Automated vs Subjective assessment of spatial and temporal MRI small bowel motility in Crohn's disease

Introduction

Recent insights from high resolution manometry[1], wireless motility pills[2] and magnetic resonance imaging (MRI)[3] have demonstrated a wide range of diverse contractile processes in the intestine, which are challenging our understanding of intestinal physiology.

MR enterography (MRE) is now widely disseminated in clinical practice, notably in patients with Crohn's disease (CD) and protocols increasingly include cine sequences aimed at capturing bowel motility. It is well established that motility is reduced in segments of bowel affected by Crohn's disease, and that this reduction is correlated with the severity of inflammation[4,5]. However recent work using automated software quantification has suggested that aberrant motility in apparently unaffected bowel, based on visual assessment, is linked to the severity of abdominal symptoms[6,7]. The ability of MRE to capture abnormal motility patterns in structurally normal bowel could prove a powerful tool in improving our understanding of gastrointestinal (GI) motility in health and disease. Applications extend beyond Crohn's disease and, for example, might include irritable bowel syndrome where a constellation of poorly resolved processes including visceral hypersensitivity, inflammation and dysmotility may be present[8].

However, there remain many unknowns. For example, it is unclear which metrics or combination thereof best capture aberrant motility, and whether radiologists can reliably detect abnormal motility without the need for specialised software. To date most researchers have examined a small number of motility metrics. For example, Menys et al.

(2016)[6] reported an association between reduced global motility variance and patient symptoms in CD while Bickelhaupt et al. (2013)[9] suggested that contraction frequencies are altered in the bowel distal to segments with inflammatory activity. There has been no previous research into the ability of radiologists to detect abnormal motility patterns as part of their conventional reporting of MRE datasets providing added value without the need for additional software.

The purpose of the current study was twofold. Firstly, we investigated both established[6] and newly proposed computer-based metrics to derive the best combination associated with CD patient abdominal symptoms. Secondly, we compared subjective grading of bowel motility by experienced radiologists with automated measurement and investigated inter- and intra-observer variation.

Materials and Methods

Patient Selection

The current study is retrospective and has been approved by both centres' ethics committees. The patients provided written informed consent for the original research studies and the requirement for consent was waived for the retrospective analysis in this study.

Data was collated from two previous studies; i) the two-centre VIGOR++ (study 1) , a prospective trial developing automated measurement of bowel wall thickness and contrast enhancement to quantify CD activity[10], and ii) a prospective single-centre study (study 2)

developing a global MRI CD activity score (MEGS)[11]. Patients recruited to both these studies completed a HBI symptom questionnaire the day of an MRE examination which included a motility sequence. The HBI is a validated symptom-based activity score in which patients grade the severity of abdominal pain and wellbeing. The other score components include the number of liquid stools per day, abdominal mass and complications (Supplementary material 1).

Patients were selected for the current study if they had a final diagnosis of CD based on clinical, biochemical, endoscopic, imaging and histopathological data and if associated dynamic MRI and HBI data was available.

A total of 185 datasets across the two parent studies were potentially available for inclusion in the current study. Datasets were excluded if the dynamic MRI sequences were inadequate (e.g. less than 3 slices, incomplete time series or motility data unavailable to this study), a HBI score was not collected or the patient had a final diagnosis other than Crohn's disease.

A proportion of patient data (n=28) used in the current study was also used in previous work investigating the relationship between two motility metrics (mean global motility and variance of global motility) and abdominal symptoms[6, 7].

Demographic data pertaining to age, sex, current medication, disease duration and surgical history of the selected patients was collected.

MRI Protocol

Patients fasted for 4 hours before ingesting oral contrast prior to undergoing MRI in the supine position on either 1.5T (Avanto, Siemens, Medical Systems, Erlangen, Germany) or 3T

(Achieva: Philips, Best, the Netherlands) units. The MRI protocol included a dynamic "cine motility" sequence acquired during a 20 second breath-hold, and prior to administration of anti-spasmodics for anatomical images. Specifically, a multi-slice 2D, coronal, balanced steady-state free precession sequence with a temporal resolution of 1 image/second and a slice thickness of 10mm was acquired coronally. Repeat coronal block acquisitions were performed to encompass the whole small bowel volume, the number of acquisitions ranging from 5 to 16 depending on the size of the patient (Supplementary material 2).

Motility Assessment

For the current study, two graphical user interfaces (GUIs) were developed using MATLAB (MathWorks, Natick, MA). A general viewer allowed for inspection of all the MRI data (as well as ROI placement and automated MRI metric measurement), and a second "radiologist viewer" which presented data in a blinded and pre-set order and facilitated subjective scoring of metrics by the study radiologists (Supplementary material 3).

For each 2D cine motility sequence, the frames were registered using a previously validated optic-flow based registration technique which produced a set of deformation fields[3]. In summary, the deformation fields' Jacobian determinants were used to derive automated motility metrics and the reference frame was used for ROI placement and derivation of the distension metric. The process of implementing the registration and generating the motility metrics for each patient is summarised in figure 1, and each metric is described in more detail below in the subsection "Automated Assessment: motility metric measurements".

In the general viewer, anonymised datasets were displayed both as a static reference image and as a "cine" movie. The radiologist viewer displayed the "cine" movie only and

was used for radiologist subjective grading of motility (described below in the subsection "Subjective Radiological Assessment").

ROI placement and slice selection

For each patient, a study coordinator (research fellow) with 6 months training in enteric MRI (RM Gollifer) and blinded to the HBI score used the general viewer to place ROIs over the small bowel on the reference image, with the cine motility movies available to aid ROI placement. The ROIs were validated by a research fellow with over 5 years MRE experience (A Menys).

In detail, for each of the motility datasets, ROIs were placed in morphologically normally appearing small bowel on all the coronal motility acquisition slices. The single coronal slice which contained the largest area of small bowel was then objectively selected based on the largest number of "small bowel" pixels that could be encompassed by a single ROI.

The ROIs excluded small bowel mesentery and Crohn's affected small bowel (SB) i.e. SB demonstrating wall thickening, abnormal T2 signal hyper enhancement etc.[12].

Automated Assessment: motility metric measurements

The automated metrics were developed to capture motility features in a single acquisition slice.

Five metrics were derived from the ROIs within the selected slice: 1) mean motility, 2) spatial variation, 3) temporal variation, 4) area of motile bowel and 5) intestinal distension.

Metrics 1, 2 and 4 were derived from the motility map generated from the standard deviation of all the deformation fields' Jacobian determinants i.e. the SD Jacobian which

summarises the local expansions and contraction on a per pixel basis throughout the entire time series. Metric 3 was derived from a temporal variation map and metric 5 was derived from the reference frame (figure 1).

In detail, metric 1 and metric 2 were derived by calculating the mean and the variance, respectively, of the SD Jacobian values. Similarly, mean motility and spatial variation metrics have been derived in a previous study, but across multiple slices[6]. Metrics 1 and 2 in the current study were applied to a single slice. Metric 4, the area of motile bowel was defined as the percentage of pixels with an SD Jacobian above a threshold of 0.11. The cut-off of 0.11 was selected based on the work of Odille et al suggesting bowel with a SD Jacobian < 0.11 is classified as immotile[3] (figure 2G-H). The temporal variation metric was derived by firstly calculating the standard deviation of the deformation fields' Jacobian determinant in multiple 5 second (or 5 frame) sliding windows, henceforth referred to as the Sliding SD Jacobian Value. For example, in a 20 second time series there would be 16 sliding windows i.e. 1-5 seconds (window 1) to 16-20 seconds (window 16). Each Sliding SD Jacobian Value is a per pixel measure of bowel expansion and contraction within a 5 second time period. The temporal variation map was generated by calculating the variance of these Sliding SD Jacobian Values which captures the difference in motility between the sliding windows for each pixel. This temporal variation metric gives an indication of variability of motility over time e.g. low temporal variation corresponds to consistent motility (either constantly high motility or constantly low motility) throughout the entire scan time while high temporal variation corresponds to a wide range of Sliding SD Jacobian Values suggesting a higher proportion of the small bowel with fluctuating motility, between low and high (figure 2E-F).

The intestinal distension metric was developed based on the intensity of the pixels within the ROIs, and their neighbours, in the reference frame. A binary mask was created with each pixel assigned a value of 1 if the signal intensities of 6 out of 9 of their neighbouring pixels (8 neighbours and the pixel being analysed) were above a threshold of 50% of the median intensity within the ROIs. The value of intestinal distension was indicated by the percentage of pixels assigned a value of 1 (a high value representing higher signal, suggesting good distension with mannitol, and a low value indicating lower signal, presumed due to small bowel collapse).

Subjective Radiological Assessment

The same five features were visually assessed using the cine motility time series for the chosen slice.

The study coordinator (RM Gollifer) conducted a training session with two experienced radiologists (10 and 12 years' experience of MRE) to explain the five metrics and what they represented in terms of different motility patterns. For example, for metric 1 (mean motility), the radiologists were told to subjectively grade the average motility of the small bowel across the slice, for metric 2 (spatial variation), they were asked to grade how variable motility was within the area of the ROI, and for metric 3 (temporal variation) they were asked to grade how the motility of the bowel changed over the 20 second time series. Fifteen datasets outside of the main study dataset were selected to demonstrate examples of different combinations of low, medium and high scores for metrics 1, 2 and 3. During the training session, these datasets were visually assessed firstly by each radiologist blinded to each other and then in consensus to agree upon a scoring scheme.

To record their grading, the radiologists would view the “cine” movie for each dataset and visually grade each of the five metrics on a sliding 0-10 scale (or % scale for area of motile bowel) discretised in increments of 0.1.

The two radiologists both scored all study datasets, blinded to the scores of the other. The datasets were presented in random order. Reading sessions typically included 15 or 30 datasets and were performed at 1 or 2 weekly intervals, respectively. One in every five datasets presented was a duplicate dataset. For example, in a 15-dataset scoring session, 12 would be original data and 3 would be duplicate data, previously scored. The 3 duplicate datasets were randomly selected and presented to the radiologists at least 2 weeks after they were originally scored.

Radiologist 1 graded the datasets in the reverse order from radiologist 2 to account for learning effect bias.

Statistical analysis

All statistical analysis was performed using MATLAB (MathWorks, Natick, MA). All data was checked for normality using a Shapiro-Wilk test (alpha = 0.05).

Intra-observer and inter-observer variability between radiologist observers was assessed using Bland-Altman plots. Mean absolute differences, 95% limits of agreement (LOA) and coefficient of variation (CV) were calculated. A low CV would be considered good and a high CV considered poor. The area of motile bowel metric was graded as a percentage and then converted to a 0-10 scale.

Univariate and multivariate regression models were tested to assess the relationship between HBI patient symptoms scores and 1) Automated motility metrics, and 2) Subjective radiological motility features (based on the mean score of radiologist 1 and 2).

In both cases, thirty combinations of the five independent variables (metric 1 – mean motility, metric 2 – spatial variation, metric 3 – temporal variation, metric 4 – area of motile bowel, metric 5 – intestinal distension) were tested against the dependent variable of HBI.

The goodness of fit in the regression analysis is reported as R2 (adjusted) to account for the varying number of independent variables being tested each time. Note that for a perfect fit, the R2 (adjusted) value would be 1.

All variables in the models and the HBI were standardised so that the estimated coefficients between metrics of different scales could be directly compared. The larger the absolute value of the standardised coefficient estimate, the higher the importance of the variable in predicting HBI.

Multicollinearity was tested to rule out models containing high collinearity between independent variables with variance inflation factor (VIF) > 5 indicating a highly collinear variable[13]. If any of the standardised coefficient estimates within a model were insignificant, then the model was rejected.

Models were also excluded if the F-statistic was insignificant at the 5% significance level ($P > 0.05$) or if the R2 (adjusted) value was negative. Both these conditions indicate that the model poorly fits the data and is inferior to a simple intercept only i.e. fitting a horizontal line.

The best accepted models met the following criteria:

- Low multicollinearity i.e. all variables in the model had a low variance inflation factor (VIF < 5)[13]
- F-statistic for the model was significant (P < 0.05)
- R2 (adjusted) value for the model was positive (R2 > 0)
- Standardised coefficient estimates for all variables in the model were significant (P < 0.05)

The following covariates: Age, sex, history of surgery (yes/no) and disease duration were added as independent variables to the accepted models to see if the standardised coefficient estimates for the metrics retained significance.

Results

Cohort Demographics

185 datasets were available from the 2 donor studies (124 donor study 1, 61 donor study 2) (Supplementary material 4).

15 patients were selected for the radiologist training session.

65 patients were excluded for the following reasons; missing motility sequences (n=24) HBI data unavailable (n=32) and non-Crohn's disease final diagnosis (n=9).

The remaining 105 patients (77 donor study 1, 28 donor study 2) formed the final study cohort (Supplementary material 5).

HBI and Motility Metrics

A summary of automated motility metrics and HBI scores is shown in table 1.

Automated Assessment

Best objective models for motility metrics v HBI

The standardised coefficient estimates for the metrics retained significance when demographic information was added so the following results discuss the models without the addition of these covariates (table 2).

The highest R2 adjusted value for a univariate model was 0.034 and consisted of standardised spatial variation of motility, which was negatively associated with standardised HBI (Coefficient estimate = -0.21, $P < 0.05$) (table 2).

The highest R2 adjusted value for a multivariate model was 0.036 and included standardised temporal variation, which was negatively associated with HBI, (Coefficient estimate = -0.23, $P < 0.05$) and standardised area of motile bowel, which was positively associated with standardised HBI (Coefficient estimate = +0.16, $P < 0.05$) (table 2).

The regression models with the original units i.e. without standardising the independent variables and the HBI showed that for 0.01 unit increase in spatial variation, there was an associated 0.61 unit (95% confidence interval [CI], 0.18-1.03) decrease in HBI.

Assuming all other variables were kept constant, for each 0.01 unit increase in temporal variation there was an associated 0.97 (95% CI, 0.30–1.63) decrease in HBI and for each 10 percent increase in area of motile bowel there was an associated 0.88 (95% CI, 0.18-1.58) unit increase in HBI.

The fitted HBI generated from each of the two models was plotted against the actual HBI as shown in figure 3.

Subjective Radiological Assessment

Inter-observer variability

Inter-observer variability for visually assessed motility metrics was poor (table 3).

The lowest coefficient of variation (CV) was 26% for area of motile bowel and the highest CV was 71% for temporal variation of motility. The absolute mean difference ranged from 0.19 for area of motile bowel to 1.3 for temporal variation of motility. The narrowest 95% limits of agreement (LOA) was for temporal variation of motility and the widest 95% LOA was spatial variation of motility. The highest agreement was for area of motile bowel (figure 4A) and the lowest agreement was for spatial or temporal variation of motility (figure 4B).

Intra-observer variability

Generally, intra-observer variability was better than inter-observer variability, with lower mean differences, narrower 95% LOAs and lower CVs as seen in table 3 for radiologist 1 and for radiologist 2.

The lowest CV was 16% for mean motility (radiologist 2) and the highest CV was 81% for temporal variation of motility (radiologist 1). The absolute mean difference ranged from 0.05 for intestinal distension (radiologist 1) to 0.37 for spatial variation (radiologist 1). The narrowest 95% LOA was for mean motility (radiologist 2) and the widest 95% LOA was for spatial variation of motility (radiologist 1). The best intra-observer agreement was for radiologist 2 scoring of mean motility (figure 5A) and the worst intra-observer agreement was for radiologist 1 scoring of spatial or temporal variation of motility (figure 5B).

Subjective models for combined observer motility scores v HBI

None of the univariate or multivariate combined models using radiologist grading of motility metrics demonstrated an association with the HBI score.

Discussion

Our study confirms there is an association between motility metrics in morphologically normal appearing small bowel captured using cine MRI and the severity of patient symptoms in CD. We also show for the first time that radiologists cannot adequately grade these motility features by simple subjective evaluation, and that software-based quantification is likely required to capture this relationship.

To date, the ability of radiologists to subjectively assess motility has not been investigated beyond 'active' and 'inactive' motility. As would be expected, intra-observer variation was better than inter-observer variation. Visual gradings for the area of motile bowel had the lowest inter- and intra-observer variability, with lower intra-observer variability also found in the visual gradings for intestinal distension and the mean motility grading by radiologist 2. Conversely, grading of spatial variation and temporal variation were highly variable. Overall, our data shows that even between experienced radiologists inter-observer variation is poor suggesting that subjective grading of motility features is unlikely to be clinically useful. Indeed, neither univariate or multivariate linear regression revealed any association between radiologist grading and HBI score.

Automated motility metrics therefore would have clear advantages over subjective assessment and provide a more consistent assessment of motility. Automated measures have already been shown to be repeatable[14], and the models performed better when tested

against patient symptoms, with several models showing a relationship with HBI. This suggests they at least in part capture the likely aberrant small bowel motility in apparently normal bowel in patients with Crohn's disease.

Variation in spatial and temporal motility are clearly the most difficult motility features to visually assess yet they seem to have the strongest relationship to symptoms. For example, in a single centre study of 53 CD subjects, Menys et al. (2016) have previously reported a significant inverse correlation between global bowel motility variance and HBI ($r = -0.45$, $P<0.001$). In the current study utilising a larger dataset of 105 CD subjects collated from 2 recruitment sites a univariate model again suggested a negative association between spatial variation and HBI (R2 adjusted = 0.034) ($r = -0.21$, $P = 0.03$). Some overlap in the datasets between the current study and that of Menys et al. (2016)[6] must however be acknowledged with 28 patients used in both studies.

The reason why decreased spatial variation, which represents more homogenous motility over the bowel (either high or low), is associated with increased symptoms is not yet certain. The best performing model was a combination of decreased temporal variation and an increased area of motile bowel. This suggests bowel health is reflected by heterogenous and patchy motility with areas of low and high motility in different segments, presumably reflecting the different roles of the proximal and distal small bowel in transit of intestinal content and nutrient absorption. It would appear that "switching off" this heterogeneity (perhaps in response to small bowel inflammation in CD) is associated with increased abdominal symptoms.

We tested a range of putative metrics which reflect the absolute level of small bowel motility as well as spatial and temporal variation. Without a "gold standard" to define

patterns of global small bowel motility in health and disease, it is possible that the metrics do not fully reflect the motility phenomena they aim to capture. Indeed, it should be noted that the association between motility metrics and abdominal symptoms was not particularly strong; the best performing model had a modest R2 adjusted of 0.036. Aberrant motility in CD is complex and although some of our tested metrics show definite promise, it is likely further refinements will be needed in the future. For example, the temporal variation metric was calculated using 5 second sliding windows, and the size of the time window could easily be modified.

Although HBI is a validated patient symptom score, it is relatively simplistic. Alternative CD questionnaires have been developed which are more detailed such as the Inflammatory Bowel Disease Questionnaire (IBDQ)[15] and this represents a limitation in this study. It would be interesting to test the motility metrics against these more complex questionnaires to see if associations are stronger. However, HBI is easier to implement clinically and considers important patient symptoms such as pain, wellbeing and diarrhoea. All measures of patient symptoms by their very nature are subjective to some degree but remain the clinically important endpoint against which to develop new methods.

Another limitation to consider is the MRE protocol for capturing motility, specifically the preparation, scan duration and the slice selection. Since the motility data is only acquired during a 20 second breath-hold, we may not be capturing the true complex nature of bowel motility which may be apparent over longer time periods. It may be more beneficial to acquire longer free breathing datasets which might allow enough time for clearer motility patterns to emerge. Software is already available to correct for motion in data acquired during free breathing[16]. Regarding preparation, mannitol is a hyperosmotic,

low calorie stimulant and it differs significantly from usual food stuffs which can provoke symptoms in patients. It is however useful for identifying areas of low motility[14]. Alternate sources of preparation should also be considered[17].

The single slice chosen in this study was objectively based on encompassing the largest areas of small bowel within a single ROI. This avoids the problem of temporal incoherence in multi-slice analysis which occurs since slices in different acquisition blocks are acquired around several seconds apart. However, it should be noted that the motility varies depending on bowel segment[18] (and by inference slice position). Further work is needed to determine if single slice analysis is sufficient or whether multi-slice protocols are preferable. Ultimately 3D acquisitions would eliminate the temporal incoherence limitation, although they are technically challenging at an adequate temporal resolution[19].

In summary, we have shown that subjective grading of MRI motility cannot reliably capture motility metrics and that objective computer-based quantification is required. Spatial and temporal variation are particularly difficult to assess visually. An association between automated motility metrics and patient symptoms is again demonstrated suggesting the metrics are at least in part capturing the likely aberrant small bowel motility presence in CD patients and have potential as a powerful non-invasive tool to interrogate bowel motility in health and disease. Further research is needed to optimise MR acquisition protocols, and further refine and validate candidate motility metrics.

References

1. Dinning PG, Wiklendt L, Gibbins I, et al. Low-resolution colonic manometry leads to a gross misinterpretation of the frequency and polarity of propagating sequences: Initial results from fiber-optic high-resolution manometry studies. Neurogastroenterol Motil. 2013;25(10).

2. Maqbool S, Parkman HP, Friedenberg FK. Wireless Capsule Motility: Comparison of the SmartPill® GI Monitoring System with Scintigraphy for Measuring Whole Gut Transit. Dig Dis Sci. 2009;54(10):2167-2174.

3. Odille F, Menys A, Ahmed A, Punwani S, Taylor SA, Atkinson D. Quantitative assessment of small bowel motility by nonrigid registration of dynamic MR images. Magn Reson Med. 2012;68(3):783-793.

4. Bickelhaupt S, Pazahr S, Chuck N, et al. Crohn's disease: small bowel motility impairment correlates with inflammatory-related markers C-reactive protein and calprotectin. Neurogastroenterol Motil. 2013;25(6):467-473.

5. Plumb AA, Menys A, Russo E, et al. Magnetic resonance imaging-quantified small bowel motility is a sensitive marker of response to medical therapy in Crohn's disease. Aliment Pharmacol Ther. 2015.

6. Menys A, Makanyanga J, Plumb A, et al. Aberrant Motility in Unaffected Small Bowel is Linked to Inflammatory Burden and Patient Symptoms in Crohn's Disease. Inflamm Bowel Dis. 2016;22(2):424-432.

7. Gollifer RM, Menys A, Makanyanga J, et al. Relationship between MRI quantified small bowel motility and abdominal symptoms in Crohn's disease patients—a validation study. Br J Radiol. 2018.

8. Spiller R, Aziz Q, Creed F, et al. Guidelines on the irritable bowel syndrome: mechanisms and practical management. Gut. 2007;56:1770-1798.

9. Bickelhaupt S, Froehlich JM, Cattin R, et al. Differentiation between active and chronic Crohn's disease using MRI small-bowel motility examinations — Initial experience. Clin Radiol. 2013;68(12):1247-1253.

10. Puylaert CAJ, Schüffler PJ, Naziroglu RE, et al. Semiautomatic Assessment of the Terminal Ileum and Colon in Patients with Crohn Disease Using MRI (the VIGOR++ Project). Acad Radiol. 2018;25(8):1038-1045.

11. Makanyanga JC, Pendsé D, Dikaios N, et al. Evaluation of Crohn's disease activity: Initial validation of a magnetic resonance enterography global score (MEGS) against faecal calprotectin. Eur Radiol. 2014;24(2):277-287.

12. Tolan DJM, Greenhalgh R, Zealley I a, Halligan S, Taylor S a. MR enterographic manifestations of small bowel Crohn disease. Radiographics. 2010;30:367-384.

13. Menard S. Applied Logistic Regression Analysis: Sage University Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.; 1995.

14. Menys A, Taylor SA, Emmanuel A, et al. Global Small Bowel Motility: Assessment with Dynamic MR Imaging. Radiology. 2013.

15. Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology. 1989;96(3):804-810.

16. Menys A, Hamy V, Makanyanga J, et al. Dual registration of abdominal motion for motility assessment in free-breathing data sets acquired using dynamic MRI. Phys Med Biol. 2014;59(16):4603-4619.

17. Khalaf A, Nowak A, Menys A, et al. Cine MRI assessment of motility in the unprepared small bowel in the fasting and fed state: Beyond the breath-hold. Neurogastroenterol Motil. 2018.

18. Menys A, Plumb A, Atkinson D, Taylor SA. The challenge of segmental small bowel motility quantitation using MR enterography. Br J Radiol. 2014;87(1040).

19. de Jonge CS, Smout AJPM, Nederveen AJ, Stoker J. Evaluation of gastrointestinal motility with MRI: Advances, challenges and opportunities. Neurogastroenterol Motil. 2018;30(1).

Supplementary Material/Supporting Information

Supplementary Material/Supporting Information 1: HBI Component breakdown with description.

Supplementary Material/Supporting Information 2: Motility balanced sequence parameters for donor site 1 and 2.

Supplementary Material/Supporting Information 3: Layout of the viewer for the visual grading of the five metrics by the study radiologists. The “cine” movie would be viewed as a repeated loop and the sliding bars to the right-hand side would be used to grade on a 0-10 scale (% scale for area of motile bowel).

Supplementary Material/Supporting Information 4: Flow chart demonstrating patient inclusions and exclusions.

Supplementary Material/Supporting Information 5: Patient Demographics with 105 patients in study cohort.

Figure Legends

Figure 1: A reference image was selected automatically from the stack of dynamic MRI images or frames (step 1). Each frame was registered to the reference image (step 2) to produce a set of deformation fields (step 3). The SD Jacobian was calculated to create a motility map (step 4B). Mean motility (metric 1), spatial variation of motility (metric 2) and area of motile bowel (metric 4) were calculated from this motility map. A temporal variation map was created by calculating the variance of the sliding SD Jacobian values map (step 4A). The temporal variation of motility (metric 3) was calculated from the temporal variation map and the intestinal distension (metric 5) was calculated from the reference frame by thresholding intensities based on 50% of the median intensity within the ROI.

Figure 2: Examples of low (first column) and high values (second column) displayed for the five metrics of mean motility (A-B), spatial variation of motility (C-D), temporal variation of motility (E-F), area of motile bowel (G-H) and intestinal distension (I-J).

Figure 3: Fitted HBI model data vs. Actual HBI data for the best automated models with A: negative association of metric 2 (spatial variation) and B: negative association of metric 3 (temporal variation) + positive association of metric 4 (area of motile bowel).

Figure 4: The visual scoring for the area of motile bowel (A - top) and for temporal variation of motility (B - bottom) are displayed on simple correlation plots (left) and Bland-Altman plots (right). The highest inter-observer agreement is seen for the area of motile bowel

visual scoring (top right) where the coefficient of variation (CV) is 26% on the Bland-Altman plot. The lowest inter-observer agreement is seen for the temporal variation of motility visual scoring (bottom right) where the CV is 71%.

Figure 5: The visual scoring for the mean motility from radiologist 2 (A - top) and for temporal variation of motility from radiologist 1 (B - bottom) are displayed on simple correlation plots (left) and Bland-Altman plots (right). The best intra-observer agreement is seen for the mean motility visual scoring from radiologist 2 (top right) where the coefficient of variation (CV) is 16% on the Bland-Altman plot. The worst intra-observer agreement is seen for the temporal variation of motility visual scoring from radiologist 2 (bottom right) where the CV is 81%.