



Published in final edited form as:

*Inflamm Bowel Dis.* 2015 October ; 21(10): 2262–2271. doi:10.1097/MIB.0000000000000496.

## Measurement of Inflammatory Bowel Disease Symptoms: Reliability of an Abbreviated Approach to Data Collection

Maria Paula Henao, MD<sup>1</sup>, Meenakshi Bewtra, MD, MPH, PhD<sup>1,2,3</sup>, Mark T Osterman, MD, MSCE<sup>1</sup>, Faten N Aberra, MD, MSCE<sup>1</sup>, Frank I Scott, MD, MSCE<sup>1,3</sup>, Gary R Lichtenstein, MD<sup>1</sup>, Jennifer Kraschnewski, MD, MPH<sup>4</sup>, and James D Lewis, MD, MSCE<sup>1,2,3</sup>

<sup>1</sup>Department of Medicine, Perelman School of Medicine at the University of Pennsylvania

<sup>2</sup>Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania

<sup>3</sup>Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania

<sup>4</sup>Departments of Medicine and Public Health Sciences, Pennsylvania State College of Medicine, Hershey, Pennsylvania

### Abstract

**Background**—The Crohn’s Disease Activity Index (CDAI) and Mayo score for ulcerative colitis (UC) require symptom recall and/or use of a symptom diary. We examined patients’ abilities to recall their symptoms and the day-to-day variability of symptoms.

**Methods**—Patients with UC or CD completed a questionnaire including items from the short CDAI (sCDAI) and the 6-point Mayo score. Patients were randomized to receive a follow-up questionnaire testing recall of the bowel symptom items between 1 to 7 days later. In a second study, patients completed a 7-day electronic diary recording their symptoms. sCDAI and 6-point Mayo scores were computed. Analyses estimated daily variability in the indices and misclassification rates when using fewer than 7 days of data.

**Results**—100%, 82%, and 90% of CD participants recalled the same disease activity status (i.e. active vs. remission) as reported on the initial survey when the follow-up questionnaire was administered 1–2, 3–5, and 6–8 days later, respectively. Compared to using 7 days of data, when using only day 7 data 3.7% of CD patients were misclassified as active or inactive. Disease activity was misclassified in 2.8%, 4.9%, and 3.3% of patients by using the last 2, 3, or 4 days, respectively. Results were similar for UC patients.

**Conclusions**—Patients with CD and UC demonstrated good recall of bowel symptoms for up to 8 days. Additionally, bowel symptoms have relatively little variability within a 7-day period allowing for accurate computation of the sCDAI and 6-point Mayo score using 1–3 days of data.

---

Corresponding author: James D Lewis, MD, MSCE, 720 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, lewisjd@upenn.edu.

Conflicts of Interest

No other authors report any potential conflicts of interest in relation to the content of this manuscript.

## Keywords

disease activity measurement; outcomes research/measurement; clinical trials; epidemiology

---

## INTRODUCTION

For decades, the gold standard to measure disease activity for Crohn's disease (CD) research has been the Crohn's Disease Activity Index (CDAI).<sup>1-3</sup> The CDAI is computed using laboratory data, physical exam findings, and self-reported CD symptoms for each of the prior 7 days.<sup>2</sup> Investigators recently validated the short CDAI (sCDAI) as an alternative method to identify symptoms and measure CD activity using only a questionnaire, which can be completed without an office visit or laboratory work.<sup>4</sup>

The most commonly used index for studies of the efficacy of ulcerative colitis (UC) therapy is the 12-point Mayo score. This composite index includes patient-reported estimates of stool frequency and bleeding, endoscopic assessment of mucosal inflammation, and the physician's global assessment of disease activity. We have previously demonstrated that the 6-point Mayo score accurately estimates disease activity using only the two components of patient-reported symptom activity (bowel frequency and bleeding).<sup>5,6</sup>

Although the CDAI and the Mayo score have been used for decades, little research has focused on the best ways to measure the patient-reported components of the indices, particularly in terms of day-to-day variability and accuracy of patient recall. Prior studies have suggested that patients often have poor adherence when using a 7-day paper diary to measure the CDAI, which can result in recall bias.<sup>7,8</sup> Likewise, there is no formal guidance on the optimal number of days to collect data on UC symptoms when computing the Mayo score. In both cases, the number of days for data collection depends on the day-to-day variability of symptoms for CD or UC, which has not been well studied. To the extent that there is limited day-to-day variability, shorter periods of data collection should be nearly as accurate as seven full days of data.

Therefore, in this study, we examined two important features of symptom recording for IBD indices: patients' abilities to recall their symptoms and the day-to-day variability of their symptoms. The former can be used to determine the maximum acceptable time window between experiencing bowel symptoms and recording these symptoms in the diary while the latter can be used to determine the minimum number of days required to collect IBD-related symptom data to accurately capture patients' disease status.

## MATERIALS AND METHODS

We conducted two prospective studies to evaluate recall and day-to-day variability of IBD symptoms. To evaluate recall of bowel symptom data, we administered a baseline and follow-up questionnaire to a convenience sample of 100 IBD patients presenting for an office visit at the Gastroenterology clinic at the University of Pennsylvania. The baseline questionnaire, completed on paper during patients' office visits and subsequently uploaded to a REDCap electronic database, included questions from the sCDAI and the Mayo score as

well as a variety of quality of life questions derived from work by the Crohn's and Colitis Foundation of America's Quality of Care task force (Supplemental Figure 1).<sup>13</sup> The latter were used as distractors to limit priming of the sCDAI and Mayo score questions for the follow-up questionnaire. Participants were asked to report their symptoms from the prior day, and were then randomized to receive the follow-up questionnaire via email on one of the subsequent seven days. The follow-up questionnaire, completed over the Internet using REDCap electronic data capture tools, consisted of an abbreviated version of the original questionnaire including only the items pertaining to the sCDAI and Mayo score, as well as one question regarding how confident participants felt about their responses. In both the baseline and the follow-up questionnaires, the participants were asked to report their symptoms on the day prior to the office visit.

To quantify day-to-day variability of bowel symptoms, we conducted a cross-sectional study of 94 patients with IBD. All patients were asked to complete a 7-day electronic diary recording their stool frequency, abdominal pain, and general wellbeing. The diary was completed over the Internet using REDCap electronic data capture tools. Patients received a daily email that provided a link to the diary. We used these data to assess the range of variability of symptoms and the relative accuracy of measuring disease activity with fewer than 7 days of data, using the full 7 days as the gold standard.

### Exclusion Criteria

For both studies we excluded participants with a history of total or sub-total colectomy, ileostomy, or colostomy given that non-invasive disease activity measures do not accurately assess disease activity in this population. Patients who did not have access to the Internet or a working email address were likewise excluded.

### sCDAI and 6-point Mayo Calculations

The sCDAI uses the same scale as the full CDAI, such that scores <150 define remission, 150–219 mild activity, 220–450 moderate activity, and >450 severe activity. Computation is straightforward:

$$sCDAI = 44 + \left( 2 * \sum_{n=1}^7 L \right) + \left( 5 * \sum_{n=1}^7 A \right) + \left( 7 * \sum_{n=1}^7 W \right)$$

where  $L$  is the number of liquid stools,  $A$  is the rating of abdominal pain (0–3, none to severe),  $W$  is the rating of general wellbeing (0–4, generally well to terrible), and  $n$  is the day of follow-up. In study 1, we estimated sCDAI scores from only 1 day of data by multiplying the individual component scores by 7. Although these estimate sCDAI scores were potentially less accurate than sCDAI scores from a full week of data, the objective was to assess the accuracy of recall between the baseline and follow-up questionnaires, and not the value of the scores themselves. In study 2, where we examined day-to-day variability and the correlation of shorter periods of data collection, participants who did not complete the full seven days of questionnaires were included in the study as long as they had completed a minimum of four days of questionnaires within a maximum of eight days from

the initial questionnaire. The sCDAI for participants who did not complete all seven days was adjusted accordingly by averaging the available data for the individual components across seven days. Similarly, when computing the projected sCDAI scores for fewer than 7 days (e.g., the last two days of data, the last three days of data, etc.), we summed the individual component scores from the corresponding days, multiplied the sum of the individual components by 7, and divided by the number of total days of data we used ( $d$ ) to adjust for the missing days:

$$sCDAI = 44 + \frac{7}{d} * \left( \left( 2 * \sum_{n=1}^d L \right) + \left( 5 * \sum_{n=1}^d A \right) + \left( 7 * \sum_{n=1}^d W \right) \right)$$

The 6-point Mayo score uses only the number of bowel movements above average and the bleeding components of the full Mayo score. Scores above 1.5 are indicative of active disease. 6-point Mayo scores were computed as previously described.<sup>5</sup> For the day-to-day variability study, the Mayo scores were calculated using the data available for each day and combined day scores (last two days, last three days, etc.) were averaged using the available data from the corresponding days.

### Statistical Analyses

All analyses were conducted separately for patients with CD and UC such that sCDAI scores were only calculated for CD patients and Mayo scores were only calculated for UC patients. In study 1, we calculated the mean and standard deviation of the difference in reported symptoms on the questionnaire completed in the office compared to the reported symptoms of the same day on the follow-up questionnaire 1–8 days later. We compared the difference in scores according to the time interval between completing the questionnaires using ANOVA. We used ANCOVA to determine whether the accuracy differs according to duration between completion of the two questionnaires before and after adjusting for the following characteristics: age; sex; and self-reported disease activity derived from a single question as remission, minimal, mild, moderate, or severely active.

In study 2, we computed the sCDAI using all 7 days of data collection and assessed the degree of variation of individual component scores for stool frequency, abdominal pain, and wellbeing as well as the sCDAI within a 1-week period by computing the difference between the minimum daily score and the maximum daily score for each patient (the weekly maximum delta). These analyses were done across the entire cohort and then stratified by whether the patient had active or quiescent disease as determined by their sCDAI or 6-point Mayo scores.

We next examined the correlation of the last day, last 2 days, last 3 days, and last 4 days with the full 7 days of data for the sCDAI, 6-point Mayo score, and individual component scores for stool frequency, abdominal pain, and wellbeing using Spearman's correlation coefficients. Patients with missing data for any of the components on the last 1–4 days of data collection were excluded from the respective analysis (i.e. pairwise deletion). Sensitivity analysis dropping any patient missing any of the component scores on any of the last 4 days (i.e. casewise deletion) produced similar results (data not shown). Additionally,

to consider the possibility that IBD symptoms do not wax and wane solely in 24-hour periods, we tested the correlation of non-consecutive days of data (days 1 and 5; days 2 and 6; days 2, 4, and 6; and days 1, 3, 5, and 7) with the full week of data. Using the full 7 days of data as the gold standard, we assessed the proportion of patients who were misclassified as active or inactive, respectively, when using different combinations of fewer than 7 days of data. This analysis was repeated categorizing the disease activity as remission, mild, moderate, and severe.

All study data were collected and managed using REDCap electronic data capture tools hosted at the University of Pennsylvania.<sup>14</sup> Statistical analyses were conducted using STATA v13.1. The study was approved by the University of Pennsylvania's Institutional Review Board. All participants provided informed consent to participate in the study. Participants were provided with a nominal compensation valued at \$5 or less for participating in the study.

## RESULTS

### Bowel Symptom Recall

Study 1 included 100 patients, of whom 50% were female and 72% had CD (Table 1). Among patients with CD, 28% were in remission at baseline when using their calculated sCDAI score for their initial visit, while 36% had mild disease activity and the remaining 36% had moderate activity. No patient with CD was classified as severe at baseline. Of the 28 patients with UC, 61% had inactive disease according to their calculated Mayo scores from their initial visit. At the time of the follow-up survey, 93% of participants reported that they were confident or very confident and 7% somewhat confident of their answers. No participant felt somewhat unsure or completely unsure of their answers. This did not differ by time of recall ( $p=0.55$ ).

Table 2 contains the mean absolute values of the difference in the initial and follow-up surveys of the sCDAI and 6-point Mayo scores and the subcomponents. Comparison of recall based on the number of days from the initial questionnaire did not identify significant differences for any of the measures, whether using all patients or only those with active disease.

Disease activity based on the sCDAI measured at the time of the office visit matched disease activity on the follow-up questionnaire in 64 (89%) of the CD patients (Figure 1A). The percent of participants with CD whose category of disease severity (remission, mild, moderate, or severe) remained identical between baseline and follow-up was 100%, 70%, and 75% for days 1–2, 3–5, and 6–8, respectively. Supplemental Table 1 provides additional detail on the degree of misclassification. Only 3% were misclassified by more than one level of disease severity. Likewise, comparing active disease versus remission of CD patients, 100%, 82%, and 90% of participants with follow-up surveys 1–2, 3–5, and 6–8 days later reported the same activity status on the initial and follow-up survey (Figure 1B). The mean absolute difference in sCDAI score from baseline to follow-up was 29.1 (s.d. 56.9) across all recall groups, with no statistically significant difference in means across groups ( $p=0.32$ ), although the range was slightly wider for those responding 3–5 days after the office visit

(Figure 1C) and in those with active disease (Figure 1D). Adjustment for age, sex, race, and self-reported disease activity did not appreciably alter the results (adjusted  $p=0.26$ ).

Agreement of categorization of disease as active or inactive at the time of the office visit and in the follow-up survey was similarly good for the 6-point Mayo score (overall 89%, 1–2 days 100%, 3–5 days 80%, 6–8 days 100%). Three patients who were categorized as inactive at the time of the office visit would have been categorized as active using the recall of data from 3–5 days prior. The difference in 6-point Mayo score computed at baseline and on subsequent recall was not statistically significantly different by study group (i.e., 1–2, 3–5, or 6–8 days later) ( $p=0.66$  unadjusted,  $p=0.13$  adjusted for age, sex, race and self-reported disease activity).

### Day-to-day Variability of Bowel Symptoms

In study 2, we assessed day-to-day variability within a 7-day period. Out of the 112 study participants who completed one or more days of the questionnaire, only 43 (38%) completed all seven days of data; 94 (84%) completed 4 or more days of bowel symptom questions and were included in the final analyses. The composition of the final cohort included 51 (54%) males and 43 (46%) females, and included 63 (67%) with CD and 31 (33%) with UC (Table 3). The mean weekly sCDAI for the 63 participants with CD was 204.2 (s.d. 67.2) and median 194 (range 128–396; IQR 147–239). Based on their sCDAI scores: 30% were in remission, 35% mild, and 35% moderate; no patients were classified as severe. The mean weekly Mayo score for the 31 patients with UC was 1.13 (s.d. 1.28) and median 0.57 (range 0–4.86; IQR 0.29–2.00). Using a Mayo score of 1.5 as the threshold for disease activity, 68% of participants with UC were considered inactive. Table 3 shows the demographic characteristics of the study participants.

We assessed the degree of variation of individual component scores for stool frequency, abdominal pain, and wellbeing as well as the sCDAI score by computing the difference between the minimum daily score and the maximum daily score for each patient. Among participants with CD, the median weekly maximum delta for sCDAI was 49 (range 0–329; IQR 28–98), bowel movement 2 (range 0–20; IQR 1–3), abdominal pain 0 (range 0–3; IQR 0–1), and wellbeing 0 (range 0–3; IQR 0–1). Among participants with UC, the median weekly maximum delta for the Mayo score was 1 (range 0–3; IQR 0–2), the median weekly maximum delta for the number of bowel movements above average was 1 (range 0–3; IQR 1–2), and the median weekly maximum delta for the bleeding score was 0 (range 0–1; IQR 0–0). The mean maximum deltas for sCDAI and Mayo scores were larger for patients with active CD ( $p<0.001$ ) and active UC ( $p=0.03$ ), respectively. There was no significant relationship between the sCDAI scores or the maximum weekly deltas in sCDAI scores and the number of days of questionnaires completed. There was a significant relationship between the maximum weekly deltas for sCDAI scores and sex ( $p=0.01$  unadjusted;  $p=0.04$  adjusted for race and age), with higher maximum weekly delta for females. No other significant relationship was found with the maximum weekly deltas for sCDAI scores.

The computed sCDAI based on fewer days of data collection correlated strongly with the full 7-day sCDAI (Table 4, Figures 2A and 2B). The Spearman's correlation coefficients were 0.86, 0.88, 0.94, and 0.97 when comparing the sCDAI of day 7, days 6–7, days 5–7,

and days 4–7 to the full week sCDAI, respectively (Table 4A). When comparing non-consecutive days of data to the weekly sCDAI scores, the Spearman's correlation coefficients were similar: 0.98 (days 1, 3, 5, and 7), 0.96 (days 2, 4, and 6), 0.94 (days 1 and 5), and 0.87 (days 2 and 6) (Table 4B). Spearman's coefficients ranged from 0.64–0.91 when comparing any individual day's projected sCDAI score with the full week sCDAI (Table 4C). Similar Spearman's coefficients were found for number of bowel movements (0.90–0.93), abdominal pain (0.74–0.89), and wellbeing (0.57–0.88) when comparing the daily scores to the weekly averages. When stratifying by sex, the Spearman correlation coefficients were comparable for males and females (data not shown). Likewise, similar results were observed for the subset of patients who had active CD as well as the patients with 7 days of electronic diary entries (Table 4).

Using the full 7 days of data as the gold standard, we assessed the proportion of patients who were misclassified as active or inactive when using only the last day, last 2 days, last 3 days, and last 4 days of data. Among patients with CD, 5% were misclassified as active or inactive using only day 7. Similarly, disease activity was misclassified in 4%, 8%, and 5% when using the last 2, 3, and 4 days, respectively. When using non-consecutive days of data, disease activity was misclassified in 0% (days 1 and 5), 8% (days 2 and 6), 6% (days 2, 4, and 6), and 0% (days 1, 3, 5, and 7). When relying on only the final day of data collection, 2 patients would have been misclassified as moderate instead of mild disease, 2 as remission instead of mild, and 3 as mild instead of moderate (Figure 2C). Using days 5–7 provided similar results (Figure 2D).

Using two-tailed, paired t-tests limited to the patients with UC, we did not observe significant differences between the Mayo scores calculated from the entire week of data and the final day ( $p=0.35$ ), the last 2 days ( $p=0.59$ ), or the last 3 days ( $p=0.83$ ) of data. Similar results were found when looking at non-consecutive days of data. The Spearman correlation of the 6-point Mayo computed using only the last day of data with the average of all available days was 0.82. This correlation improved slightly using the average Mayo scores across the last 2 ( $\rho=0.86$ ) or last 3 days ( $\rho=0.91$ ) of data. Using data from non-consecutive days to calculate the Mayo score produced similar correlation coefficients when compared to the weekly average scores. Among the 18 participants with UC who completed all 7 days of data, disease categorization as active or inactive was unchanged in 100% of participants when comparing the weekly average Mayo scores with the scores from the final day or final 2 days; 94% were unchanged using the final 3 days of data.

## DISCUSSION

Measurement of disease activity within clinical studies of IBD is in a state of transition. The full CDAI and sCDAI are cumbersome for patients and researchers alike due to the requirement of 7-day diaries. Possible solutions to make data collection more practical include shortening the number of days of data required or allowing patients to report on multiple days at a single time. The former approach has been suggested previously by Harvey and Bradshaw.<sup>15</sup> For UC, there is no standard for the number of days of symptoms to record. Our results demonstrate reasonable accuracy in patient's recall abilities for up to 8 days in both CD and UC, suggesting that it is feasible to have patients report on several days

of symptoms at a single time. In our second study, fewer than half of all patients (39%) completed all 7 days of the electronic diary on schedule, which confirms the difficulty of requiring full 7-day diaries with each day's questionnaire completed on discrete occasions. Importantly, we showed that bowel symptoms had relatively little variability within a 7-day period, supporting the approach of collecting fewer than 7 days of data. Although females with CD had significantly higher weekly maximum deltas in sCDAI scores, stratifying the analyses by sex did not reveal any noteworthy differences in the correlation of using fewer days of data to calculate sCDAI scores. Taken together, these data suggest that the sCDAI can be accurately computed using as little as 1 day of data, that there is slight improvement in accuracy with 2–4 days, and that this score remains accurate when allowing several days to recall symptoms.

These results will become increasingly important as date and time stamps are more commonly used in electronic questionnaires, forcing investigators to determine when too many days have elapsed for bowel symptom data to continue to be accurate. With this information, we infer that data on bowel symptoms for UC and CD can be used if a participant completes the data recall within up to 8 days, although we suspect that the data are likely to be most accurate if recall is completed within 1–2 days.

To our knowledge, this study is the first to assess recall of CD or UC activity. Previous data on gastrointestinal symptoms come primarily from infectious diarrhea parental recall in underdeveloped nations and demonstrate that parents are more likely to under-report diarrhea if the recall period was longer than 2–3 days.<sup>16–19</sup> Our data on bowel symptom recall contradicted this notion in IBD patients, demonstrating accurate recall for up to 8 days. This discrepancy is likely due to the nature of the diseases that were studied. Infectious diarrhea is an acute illness that results in dramatic change in bowel frequency but resolves quickly. CD and UC are chronic diseases and, as we have demonstrated, there is relatively little day-to-day variability in symptoms. For patients with an identical bowel pattern each day, it is not difficult to accurately recall symptoms over longer periods. Furthermore, patients with IBD may become attuned to their daily bowel symptoms, which may improve recall. Although not statistically significant, it appeared that recall of symptoms among patients with active disease was slightly less accurate than in those in remission. This may be due to greater day-to-day variability, particularly if a new treatment was initiated, as is seen in patients with infectious diarrhea.

Similar studies in other chronic diseases have generally been consistent with our results. For example, a 7-day recall study of respiratory symptoms in patients with Chronic Obstructive Pulmonary Disease showed that scores in a 7-day recall were higher compared to the daily recall scores; nevertheless, they were equivalent in detecting change over time.<sup>9</sup> Other studies have examined the recall of pain in patients with rheumatological diseases.<sup>10–12,20</sup> Although recall rating of pain was higher than momentary pain assessment overall, it was shown to correlate well with momentary ratings for the first 2 days, becoming unreliable after the second day.<sup>10</sup> Similar studies on pain recall have also shown that patients with higher variability of momentary pain report higher levels of recalled pain compared to those with low variability.<sup>11</sup>



Our studies had several limitations. In our evaluation of recall accuracy, participants did not have to recall all 7 days at once, but rather recalled only their original baseline day. We focused on the ability to recall a specific day so as to not prime the participant's memory by collecting daily symptom data for a full week before asking the patient to recall the data. Nonetheless, if our baseline questionnaire primed the participant's memory, we may have over-estimated the accuracy of recall. We tried to minimize this overestimation by including other questions related to IBD in the questionnaires (Supplemental Figure 1).

Our studies used the sCDAI and 6-point Mayo scores as the gold standard. We did not assess more objective markers of disease activity such as endoscopy, imaging, or biomarkers in the blood and stool. However, the purpose of the study was not to assess the accuracy of the indices, but rather the ability to capture the same data more efficiently. We anticipate that similar effects would apply to measures of patient reported outcomes that are currently under development.

By chance rather than design, we did not have any patients with severely active disease. However, it is unlikely that recall of symptoms for patients with severely active disease would be sufficiently inaccurate to misclassify them as remission or mild disease. Likewise, while there may be greater day-to-day variability in patients with active disease, it is unlikely that their symptoms over any 1–3 days would not capture the active disease given the low degree of misclassification of those with moderate disease when using only 1 or 3 days of symptoms. Similarly, the study neither excluded nor analyzed data specifically from patients who were changing therapies between initial and follow-up questionnaires. As such, our results may slightly overestimate the day-to-day variability in symptoms among patients on stable regimens.

Our study population was highly educated with approximately two-thirds having a college or graduate degree and 50% working in professional occupations. If higher education is a surrogate for better recall, this may limit the generalizability of these results to other populations.

In summary, these data provide guidance on ways to simplify data collection for IBD symptoms. Given the stability of disease activity described in our results over a 7-day period, fewer than 7 days of data can be used to accurately document current symptoms of IBD. Even with only 1 day of data, misclassification is minimal when examining whether a patient's disease is active or inactive. The slight decrease in accuracy that results from shorter data collection periods may be outweighed by the potential benefits, including decreased research costs, decreased burden on study participants and research staff, and less missing data. Alternatively, our data suggest that patients can recall symptoms from several days prior with reasonable accuracy, thereby allowing for collection of several days of data at one time. These principles can also be applied to the use of the sCDAI and Mayo score and to the development of new patient reported outcome measures.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Source of Funding:

This work was supported by NIH grants K24-DK078228 (JDL), K08-DK-095951 (FIS) and T32-DK007740 (JDL, MPH),

Dr. Lewis reports having served as a paid consultant for Amgen and AstraZeneca for work related to patient reported outcome measures. He has served as a paid consultant for work outside of the scope of this manuscript for Shire, Lilly, Janssen, AbbVie, Immune Pharmaceuticals, Pfizer, Takeda, Merck, MedImmune. Dr. Lewis has received research funding from Shire, Takeda Bayer, and Nestle Health Science.

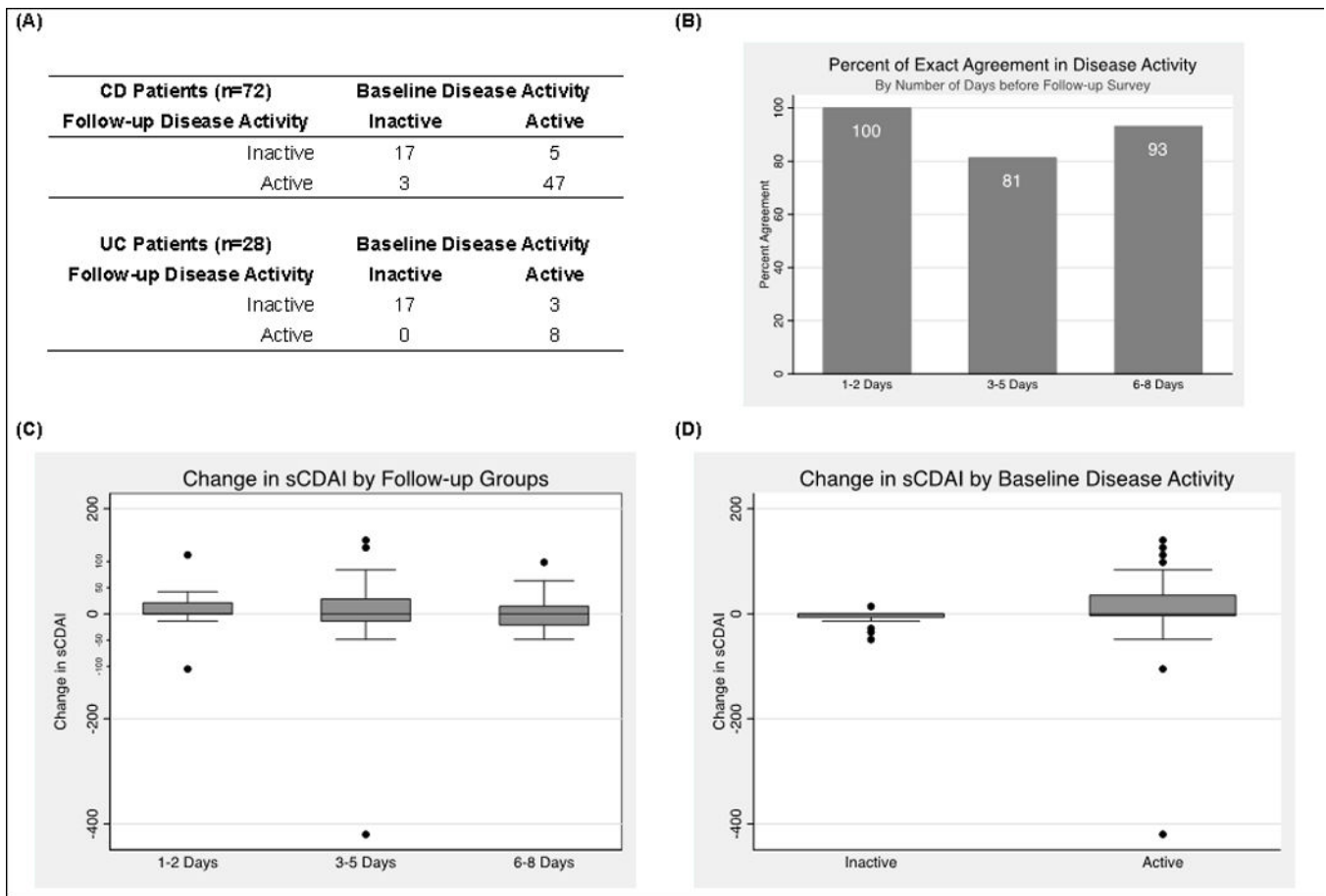
Dr. Lichtenstein reports having served as a paid consultant for work outside of the scope of this manuscript for the following: Abbott Corporation / Abbvie, Actavis, Alaven, Ferring, Hospira, Janssen Orthobiotech, Luitpold / American Regent, Pfizer Pharmaceuticals, Prometheus Laboratories, Inc., Salix Pharmaceuticals, Santarus, Shire Pharmaceuticals, Takeda, UCB, and Warner Chilcotte. He reports research support from Ferring, Janssen Orthobiotech, Prometheus Laboratories, Inc., Salix Pharmaceuticals, Santarus, Shire Pharmaceuticals, UCB, and Warner Chilcotte. He has received honorarium for participation in CME events from Ironwood and Luitpold / American Regent.

Dr. Bewtra reports having received research support from Janssen Orthobiotech.

## REFERENCES

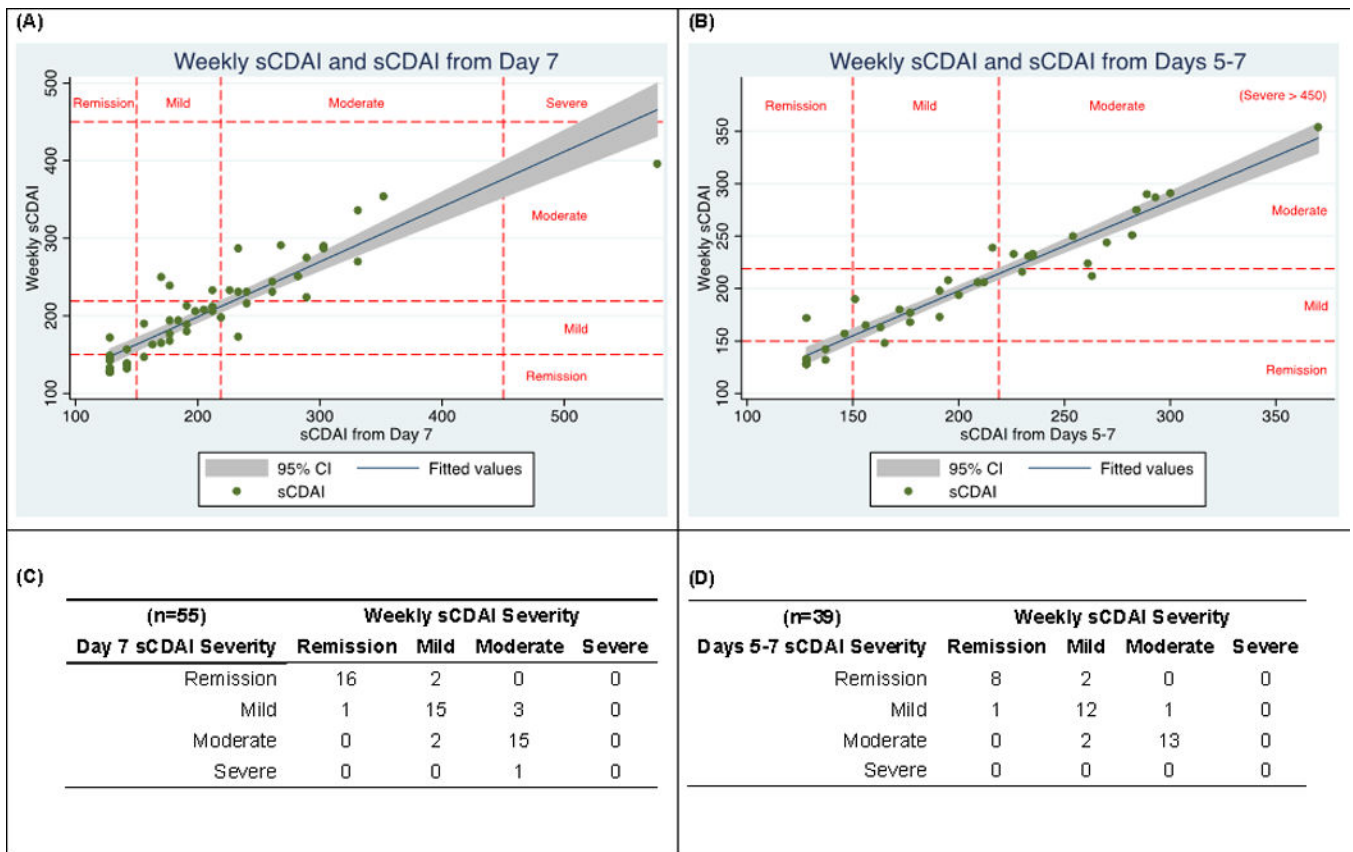
1. Best W, Beckett J, Singleton J, et al. Development of a Crohn's Disease Activity Index. National Cooperative Crohn's Disease Study. *Gastroenterology*. 1976; 70:439–444. [PubMed: 1248701]
2. Yoshida E. The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire: A review of instruments to assess Crohn's disease. *Can. J. Gastroenterol*. 1999; 13:65–73. [PubMed: 10099817]
3. Sandborn W, Feagan B, Hanauer S, et al. A Review of Activity Indices and Efficacy Endpoints for Clinical Trials of Medical Therapy in Adults with Crohn's Disease. *Gastroenterology*. 2002; 122:512–530. [PubMed: 11832465]
4. Thia K, Faubion WA, Loftus EV, et al. Short CDAI: Development and validation of a shortened and simplified Crohn's disease activity index. *Inflamm. Bowel Dis*. 2011; 17:105–111. [PubMed: 20629100]
5. Lewis JD, Chuai S, Nessel L, et al. Use of the Non-invasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis. *Inflamm. Bowel Dis*. 2008; 14:1660–1666. [PubMed: 18623174]
6. Bewtra M, Brensinger CM, Tomov VT, et al. An Optimized Patient-reported Ulcerative Colitis Disease Activity Measure Derived from the Mayo Score and the Simple Clinical Colitis Activity Index: *Inflamm. Bowel Dis*. 2014; 20:1070–1078. [PubMed: 24810138]
7. Litcher-Kelly L, Kellerman Q, Hanauer S, et al. Feasibility and Utility of an Electronic Diary to Assess Self-Report Symptoms in Patients With Inflammatory Bowel Disease. *Ann. Behav. Med*. 2007; 33:207–212. [PubMed: 17447873]
8. Stone AA, Shiffman S, Schwartz JE, et al. Patient non-compliance with paper diaries. *BMJ*. 2002; 324:1193. [PubMed: 12016186]
9. Bennett AV, Amtmann D, Diehr P, et al. Comparison of 7-day recall and daily diary reports of COPD symptoms and impacts. *Value Health*. 2012; 15:466–474. [PubMed: 22583457]
10. Broderick JE, Schwartz JE, Vikingstad G, et al. The accuracy of pain and fatigue items across different reporting periods. *Pain*. 2008; 139:146–157. [PubMed: 18455312]
11. Stone AA, Schwartz JE, Broderick JE, et al. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers. Soc. Psychol. Bull*. 2005; 31:1340–1346. [PubMed: 16143666]
12. Stone AA, Broderick JE, Schwartz JE. Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue. *Contemp. Clin. Trials*. 2010; 31:483–490. [PubMed: 20620239]

13. Melmed GY, Siegel CA, Spiegel BM, et al. Quality Indicators for Inflammatory Bowel Disease: Development of Process and Outcome Measures. *Inflamm. Bowel Dis.* 2013; 19:662–668. [PubMed: 23388547]
14. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf.* 2009; 42:377–381.
15. Harvey RF, Bradshaw JM. A Simple Index of Crohn's-Disease Activity. *The Lancet.* 1980; 315:514.
16. Alam N, Henry F, Rahaman M. Reporting errors in one-week diarrhoea recall surveys: experience from a prospective study in rural Bangladesh. *Int. J. Epidemiol.* 1989; 18:697–700. [PubMed: 2807676]
17. Zafar S, Luby S, Mendoza C. Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. *Epidemiol. Infect.* 2010; 138:264–269. [PubMed: 19653923]
18. Byass P, Hanlon P. Daily morbidity records: recall and reliability. *Int. J. Epidemiol.* 1994; 23:757–763. [PubMed: 8002190]
19. Ramakrishnan R, Venkatarao T, Koya P, et al. Influence of recall period on estimates of diarrhoea morbidity in infants in rural Tamilnadu. *Indian J. Public Health.* 1999; 43:136–139. [PubMed: 11243060]
20. Stone AA, Broderick JE, Kaell AT. Single momentary assessments are not reliable outcomes for clinical trials. *Contemp. Clin. Trials.* 2010; 31:466–472. [PubMed: 20580945]



**Figure 1. Changes in disease activity in study 1**

(A) The number of participants with CD (n=72) and UC (n=28) who switched from active to inactive or vice versa; (B) Percent agreement of CD activity status stratified by number of days from baseline measurement to recall of symptoms; (C and D) Box and whiskers plot of the difference of sCDAI scores at baseline and follow-up, stratified by timing of follow-up survey (C) and baseline disease activity (D).



**Figure 2. Agreement of sCDAI computed with 7 days of data and with fewer days of data collection**

(A) Correlation of the 7-day sCDAI with only day 7 ( $r^2=0.83$  for the regression line shown); (B) Correlation of the 7-day sCDAI with days 5–7 ( $r^2=0.93$ ); (C and D) Agreement of categorization of disease activity using all 7 days of data with available day 7 data (C) and available days 5–7 data (D).

**Table 1**

Characteristics of participants in study 1 examining patient recall of symptoms.

	Crohn's Disease* (n=72)	Ulcerative Colitis* (n=28)	Total (n=100)
<b>GENDER</b>			
Male	34 (47%)	16 (57%)	50 (50%)
Female	38 (53%)	12 (43%)	50 (50%)
<b>AGE</b>			
Mean Age (S.D.)	35.8 (14.1)	41.1 (19.2)	37.3 (15.8)
<b>RACE</b>			
Asian	4 (6%)	2 (7%)	6 (6%)
Black or African America	6 (8%)	0 (0%)	6 (6%)
White or Caucasian	60 (83%)	26 (93%)	86 (86%)
More than one race	1 (1%)	0 (0%)	1 (1%)
Other / Do not wish to report	1 (1%)	0 (0%)	1 (1%)
<b>ETHNICITY</b>			
Hispanic or Latino	0 (0%)	0 (0%)	0 (0%)
Not Hispanic or Latino	67 (93%)	28 (100%)	95 (95%)
Unknown / Not reported	5 (7%)	0 (0%)	5 (5%)
<b>EDUCATION</b>			
Less than High School Degree	0 (0%)	1 (4%)	1 (1%)
High School Diploma	5 (7%)	1 (4%)	6 (6%)
Some College	20 (29%)	3 (11%)	23 (23%)
Bachelors Degree	28 (40%)	12 (43%)	40 (40%)
Graduate Degree	17 (24%)	11 (39%)	28 (28%)
Not reported	2 (3%)	0 (0%)	2 (2%)
<b>OCCUPATION</b>			

	Crohn's Disease* (n=72)	Ulcerative Colitis* (n=28)	Total (n=100)
Retired	4 (6%)	3 (11%)	7 (7%)
Unemployed	1 (1%)	1 (4%)	2 (2%)
Student	12 (17%)	4 (14%)	16 (16%)
Manual Laborer	2 (3%)	1 (4%)	3 (3%)
Office Worker	6 (9%)	3 (11%)	9 (9%)
Homemaker	3 (4%)	2 (7%)	5 (5%)
Professional	34 (49%)	14 (50%)	48 (48%)
Other	8 (11%)	0 (0%)	8 (8%)
Not reported	2 (3%)	0 (0%)	2 (2%)
<b>TOBACCO†</b>			
Tobacco Users	15 (21%)	2 (7%)	17 (17%)

\* Diagnosis with CD or UC was based on physician diagnosis as recorded in the medical record.

† Data for tobacco use were not available for 2 participants (2%) in Study 1, both with CD (3% of CD participants).

**Table 2**

The mean absolute values of the difference between initial and follow-up responses

	1-2 Days Follow-Up (n=23)		3-5 Days Follow-Up (n=48)		6-8 Days Follow-Up (n=29)		P Value*
Participants with CD (n)	19	33	20	33	20	33	
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	
<b>sCDAI</b>	19.15 (34.52)	40.09 (76.79)	20.30 (25.48)	20.30 (25.48)	20.30 (25.48)	20.30 (25.48)	0.32
<b>Bowel Movements</b>	0.47 (0.47)	0.85 (1.20)	1.25 (2.10)	1.25 (2.10)	1.25 (2.10)	1.25 (2.10)	0.26
<b>Abdominal Pain</b>	0.11 (0.32)	0.18 (0.39)	0.15 (0.37)	0.15 (0.37)	0.15 (0.37)	0.15 (0.37)	0.78
<b>Wellbeing</b>	0.21 (0.54)	0.38 (0.79)	0.15 (0.37)	0.15 (0.37)	0.15 (0.37)	0.15 (0.37)	0.35
<b>Liquid Stools</b>	0.58 (1.12)	1.09 (2.85)	0.85 (0.88)	0.85 (0.88)	0.85 (0.88)	0.85 (0.88)	0.69
<b>Participants with Active CD (n)</b>	<b>14</b>	<b>24</b>	<b>14</b>	<b>24</b>	<b>14</b>	<b>24</b>	
<b>sCDAI</b>	25.00 (38.70)	51.63 (87.01)	25.00 (28.63)	25.00 (28.63)	25.00 (28.63)	25.00 (28.63)	0.34
<b>Bowel Movements</b>	0.50 (1.09)	0.96 (1.30)	1.57 (2.44)	1.57 (2.44)	1.57 (2.44)	1.57 (2.44)	0.23
<b>Abdominal Pain</b>	0.14 (0.36)	0.17 (0.38)	0.21 (0.43)	0.21 (0.43)	0.21 (0.43)	0.21 (0.43)	0.88
<b>Wellbeing</b>	0.29 (0.61)	0.54 (0.88)	0.21 (0.43)	0.21 (0.43)	0.21 (0.43)	0.21 (0.43)	0.34
<b>Liquid Stools</b>	0.71 (1.27)	1.46 (3.28)	0.93 (0.92)	0.93 (0.92)	0.93 (0.92)	0.93 (0.92)	0.62
<b>Participants with UC (n)</b>	<b>4</b>	<b>15</b>	<b>9</b>	<b>15</b>	<b>9</b>	<b>15</b>	
<b>Mayo Score</b>	0.25 (0.50)	0.60 (0.83)	0.44 (0.53)	0.44 (0.53)	0.44 (0.53)	0.44 (0.53)	0.66
<b>Stool Frequency Above Normal</b>	0.25 (0.50)	0.23 (0.46)	0.33 (0.50)	0.33 (0.50)	0.33 (0.50)	0.33 (0.50)	0.93
<b>Participants with Active UC (n)</b>	<b>1</b>	<b>10</b>	<b>0</b>	<b>10</b>	<b>0</b>	<b>10</b>	
<b>Mayo Score</b>	0.00 (0.00)	0.60 (0.97)	-	0.60 (0.97)	-	0.60 (0.97)	0.57
<b>Stool Frequency Above Normal</b>	0.00 (0.00)	0.20 (0.42)	-	0.20 (0.42)	-	0.20 (0.42)	0.66*

\* P-values are from ANOVA across categories of time from baseline to follow-up survey.



Characteristics of patients included in study 2 examining day-to-day variability of bowel symptoms.

**Table 3**

	Crohn's Disease* (n=63)	Ulcerative Colitis* (n=31)	Total (n=94)
<b>GENDER</b>			
Male	31 (49%)	20 (65%)	51 (54%)
Female	32 (51%)	11 (35%)	43 (46%)
<b>AGE</b>			
Mean Age (S.D.)	37.6 (14.8)	41.5 (17.5)	38.8 (15.7)
<b>RACE</b>			
Asian	5 (8%)	3 (10%)	8 (9%)
Black or African America	5 (8%)	0 (0%)	5 (5%)
White or Caucasian	52 (81%)	28 (90%)	80 (85%)
More than one race	1 (2%)	0 (0%)	1 (1%)
Not Reported	1 (2%)	0 (0%)	1 (1%)
<b>ETHNICITY</b>			
Hispanic or Latino	0 (0%)	0 (0%)	0 (0%)
Not Hispanic or Latino	63 (100%)	31 (100%)	94 (100%)
<b>EDUCATION</b>			
Less than High School Degree	0 (0%)	1 (3%)	1 (1%)
High School Diploma	2 (3%)	3 (10%)	5 (5%)
Some College	17 (27%)	2 (6%)	19 (20%)
Bachelors Degree	21 (33%)	9 (29%)	30 (32%)
Graduate Degree	18 (29%)	14 (45%)	32 (34%)
Not Reported	5 (8%)	2 (6%)	7 (7%)
<b>OCCUPATION</b>			

	Crohn's Disease* (n=63)	Ulcerative Colitis* (n=31)	Total (n=94)
Retired	3 (5%)	3 (10%)	6 (6%)
Unemployed	0 (0%)	0 (0%)	0 (0%)
Student	10 (16%)	3 (10%)	13 (14%)
Manual Laborer	1 (2%)	1 (3%)	2 (2%)
Office Worker	4 (6%)	3 (10%)	7 (7%)
Homemaker	1 (2%)	3 (10%)	4 (4%)
Professional	34 (54%)	14 (45%)	48 (51%)
Other	5 (8%)	2 (6%)	7 (7%)
Not Reported	5 (8%)	2 (6%)	7 (7%)
<b>TOBACCO<sup>†</sup></b>			
Tobacco Users	10 (16%)	2 (6%)	12 (13%)

\* Diagnosis with CD or UC was based on physician diagnosis as recorded in the medical record.

<sup>†</sup> Data for tobacco use were not available for 7 participants (7%) in Study 2; 5 with CD (8%) and 2 with UC (6%).

Correlation of sCDAI or Mayo scores computed from fewer than 7 days of data compared to one week of data collection.\* (A) Correlation of the last 1–4 days using consecutive days of data. (B) Correlation using non-consecutive days of data. (C) Correlation of individual days of data.

**Table 4**

	Day 7	Days 6–7	Days 5–7	Days 4–7
<b>CD patients</b>	0.91	0.93	0.96	0.98
<i>Number of observations</i>	55	48	39	38
<b>CD patients with active disease<sup>†</sup></b>	0.81	0.87	0.94	0.95
<i>Number of observations</i>	38	35	30	29
<b>CD patients with 7 full days entries</b>	0.86	0.88	0.94	0.97
<i>Number of observations</i>	25	25	25	25
<b>UC patients</b>	0.81	0.85	0.91	0.95
<i>Number of observations</i>	27	24	22	21
<b>UC patients with active disease<sup>#</sup></b>	0.76	0.86	0.95	0.98
<i>Number of observations</i>	8	8	8	8
<b>UC patients with 7 full days entries</b>	0.75	0.79	0.87	0.92
<i>Number of observations</i>	18	18	18	18

	Days 1 & 5	Days 2 & 6	Days 2, 4, & 6	Days 1, 3, 5, & 7
<b>CD patients</b>	0.96	0.95	0.98	0.99
<i>Number of observations</i>	44	50	47	33
<b>CD patients with active disease<sup>†</sup></b>	0.91	0.90	0.96	0.97
<i>Number of observations</i>	32	37	34	24
<b>CD patients with 7 full days entries</b>	0.94	0.87	0.96	0.98
<i>Number of observations</i>	25	25	25	25

(B)

	Days 1 & 5	Days 2 & 6	Days 3, 4, & 6	Days 1, 3, 5, & 7
<b>UC patients</b>	0.87	0.93	0.96	0.89
<i>Number of observations</i>	23	23	22	20
<hr/>				
<b>UC patients with active disease<sup>#</sup></b>	0.95	0.96	0.99	0.99
<i>Number of observations</i>	7	7	7	6
<hr/>				
<b>UC patients with 7 full days entries</b>	0.86	0.94	0.95	0.90
<i>Number of observations</i>	18	18	18	18

(C)

	1	2	3	4	5	6	7
<hr/>							
<b>CD patients</b>	0.80	0.91	0.92	0.94	0.95	0.89	0.91
<i>Number of observations</i>	55	57	54	59	52	54	55
<hr/>							
<b>CD patients with active disease<sup>†</sup></b>	0.60	0.79	0.81	0.90	0.91	0.83	0.81
<i>Number of observations</i>	39	40	36	40	37	39	38
<hr/>							
<b>CD patients with 7 full days entries</b>	0.64	0.77	0.86	0.89	0.91	0.82	0.86
<i>Number of observations</i>	25	25	25	25	25	25	25
<hr/>							
<b>UC patients</b>	0.86	0.83	0.75	0.86	0.81	0.86	0.81
<i>Number of observations</i>	29	30	27	29	25	24	27
<hr/>							
<b>UC patients with active disease<sup>#</sup></b>	0.83	0.90	0.68	0.83	0.80	0.75	0.76
<i>Number of observations</i>	9	9	8	10	8	8	8
<hr/>							
<b>UC patients with 7 full days entries</b>	0.83	0.85	0.74	0.81	0.74	0.83	0.75
<i>Number of observations</i>	18	18	18	18	18	18	18

\* Data are reported as Spearman correlation coefficients.

<sup>†</sup> sCDAI based on 1 week of data > 150.

<sup>#</sup> Mayo score based on 1 week of data > 1.