eigenvalue 4.3; (iii) PC3: stool frequency symptoms, 5% of total variance, eigenvalue 3.6; and (iv) PC4: satisfaction relating to disorder (including treatment), 4% of total variance, eigenvalue 3.1. A full list of PCs with respective eigenvalues > 1 and cumulative variance is shown in Figure 1, a–c. We also considered the possibility that these 23 retrieved features may not adequately capture all important variance characteristics between IBS-C and FC, given a variance coverage of 69%. Therefore, for subsequent machine learning, we also expanded our syndromic model to include all PCs to fulfill 95% of the total variance.

We reviewed the similarities, and dissimilarities, of the PCs between IBS-C and FC. There were only 2 significant differences between the diagnostic labels: (i) PC6: abdominal pain (which specifically included abdominal pain related to stool frequency, appearance of bowel movements, and association to defecation, favoring patients with IBS-C to have more of these features (t 16.30; corrected $P < 0.0001$) and (ii) PC2: viscerosensory

favoring patients with IBS-C to report more viscerosensory symptoms (t 4.33, corrected $P < 0.0001$). Individual parameter constituents for these 2 PCs are available as Supplementary Data, (see Supplementary Digital Content 4, http://links.lww.com/AJG/B620). Notably, PC6 was not the *sole* component featuring visceral pain either—PC23: rectal pain/bleeding showed no significant difference between the diagnostic groups (t 2.40, corrected $P > 0.05$). The remaining PCs were not significantly different, which included symptom domains, QOL measures, demographics, stool frequency measures, and transit time. Strip plots of individual patients, per diagnosis, per PC (with eigenvalue > 1), are shown in Figure 1d, which also highlights the considerable overlap between the 2 groups in nearly all domains.

Abdominal pain alone is necessary and sufficient for differentiating IBS-C and FC

Having dimension reduce the data into distinct features (PCs), we used machine learning to compare the accuracy of diagnostic models based on single differentiating characteristics (unisymptomatic) vs multiple characteristics (syndromic). The specific predictive aim was to classify patients into their Rome III-derived diagnostic label of either IBS-C or FC.

First, we trained a unisymptomatic classifier to predict a diagnosis of IBS-C or FC using PC6: abdominal pain alone. All models performed near perfectly at distinguishing IBS-C and FC (all area under the curves [AUCs] 0.97) (Figure 2a). Second, we trained a classifier using PC2: viscerosensory alone. These models performed poorly, the "best" performer of which was the neural network (AUC 0.64) (Figure 2b). Next, we trained a syndromic classifier using all PCs, which totaled 95% of the cumulative variance of the total dataset. This included the 23 PCs with eigenvalues > 1 (as depicted in Figure 1), but additionally a further 31 PCs (all of which had eigenvalues ranging from 0.44 to 0.99). Best model performance was the neural network (AUC 0.98) (Figure 2c). Diagnostic accuracy of the best performing syndromic model did not differ significantly from the best performing unisymptomatic model using PC6: abdominal pain alone ($P > 0.15$). Of note, the abdominal pain unisymptomatic K-nearest neighbor model performed significantly better than the fully syndromic model ($P < 0.0008$), suggesting that, in fact, the presence of features other than abdominal pain was confusing to the model in differentiating the 2 disorders, rather than being beneficial.

We then evaluated the performance of the syndromic model using all PCs (totaling 95% of cumulative variance), *except* we withheld the PC6: abdominal pain feature. These models all performed poorly with chance-level accuracy, the best model of which was a Gaussian naive Bayes (AUC 0.63). This syndromic model performed significantly worse than the unisymptomatic PC6: abdominal pain model ($P < 0.0001$) (Figure 2d). We reviewed the individual patients misclassified by the full-feature syndromic model (Figure 2, e and f). Patients who were wrongly classified by the model seemed to rely strongly on PC6, the abdominal pain component. Individuals who were wrongly classified by the model as FC (Figure 1e; blue dots), but actually had a clinician-provided diagnosis of IBS-C, seemed to have degrees of pain which overlapped that of the accurately classified patients with FC (Figure 1e; orange dots). Similarly, those who were wrongly classified by the model as having IBS-C (Figure 1e; pink dots), but actually had a clinician-provided diagnosis of FC, seemed to have degrees of pain which overlapped the accurately classified patients with IBS-C (Figure 1e; green dots). This pattern

FUNCTIONAL GI DISORDERS
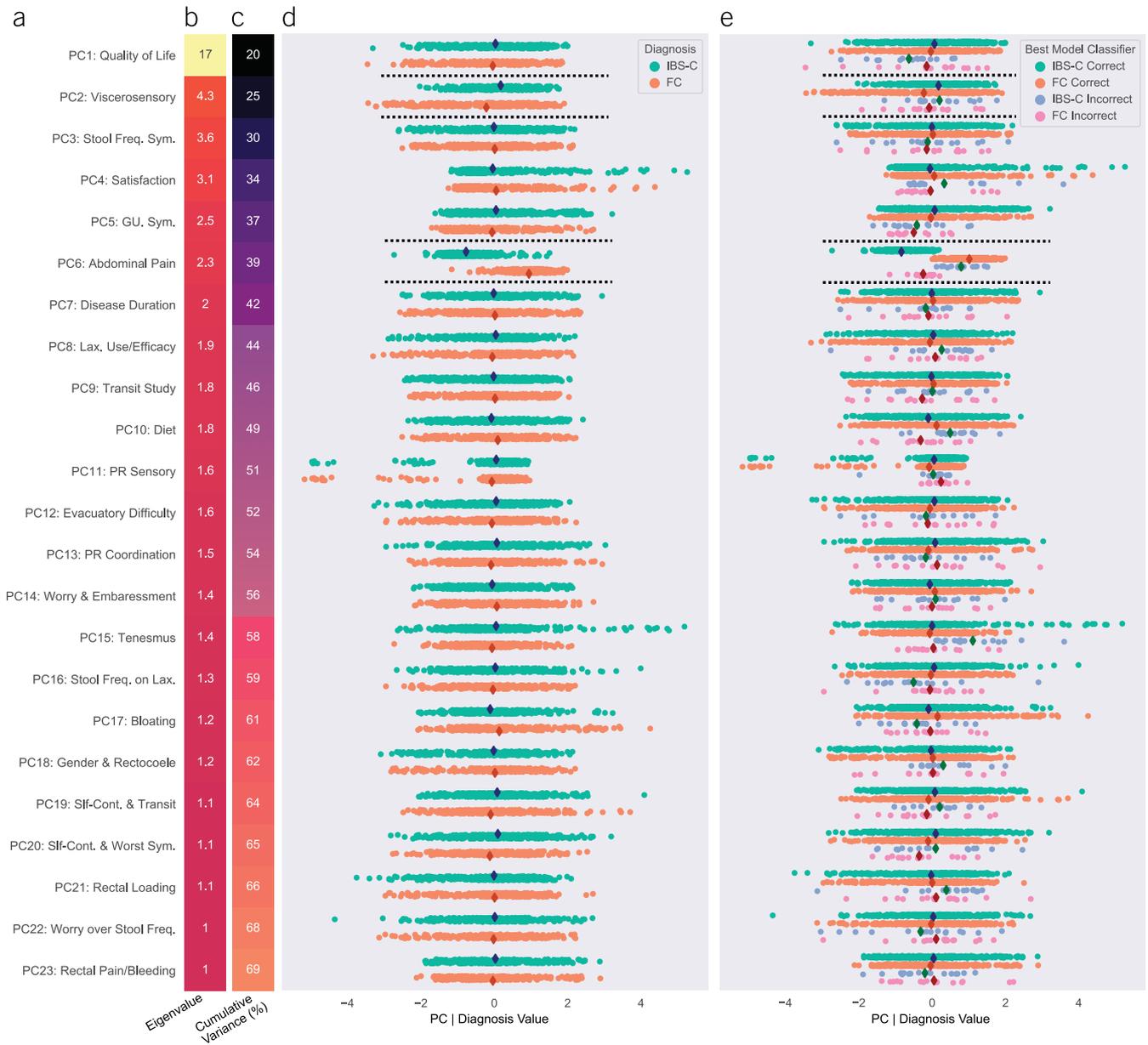
Figure 1. All disease components of chronic constipation are equivalent between IBS-C and FC, except for abdominal pain and viscerosensory measures. (**a**) PCA identified 23 meaningful components listed in descending order of eigenvalue (**b**), which cumulatively accounted for 69% of the total variance (**c**). (**d**) PCs were compared between IBS-C and FC groups, which revealed the groups to be largely similar, with the exception of PC2: viscerosensory and PC6: abdominal pain (both demarcated with dotted lines). (**e**) Per-component results of the best-performing machine learning classifier in diagnosing IBS-C or FC (Figure 2). Patients who are wrongly diagnosed by the model seem to hinge exclusively on PC6: abdominal pain, wherein if a patient was wrongly diagnosed as IBS-C (i.e., actually having FC, pink dots), their PC6 pain value was closer to the IBS-C group (green dots) and *vice versa*. FC, functional constipation; Freq., frequency; GU, genitourinary; IBS-C, irritable bowel syndrome with constipation; Lax, laxative; PC, principal component; PCA, principal component analysis; PR, per rectum; Slf-Cont, self-control; sym, symptoms.

was not apparent for any of the remaining PCs, wherein there was considerable overlap.

Pain is only discriminative for IBS-C over FC if it directly features in diagnostic criteria

Having ascertained PC6: abdominal pain as the key feature in whether a model would diagnose IBS-C accurately or inaccurately (Figure 2), we investigated this specific component and its individual constituents in greater detail. We identified that by only using the pain measures in PC6 that do not directly

factor into the diagnostic criteria for IBS, the model then becomes unable to separate the patient groups accurately; the accuracy of the best performing model was 53% (see Supplementary Material, Supplementary Digital Content 3, http://links.lww.com/AJG/B619 and Supplementary Figure 4, Supplementary Digital Content 8, http://links.lww.com/AJG/B624). Therefore, in our sample, the only factors which allow IBS-C and FC to be distinguished by machine learning were abdominal pain measures that form part of the diagnostic criteria.
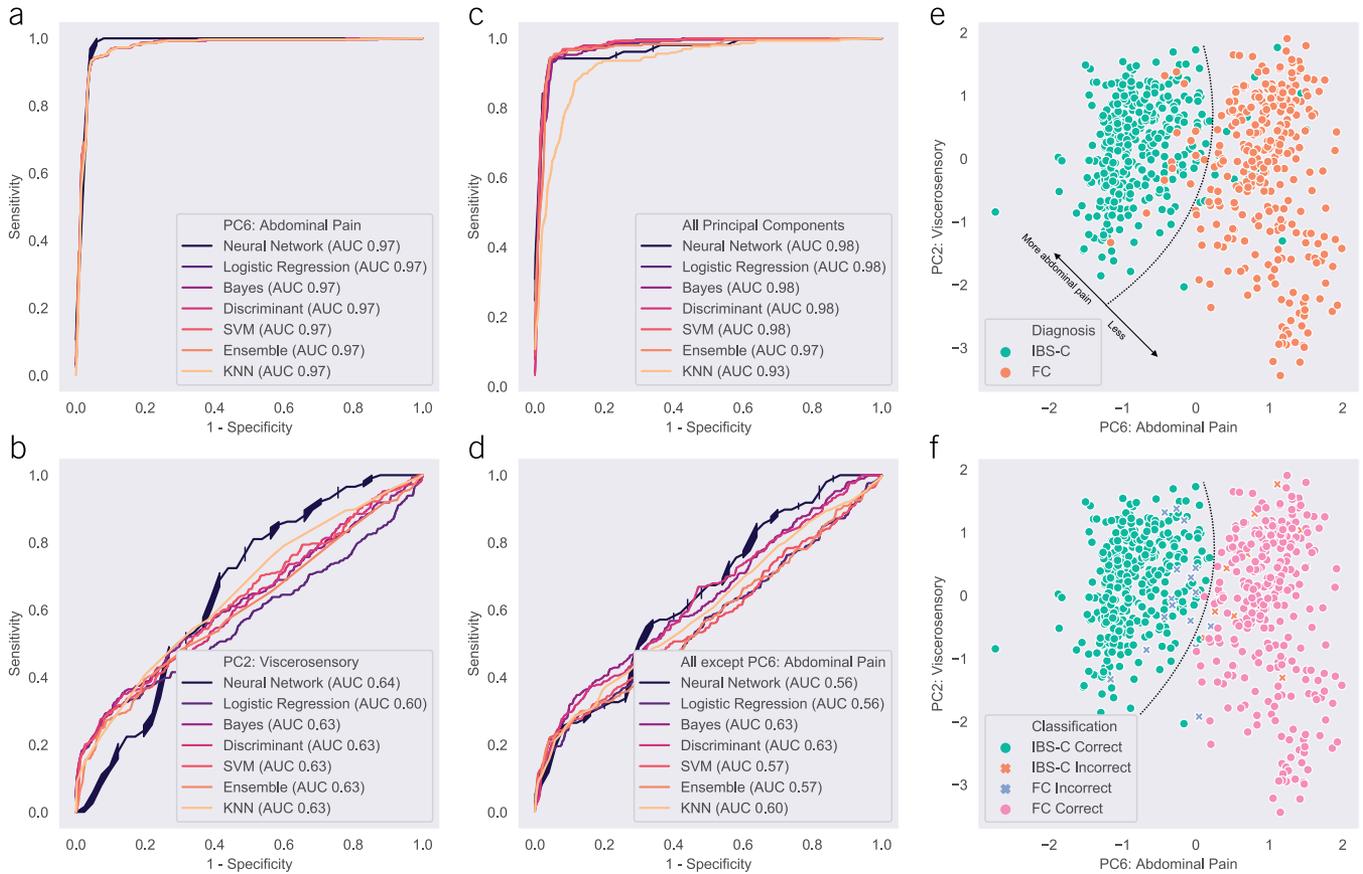
Figure 2. Abdominal pain alone is necessary and sufficient for differentiating IBS-C and FC. (**a**) Unisymptomatic models of abdominal pain (PC6) achieved near-perfect accuracy in distinguishing IBS-C and FC. (**b**) Models of viscerosensory measures (PC2) perform poorly. (**c**) There is no significant improvement in model accuracy when a syndromic feature set is used in place of a unisymptomatic pain feature. (**d**) Syndromic feature set models, when excluding PC6: abdominal pain, shows chance-level accuracy. (**e**) Two-dimensional plot of the 2 components which significantly differed between patients with IBS-C and FC, PC2: viscerosensory and PC6: abdominal pain, with patients with IBS-C having both worse pain and other viscerosensory measures. The 2 diagnoses arguably seem distinct with these data alone, illustrated with dotted line approximately separating the groups. (**f**) Two-dimensional plot of the 2 components, with the results of the best-performing classifier plotted. Namely, patients wrongly classified as either diagnosis seem to fall on the "wrong side" of the diagnostic line, not conforming to the stereotype of the diagnosis of FC or IBS-C (i.e., patients with FC having abdominal pain features expected in IBS-C, and patients with IBS-C having abdominal pain features expected in FC). FC, functional constipation; IBS-C, irritable bowel syndrome with constipation; PC, principal component.

Unsupervised machine learning clusters patients into syndromes by disease impact

We sought to determine if syndromic segregation of chronic constipation patients was apparent using unsupervised learning to do so in a data-driven way. Two-step cluster analysis identified the best silhouette fit of possible outcomes was the parcellation of patients into 2 disease clusters. Cluster 1 contained 421 (63.7%) patients, with the remaining 240 patients allocated to cluster 2 (36.3%). The feature weights which determined allocation to cluster 1 or 2 were reviewed. These identified that disease impact and QOL domains held the highest importance, including adverse effect on dietary habit, worry, and anxiety over stool frequency, viscerosensory measures, embarrassment, and QOL itself (Figure 3, a and b). A diagnostic label of IBS-C or FC was weakly predictive only, which had equivalent predictive power to PC6: abdominal pain, an expected outcome given that the 2 disorders are seemingly differentiated by this measure alone. By using the Fisher exact test, a diagnosis of IBS-C over FC was weakly significant in allocation to cluster (corrected $P <$ 0.03). We additionally compared cluster allocation, aligned to diagnostic label and individual PC features, which showed

considerable overlap between the diagnoses in all domains except for abdominal pain, where the 2 separated, as would be expected (Figure 3c). We cross-compared the findings of our 2-step cluster analysis, which used all 23 PCs, to an independent analysis of all raw data using UMAP (26). This confirmed these findings, namely that 2 clusters exist that hinge on the impact of disease on daily life, which held negligible relationship to diagnostic label (Figure 3, d–e).

We examined the cluster profiles for patients using aforementioned feature (node) importance and feature correlations (edges). Stochastic block modeling was used to identify communities of similar patients. This was exponentially weighted by feature importance and feature-feature correlation, which generated undirected weighted networks of how constipation features interact as a network (Figure 4). These enabled us to review the complex network-like interaction between many features in a syndromic fashion. For instance, the cluster 1 node of dietary disturbance strongly connected to greater worry and embarrassment that strongly related to worse overall QOL, which in itself was also related directly to dietary disturbance. For example, in cluster 2, less dietary disturbance was strongly linked to less

FUNCTIONAL GI DISORDERS
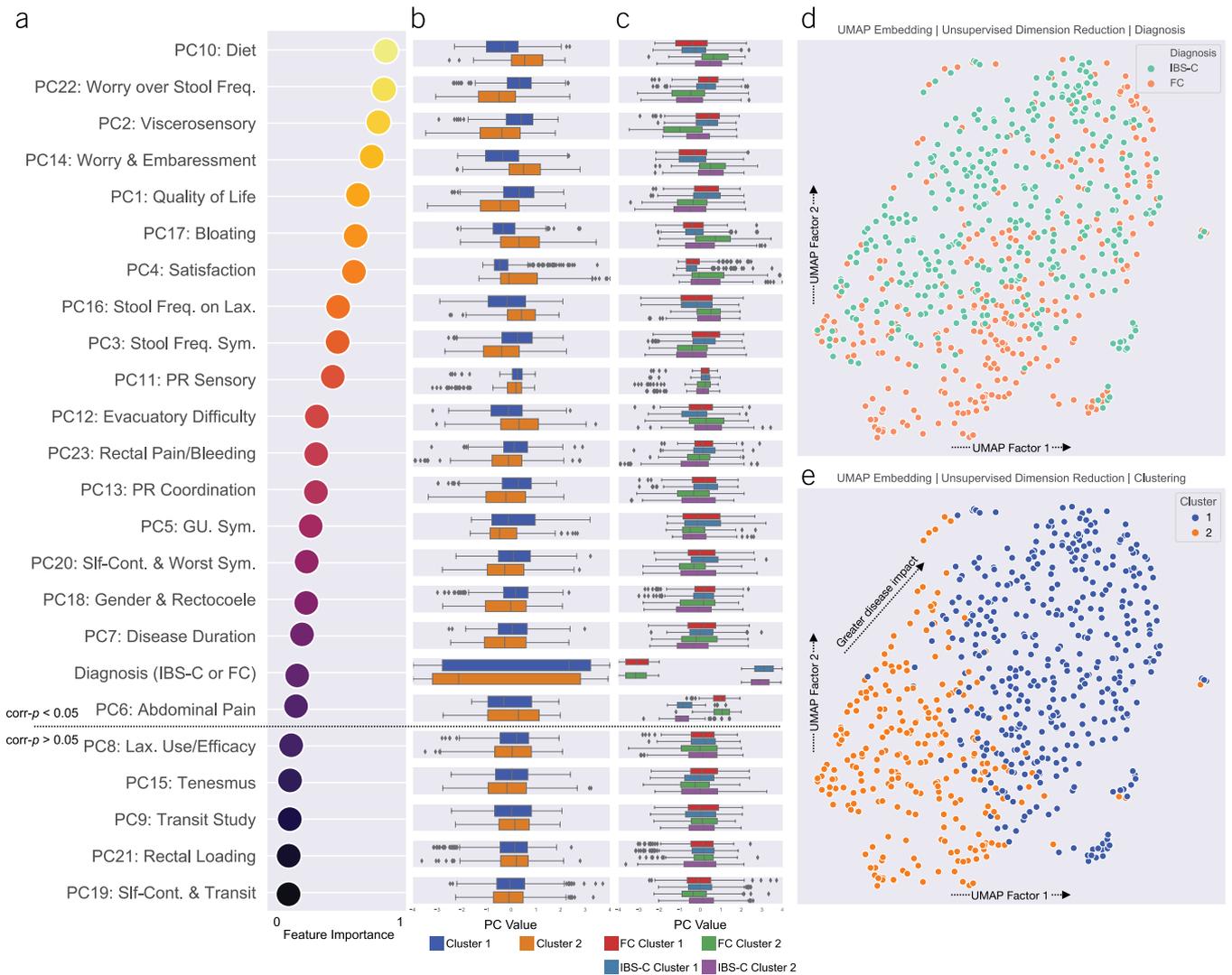
Figure 3. Unsupervised learning clusters patients with chronic constipation patients into syndromes by disease impact. (**a**) Feature importance, by PC, in allocation to cluster 1 or cluster 2, where values tending to 1 are of higher importance. (**b**) Box and whisker plots illustrating the differences between clustered groups. (**c**) Box and whisker plots illustrating the differences between diagnostic labels and cluster groups show that diagnoses only differ with abdominal pain. (**d**) UMAP embeddings of all raw features illustrate no convincing visual distinction between IBS-C and FC. (**e**) Rather, UMAP embeddings align well to our independent cluster derivation. FC, functional constipation; Freq., frequency; GU, genitourinary; IBS-C, irritable bowel syndrome with constipation; Lax, laxative; PC, principal component; PCA, principal component analysis; PR, per rectum; Slf-Cont, self-control; sym, symptoms; UMAP, uniform manifold approximation and projection.

bloating that was linked to less viscerosensory symptoms, which was strongly linked to less abdominal pain.

To formally test the validity of these newfound clusters as syndromic segregations, we trained classification models with sequential feature addition and statistically compared the performance (Figure 5). First, a unisymptomatic (single feature) model using PC10: diet (statistically the most predictive feature, see Figure 3) achieved a best possible accuracy with a logistic regression (AUC 0.74) (Figure 5a). We next retrained the classifier using the 5 most predictive features (Figure 3a), the best model being a linear support vector machine (SVM) (AUC 0.87), performing significantly better than the single feature model ($P < 0.0001$) (Figure 5b). This performance was further improved when using the 10 most predictive features, the best model being a coarse Gaussian SVM (AUC 0.94),

significantly more accurate than the 5-feature model ($P < 0.0003$) (Figure 5c). We finally evaluated model performance using all PCs which, after multiple comparison, significantly differed between cluster 1 and 2 (Figure 3a), an 18 feature model. This fully syndromic model performed well, the best model being a quadratic SVM (AUC 0.99), significantly more accurate than the previous 10-feature model ($P < 0.0001$) (Figure 5d).

DISCUSSION

A valid clinical classification is either based on a known pathogenesis or clear differences in diagnostic (clinical or investigatory) criteria. A definition of a syndrome as a unique disease entity is one that differs from another disorder according to multiple features. A valuable classification is one that defines treatment

Figure 4. Network-based cluster profiling in chronic constipation hinges on disease impact. Disease networks of cluster 1 (**a**) and 2 (**b**). Node size and color indicate importance of a particular node in the profile, with edge color and width corresponding to correlation strength between nodes (key). Abdo., abdominal; Dis, disease; Diff., difficulty; Disturb., disturbance; Freq, frequency; Impair., impairment; Embar., embarrassment; F, female; GU, genitourinary; Lax., laxative; o/, over; PR, per rectal; QOL, quality of life; Rc., rectocele; Sens., sensory; SC., self-confidence; sym, symptoms; Ts, transit study speed; Tx, treatment; w/, with.

responses. The classification of constipation into FC and IBS-C is necessarily an arbitrary one because the pathogenesis of these conditions remains unclear. One of the difficulties with this classification is the need for an exclusion clause because many patients fulfil both FC and IBS-C criteria, in which case the IBS-C diagnosis takes priority. This explains much of the overlap, but the question still remains as to whether the classification identifies any other meaningful differences.

The initial dichotomous use of the classification has been tempered in recent years with the Rome IV criteria, suggesting that they represent a spectrum (27), one seemingly focused on

abdominal pain, which our data support. The classification has also been shown to be temporally unstable, with approximately one-third of patients across either disease label switching to the other label 12 months later (8). However, other researchers claim important differences between the subtypes (28,29), and a recent study proposed subtype differences that hinge on viscerosensory, extraintestinal, somatic, and psychological variables (28), a hypothesis that our clustering findings also support.

We demonstrate that chronic constipation, as a disease entity, comprises a number of features. These components are multifaceted and vary in nature, from QOL domains, stool frequency measures,

Figure 5. Clustering of patients by disease impact is a valid syndromic classification. ROC plots of model performance in distinguishing patients from cluster 1 vs cluster 2. (**a**) Single feature (unisymptomatic) model ROCs. (**b**) 5-feature model ROCs. (**c**) 10-feature model ROCs. (**d**) 18 feature (fully syndromic) model ROCs, which have best predictive accuracy. ROC, receiver operator characteristic.

symptom measures, viscerosensory (including pain) measures, clinical examination, and investigatory findings. Using a machine learning approach to investigate whether FC and IBS-C represent distinct subgroups, we show that of the dimensions that seem to make up chronic constipation, only abdominal pain and viscerosensory measures differ significantly between these putatively distinct disorders. Moreover, using machine learning, a model is able to differentiate accurately between the 2 disorders only when abdominal pain data that directly factors into current accepted diagnostic criteria are supplied; without these data, the model performs at a level of chance. Other measures of pain that do not feature in the diagnostic classification for IBS show no meaningful predictive value during machine learning either. We show that patients with symptoms of chronic constipation exhibit syndromic clustering, but this is defined most importantly by the impact of the disease on their life, not by the diagnostic label these patients are given. It would be prudent to evaluate these findings in other geographical regions and nonspecialist centers to ensure generalizability. Of limitation in this study, there are some quantitative physiological measures we did not ascertain, such as gut microbiota or colonic MRI, although it was felt that the inclusion of methods not routinely used in current clinical practice were not as justifiable and would limit generalizability outside of a tertiary center. Future studies should establish the additive value insofar as predicting one diagnosis over the other or whether this would influence our newfound clustering pattern.

Clinically, the treatment of the 2 subgroups is often highly similar, with drugs that have laxative effects. This is with the caveat that by virtue of these diagnostic criteria, an evidence base for different treatment options exists because a literature of randomized controlled trials for "people with IBS," i.e., those with more abdominal pain, and to a lesser extent, "people with FC," i.e., those with less abdominal pain, has been developed. We argue that the presumption the 2 are distinct disorders can lead to a situation where it is unclear whether some treatments are helpful or not see Table 2 of ref. 7. An example of this in the United Kingdom is the National Institute for Health and Care Excellence licensing of linaclotide for IBS-C (30), whereas lubiprostone is licensed for chronic idiopathic constipation (31), despite the literature illustrating efficacy of linaclotide for both conditions (32). This is especially the case for treatments where effectiveness has been ascertained principally by stool frequency measures, such as spontaneous bowel movements, as opposed to a specific focus on pain. Given that we fail to show differences in these stool frequency or transit measures between patients with FC and IBS-C, we therefore question the appropriateness of their segregation. Of course, treatments specific to attenuating visceral pain may differ, but it could be construed the reason for this is because the 2 disorders are segregated by the presence of visceral pain anyway, arguably a self-fulfilling difference.

Although we show that the diagnoses of IBS-C and FC are essentially inseparable in our data set, other than for the presence of pain, we identify a clustering pattern of patients whose chronic constipation, in general, has a greater impact on their everyday life. These patients, belonging to cluster 1 exhibited markedly greater dietary disturbance, greater worry and embarrassment regarding their stool frequency, and greater viscerosensory disturbance, as well as numerous other measures, with the converse profiling being true for patients in cluster 2. Furthermore, these domains, although individual in

their own right (such as rectal pain), all inter-relate to provide a highly complex and organized disease network, such as rectal pain connecting to tenesmus, which in turn connects to a lower treatment satisfaction *and so on*. We validate that this network of features is in keeping with syndromic segregation, wherein predictive accuracy of models improves significantly with the incremental addition of features (therefore favoring the syndromic definition of disease as opposed to solely unisymptomatic difference). Future studies should seek pathophysiological differences and investigate best therapeutic algorithms related to these clusters.

In summary, we report that in a large number of patients with chronic constipation labeled with either IBS-C or FC, the 2 groups show negligible differences across all demographic data, symptomatology, examination, and transit studies. We show that, using machine learning, algorithms can accurately diagnose IBS-C over FC only when abdominal pain measures directly factored into the Rome III criteria are available. Machine learning supports the 2 disorders to be unisymptomatically different, i.e. disorders lying on a one-dimensional spectrum, rather than disparate syndromes. Treatments currently considered for either IBS-C or FC may be equally effective for the alternate diagnostic label, especially where the primary aim is to increase spontaneous bowel movements, which needs future evaluation in clinical trials. Furthermore, we illustrate with machine learning and graphical networks that patients with chronic constipation do separate into 2 clusters, but these clusters are a syndromic network of inter-relating disease factors that impact on a patient's life, as opposed to diagnostic label. Further research is required to evaluate the value of this diagnostic classification system, whether pathophysiological differences are apparent and whether effective treatments differ between these newly identified syndromic clusters. In particular, treatments that moderate abnormal sensation should be sought, not just those that increase transit and stool production. The outcome measures for clinical trials may need modification. Personalization of treatment regimens for chronic constipation based on these novel, network-based, syndromic clusters may be beneficial.

Study Highlights

WHAT IS KNOWN

✓ Chronic constipation is presently classified into 2 main syndromes—IBS-C and FC.

✓ The segregation of these as separate syndromes assumes that they differ along multiple clinical characteristics and are plausibly of distinct pathophysiology.

✓ Previous studies have questioned the mutual exclusivity of IBS-C and FC as disease entities, suggesting instead that the 2 lie on a single spectrum.

WHAT IS NEW HERE

✓ Using machine learning, IBS-C and FC differ only by a single feature, abdominal pain, a self-fulfillling difference given that abdominal pain inherently distinguishes the 2 in current diagnostic criteria. These disorders therefore do not differ *syndromically*, but *unisymptomatically.*

✓ In chronic constipation, 2 clusters of patients *do exist* but bear negligible resemblance to the diagnostic labels of IBS-C and FC. Rather, these clusters of patients are separated by a complex network of features, including diet disturbance, viscerosensory disturbance and QOL.

✓ This has significant implications for both diagnostic classification systems and treatment algorithms. We propose an alternative, data-driven and validated as syndromic, segregation of patients with chronic constipation.

REFERENCES

1. Suares NC, Ford AC. Prevalence of, and risk factors for, chronic idiopathic constipation in the community: Systematic review and meta-analysis. Am J Gastroenterol 2011;106:1582–91; quiz 1581–92.
2. Drossman DA. The functional gastrointestinal disorders and the Rome III process. Gastroenterology 2006;130:1377–90.
3. Drossman DA, Hasler WL. Rome IV-functional GI disorders: Disorders of gut-brain interaction. Gastroenterology 2016;150:1257–61.
4. Schmulson MJ, Drossman DA. What is new in Rome IV. J Neurogastroenterol Motil 2017;23:151–63.
5. Longstreth GF, Thompson WG, Chey WD, et al. Functional bowel disorders. Gastroenterology 2006;130:1480–91.
6. Merriam-Webster. Dictionary Definition: Syndrome (https://www.merriam-webster.com/dictionary/syndrome). Accessed Septemebr 14, 2019.
7. Siah KT, Wong RK, Whitehead WE. Chronic constipation and constipation-predominant IBS: Separate and distinct disorders or a spectrum of disease? Gastroenterol Hepatol (N Y) 2016;12:171–8.
8. Wong RK, Palsson OS, Turner MJ, et al. Inability of the Rome III criteria to distinguish functional constipation from constipation-subtype irritable bowel syndrome. Am J Gastroenterol 2010;105:2228–34.
9. Shekhar C, Monaghan PJ, Morris J, et al. Rome III functional constipation and irritable bowel syndrome with constipation are similar disorders within a spectrum of sensitization, regulated by serotonin. Gastroenterology 2013;145:749–57;quiz e13–4.
10. Zhao YF, Ma XQ, Wang R, et al. Epidemiology of functional constipation and comparison with constipation-predominant irritable bowel syndrome: The systematic investigation of gastrointestinal diseases in China (SILC). Aliment Pharmacol Ther 2011;34:1020–9.
11. Heidelbaugh JJ, Stelwagon M, Miller SA, et al. The spectrum of constipation-predominant irritable bowel syndrome and chronic idiopathic constipation: US survey assessing symptoms, care seeking, and disease burden. Am J Gastroenterol 2015;110:580–7.
12. Rey E, Balboa A, Mearin F. Chronic constipation, irritable bowel syndrome with constipation and constipation with pain/discomfort: Similarities and differences. Am J Gastroenterol 2014;109:876–84.
13. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology: promises and pitfalls. Am J Gastroenterol 2019;114:422–8.
14. Topol E. The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future. Health Education England, NHS: London, 2019.
15. Goodfellow I, Bengio Y, Courville A. Deep Learning: MIT Press: Cambridge, 2017.
16. Ruffle JK, Farmer AD, Aziz Q. Chapter 33: Artificial intelligence in gastroenterology. In: Faintuch J, Faintuch S (eds). Precision Medicine for Investigators, Practitioners and Providers: Academic Press-Elsevier: Cambridge, 2020, pp 343–50.
17. Ruffle JK, Patel A, Giampietro V, et al. Functional brain networks and neuroanatomy underpinning nausea severity can predict nausea susceptibility using machine learning. J Physiol 2019;597:1517–1529.
18. Agachan F, Chen T, Pfeifer J, et al. A constipation scoring system to simplify evaluation and management of constipated patients. Dis Colon Rectum 1996;39:681–5.
19. Marquis P, De La Loge C, Dubois D, et al. Development and validation of the patient assessment of constipation quality of life questionnaire. Scand J Gastroenterol 2005;40:540–51.
20. Frank L, Kleinman L, Farup C, et al. Psychometric validation of a constipation symptom assessment questionnaire. Scand J Gastroenterol 1999;34:870–7.
21. Tack J, Muller-Lissner S, Stanghellini V, et al. Diagnosis and treatment of chronic constipation—A European perspective. Neurogastroenterol Motil 2011;23:697–710.
22. Metcalf AM, Phillips SF, Zinsmeister AR, et al. Simplified assessment of segmental colonic transit. Gastroenterology 1987;92:40–7.
23. Benjamini Y, Yekutieli D. False discovery rate–adjusted multiple confidence intervals for selected parameters. J Am Stat Assoc 2005;100:71–81.
24. McDonald JH. Handbook of Biological Statistics. 3rd edn. Sparky House Publishing: Baltimore, MD; 2014.
25. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653–60.
26. McInees L, Healy J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv e-prints 2018;1802–03426 (https://arxiv.org/abs/1802.03426).
27. Mearin F, Lacy BE, Chang L, et al. Bowel disorders. Gastroenterology 2016;150:P1393–1407.E5.
28. Polster A, Van Oudenhove L, Jones M, et al. Mixture model analysis identifies irritable bowel syndrome subgroups characterised by specific profiles of gastrointestinal, extraintestinal somatic and psychological symptoms. Aliment Pharmacol Ther 2017;46:529–39.
29. Lam C, Chaddock G, Marciani L, et al. Colonic response to laxative ingestion as assessed by MRI differs in constipated irritable bowel syndrome compared to functional constipation. Neurogastroenterol Motil 2016;28:861–70.
30. NICE. Linaclotide—National Institute for Health and Care Excellence, British National Formulary. BNF, British National Formulary. (https://bnf.nice.org.uk/drug/linaclotide.html). Accessed October 22, 2019.
31. NICE. Lubiprostone—National Institute for Health and Care Excellence, British National Formulary. BNF, British National Formulary. (https://www.nice.org.uk/guidance/ta318/documents/constipation-chronic-idiopathic-lubiprostone-final-appraisal-determination-document2). Accessed October 22, 2019.
32. Lembo AJ, Kurtz CB, MacDougall JE, et al. Efficacy of linaclotide for patients with chronic constipation. Gastroenterology 2010;138:886–95.e1.