Development of a Magnetic Resonance Imaging Index for Fistulising Crohn's Disease

Running title: *MRI Index for Fistulising CD*

Authors: Mark A Samaan,[1, 2] Carl A J Puylaert,[3] Barrett G Levesque,[2, 4] Guangyong Zou,[2,] Larry Stitt,[2] Stuart A Taylor,[1] Lisa M Shackelton,[2] Margaret K Vandervoort,[2] Reena Khanna,[2,] Cynthia Santillan,[4] Jordi Rimola,[5] Pieter Hindryckx,[2, 6] C Yung Nio,[3] William J Sandborn,[2, 4] Geert D'Haens,[2, 3] Brian G Feagan,[2] Vipul Jairath,[2] Jaap Stoker[3]

Affiliations: [1]London, United Kingdom, [2]London, Canada, [3]Amsterdam, The Netherlands, [4]La Jolla, USA, [5]Barcelona, Spain, [6]Ghent, Belgium,

Correspondence to: Vipul Jairath, DPhil, MRCP**,** Western University**,** London, Ontario, Canada, Email: vjairath@uwo.ca

Key words: Perianal, fistula, MRI

Abbreviations: MRI, magnetic resonance imaging; CD, Crohn's disease; ICCs, intraclass correlation coefficients; TSE, turbo spin echo; VAS, visual analogue scale; IBD, inflammatory bowel disease

Word count: 4347

Background: Magnetic resonance imaging (MRI) is the gold standard for assessment of perianal fistulising Crohn's disease (CD). The Van Assche index is the most commonly used MRI fistula index.

Aims: Assess the reliability of the Van Assche index, modify the instrument to improve reliability and create a novel index for fistulizing CD.

Methods: A consensus process developed scoring conventions for exisiting Van Assche index component items and new items. Four experienced radiologists evaluated 50 MRI images in random order on three ocasions. Reliability was assessed by estimates of intraclass correlation coefficients (ICCs). Common sources of disagreement were identified and recommendations made to minimise disagreement. A mixed effects model used a 100 mm visual anologue scale (VAS) for global severity as outcome and component items as predictors to create a modified Van Assche index.

Results: Intraclass correlation coefficients (95% confidence intervals) for intra-rater reliability of the original and modified Van Assche indices and the VAS were 0.86 (0.81–0.90), 0.90 (0.86–0.93) and 0.86 (0.82–0.89). Corresponding ICCs for inter-rater reliability were 0.66 (0.52–0.76), 0.67 (0.55–0.75) and 0.58 (0.47–0.66). Sources of disagreement included number, location, and extension of fistula tracts, and rectal wall involvement. A modified Van Assche index (range 0–24) was created that included seven component items.

Conclusions: Whereas "almost perfect" intra-rater reliability was observed for the assessment of MRI images for fistulising CD using the Van Assche index, inter-rater reliability was considerably lower. Our modification of this index should result in a more optimal instrument.

INTRODUCTION

Approximately one-third of patients with Crohn's disease (CD) will develop a perianal fistula in their lifetime.[1,2] This relatively common disease complication is associated with significant morbidity and substantially impaired quality of life. Medical therapy is only partially effective, such that patients frequently undergo surgery to control symptoms and disease-related complications. Therefore, more effective treatment approaches are needed that will require evaluation in well-designed randomised controlled trials.

In this respect, the lack of a robust and validated outcome measure has constrained research in this area. In clinical practice, pelvic magnetic resonance imaging (MRI) has an established role for the evaluation of patients with perianal fistulas and is used to make clinical decisions and assess changes in disease state.[3,4] While several scoring systems, such as the Fistula Drainage Assessment and the partially validated Perianal Disease Activity Index, have been used to quantify clinical parameters, far fewer have been developed for MRI and they also lack standardised definitions of descriptor items.[5] The most frequently used MRI index is the Van Assche index, originally developed to fulfill the need for an instrument that could measure response of perianal fistulising CD to medical therapy.[6] The components of the index, based primarily upon radiological expertise and the classification of perianal fistulas by Parks *et al*, were shown to be reliable in the initial small (n=18) study.[7,6] In a subsequent small study, in addition to the original Van Assche index components Horsthuis *et al* included assessment of T1-weighted post-gadolinium hyperintensity and the presence of an infiltrate as potential measures of inflammatory activity. Although this study confirmed the Van Assche index was suitable for use as measure of response to therapy in clinical practice, the

validity of the individual index components was not assessed due to sample size limitations, and no conclusions could be made on the value of the additional inflammatory components.[8] Ng *et al* also found the total Van Assche index partially responsive to change with treatment with anti-tumor necrosis factor therapy, although they did not examine the operating properties of the individual index items and it was evident that the overall index was insensitive to change in some patients, including those with a 50%-80% reduction in track volume.[9]

In 2013, the World Congress of Gastroenterology working group formulated a multi-disciplinary consensus statement for classifying, diagnosing, and treating fistulising CD and identified the need for a validated index for measuring response to therapy as a high priority.[10] In the first steps towards this goal we: (1) developed standardized scoring conventions for existing components of the Van Assche index and assessed their reliability in a convenience sample of MRI examinations in patients with fistulising perianal CD; (2) identified items with highest disagreement in the Van Assche index, developed modifications to these and included new items through a formal consensus process, and then (3) created a modified Van Assche index.

MATERIALS AND METHODS

Study population

Fifty MRI scans carried out in 50 patients with active perianal fistulising CD between July 2011 and November 2013 at the Academic Medical Center, Amsterdam, The Netherlands were evaluated. The sample included a wide range of disease complexity and anatomical classification (see Supplementary **Table 1**). The indications for which the scans were performed included active perianal symptoms and evaluation of response to previous medical or surgical interventions. Scans with

missing sequences, and those from patients who did not receive intravenous contrast due to allergy or who had undergone ileoanal pouch formation for colonic CD were excluded (Supplementary **Figure 1**). Of the 50 scans, 45 were collected in a consecutive manner with a further 5 scans hand-selected to ensure that all of the fistula classifications were represented in our cohort. These additional scans were selected based on the description of the reporting radiologist for the purposes of clinical practice, and were not selected by the expert central readers and had not been previously reviewed by the readers for other purposes. These scans included intersphincteric, extrasphincteric, suprasphicteric, and two transsphincteric fistulas. One of these was considered complex. Ideally, more of the less commonly occurring fistula subtypes would have been included in our final set of scans but they, by their very nature, occur infrequently and were therefore difficult to identify.

MRI acquisition technique

MRI scans were performed on a 1.5-T scanner (Magnetom Avanto, Siemens, Erlangen, Germany) according to a standardized protocol for perianal fistulising disease, consisting of the following sequences: T2-weighted turbo spin echo (TSE) sequences in the sagittal, coronal and transverse plane, a fat-saturated T2-weighted TSE spectral adiabatic inversion recovery sequence and a post-contrast fat-saturated T1-weighted TSE sequence in the transverse plane (Supplementary **Table 2**). A combination of saturated and unsaturated T2 sequencing was used to discriminate between fibrosis, edema and fat.

The post contrast sequences were performed 60 seconds after intravenous administration of 0.1 mL/kg gadobutrol (Gadovist 0.1 mmol/mL, Bayer Schering Pharma, Berlin, Germany). The coronal and axial sequences were parallel and orthogonal to the anal sphincter axis, respectively. A spasmolytic agent (Buscopan,

Boehringer Ingelheim, Ingelheim, Germany) was used to reduce bowel peristalsis and motion artifacts. All MRI examinations were anonymized and uploaded to a secure central viewing system.

Scoring conventions and modification to the original Van Assche index

Based upon face validity and expert opinion of radiologists in this field (CS, CYN, JR, JS, SAT), we developed standardized scoring conventions for component items of the original Van Assche index, and included additional descriptors to these items as modifications. We also incorporated novel items considered potentially important based on face validity that might also be responsive to change after a therapeutic intervention through the same expert opinion process. A pilot study followed whereby four expert abdominal radiologists (CS, CYN, JR, SAT) with experience in fistula MRI evaluated four scans (that were not subsequently included in the 50 study scans), before refining the items/descriptors during subsequent discussion, which took place prior to initiation of the reliability study. These discussions were conducted in a manner consistent with a Delphi process although no formal voting was carried out. The adapted version of the Van Assche index is henceforth referred to as the modified Van Assche index.

Reliability study

Four expert abdominal radiologists (CS, CYN, JR, SAT) with experience in fistula MRI and blinded to clinical information independently reviewed 50 MRI examinations in triplicate in random order and assessed disease activity using both the original and the modified Van Assche index (**Table 1**).

The study radiologists also completed three visual analogue scale (VAS) assessments for each scan as part of their evaluations. The first was aimed at evaluating the overall severity of the inflammatory component of the perianal disease

(0 mm = no active inflammatory disease, 100 mm = severe active inflammatory disease). The second assessed the overall complexity (0 mm = minimal complexity, 100 mm = highly complex). The final VAS assessment was a combined global assessment that took account of both inflammatory activity and complexity (0 mm = no perianal disease, 100mm = worst perianal disease encountered), and was the score used as the gold standard comparator. In addition, the presence or absence of imaging artifacts was recorded, as was image quality, using a three-point scale (good, adequate, inadequate).

Intra-and inter-rater reliability statistics were calculated and compared for the total index and the component items of the original and modified versions of the Van Assche index and the three features assessed with the VAS. Index components with ”fair” or ”poor” inter-rater reliability based on the criteria of Landis and Koch, whereby intraclass correlation coefficients (ICCs) of <0.0, 0.0–0.20, 0.21–0.4, 0.41–0.6, 0.61-0.8, and >0.81 constitute ‘poor,’ ‘slight,’ ‘fair,’ ‘moderate,’ ‘substantial’ and ‘almost perfect’ reliability, respectively, were subsequently discussed during a second consensus meeting of the study radiologists. Consensus statements were generated during this meeting, and RAND appropriateness methodology was used to refine items accordingly.[11] RAND appropriateness methodology uses a modified Delphi panel approach to combine the best available evidence and personal clinical experience of experts. The panel facilitates decision making through an iterative process in which questions are raised and viewpoints discussed. Experts then vote on the appropriateness of statements developed during this process with an aim to reach consensus according to predefined criteria. As part of the process, rules were developed that would improve consistency of reading of MRI images of perianal fistulas. Other potential sources of disagreement among the radiologists were

explored by identifying “outlier” images that led to the poorest inter-rater reliability. These outlier images were reviewed by one expert reader (JS) for potential common sources of variance.

Statistical methods

Clinical characteristics were assessed using descriptive statistics. Intra- and inter-rater reliability were quantified using intraclass correlation coefficients (ICCs), which are equivalent to weighted Kappa statistics in the case of ordinal data.[12] For each component item, point estimates of intra- and inter-rater ICCs were concurrently estimated using a two-way random effects model with interaction.[13] Associated two-sided 95% CIs were obtained using the non-parametric percentile bootstrap method with 2000 samples obtained with replacement at the level of the image to maintain data structure. This approach is commonly known as the cluster bootstrap method.[14] The degree of reliability was interpreted based on Landis and Koch benchmarks. These empirical benchmarks were originally developed for grading kappa statistics and have now become widely adopted for assessment of ICCs.

The modified Van Assche index was created using the VAS for global assessment of severity as the outcome criterion. To account for the data structure that each MRI scan was read three times by each of the four readers, a mixed effects model was adopted with fixed effects including the nine modified Van Assche items and random effects including MRI scan, reader and their interaction. The model fit was assessed using residual diagnostics. The Bayesian Information Criterion (BIC) using maximum likelihood estimation was used to assess the quality of fit of the model. Regression coefficients were standardized by dividing by the smallest coefficient to facilitate easy calculation of the modified Van Assche index. Sample size was estimated using the method proposed by Zou.[15] Assuming a true

ICC of 0.75, rating of 50 MRI sequences from 50 patients by 4 radiologists would yield an 92% chance of obtaining a two-sided 95% lower confidence bound for the ICC of 0.55, a value considered to be ‘moderate’ according to the Landis and Koch benchmarks.

Ethical considerations

The use of the scans for the purposes of this study was approved by the medical ethics committee at the Academic Medical Centre and additional patient consent for use of images was not required.

RESULTS

Modified Van Assche index

Scoring conventions, modifications, and addition of novel items to the original Van Assche index based upon the expert consensus to create the modified Van Assche index are shown in **Table 1**. Core items from the Van Assche index were all retained, although the “Collections” item was modified and incorporated into a new item “Inflammatory mass” as described in **Table 1** and below. Modifications to the original Van Assche index included: (1) the addition of “submucosal” to the location component, as well as creation of separate categories for extra- and intersphincteric locations; (2) addition of “horseshoe configuration” to the extension component; (3) inclusion of extensions in addition to the primary tract in the assessment of the hyperintensity of T2-weighted images; (4) addition of “increased signal intensity” to the rectal wall involvement component. New items added to the Van Assche index were: (1) presence of a recto/anovaginal tract; (2) presence of an inflammatory mass in conjunction with assessment of size of collections; (3) hyperintensity of the primary tract or extensions on post-contrast fat saturated T1-weighted images; (4) assessment of the dominant feature of the primary tract and extensions.

Participant characteristics

Participant demographics are outlined in Supplementary **Table 3**. The mean age of the patients was 39 (range, 19-75 years), 48% (24/50) were male, the mean CD disease duration was 14 years (range, 0.5-47 years) and the mean duration from first fistula diagnosis was 9 years (range, 1 month-39 years). Most participants had prior perianal surgery (66%), and more than half (56%) were under treatment with an anti-tumour necrosis factor (TNF) agent. The mean overall assessment of disease severity based on the VAS was 4.36 (range 0.20 to 9.80).

MRI image quality

A total of 600 reads were performed by the four expert radiologists. Of these, 565 (94.2%) were considered to be of good quality, 30 (5.0%) were adequate and 5 (0.8%) were considered inadequate. Only 2.2% of reads reported a missing sequence or plane and imaging artefact was present in 5%. Four cases contained ano-vaginal fistulas.

Reliability results for overall indices

Intraclass correlation coefficients and 95% confidence intervals for intra- and inter-rater reliability for scoring of the VAS and the total original and modified Van Assche indices and their components are shown in **Table 2**.

While almost perfect intra-rater reliability was observed for all of the VAS-based assessments, inter-rater reliability ranged from moderate to substantial. The lowest inter-rater ICC was observed for assessment of complexity of disease (0.41, 95% CI 0.28-0.51), followed by global assessment (0.58, 95% CI 0.47-0.66) and severity of inflammatory perianal disease activity (0.64, 95% CI 0.53-0.72). Intra- and

inter-rater reliability for the original Van Assche index were almost perfect (ICC 0.86, 95% CI 0.81-0.90) and substantial (0.66, 95% CI 0.52-0.76). Intra- and inter-rater reliability for the modified Van Assche index) were both numerically higher than for the original Van Assche index, although the degree of intra- and inter-rater reliability according to the Landis and Koch criteria was unchanged and remained almost perfect (0.90, 95% CI 0.86-0.93) and substantial (0.67, 95% CI, 0.55-0.75), respectively.

Reliability of index component items

Intra-rater reliability was substantial to almost perfect for all component items of the original Van Assche index except for fistula location which was moderate (0.59, 95% CI 0.48-0.70) prior to application of standardized scoring conventions adapted from the St Mark's classification, and which improved to substantial after application of the conventions (0.70, 95% CI 0.59-0.79) (**Table 2**).

Inter-rater reliability was slight to moderate for all component items of the original Van Assche index, with the exception of collections, which was substantial (0.61, 95% CI 0.46-0.73). Inter-rater reliability for the assessment of fistula location improved from fair to moderate (0.30, 95% CI 0.18-0.41 vs. 0.46, 95% CI 0.29-0.59 respectively) and fistula extension improved from slight to moderate (0.17, 95% CI 0.06-0.28 vs. 0.48, 95% CI 0.33-0.61 respectively) (**Table 2** and Supplementary **Table 4**) with the modified version of the Van Assche index. Inter-rater reliability for the assessment of hyperintensity of tracts on fat-saturated T2-weighted images remained unchanged, and decreased for the assessment of rectal wall involvement with the modified van Assche index despite modification and inclusion of standardized scoring conventions.

Four new items were incorporated into the modified van Assche index based upon the expert consensus, specifically: (1) presence of a recto/anovaginal tract; (2) inflammatory mass (which incorporated the item of collections from the original van Assche index plus additional features); (3) hyperintensity of the primary tract or extensions on post-contrast fat saturation T1-weighted images; and (4) dominant feature of the primary tract and extensions. Assessment of these new items was associated with slight-to-moderate inter-rater reliability (**Table 2** and Supplementary **Table 4**), with the lowest ICCs observed for the presence of an ano/rectovaginal tract (0.15, 95% CI 0.02-0.25). Higher inter-rater ICCs were observed for assessment of hyperintensity of the primary tract or extensions seen on post-contrast fat saturated T1-weighted images (0.40, 95% CI 0.27-0.50) and assessment of the dominant feature of the primary tract and extensions (0.37, 95% CI 0.23-0.49). The highest inter-rater ICC was observed for inflammatory mass (0.59, 95% CI 0.43-0.71).

Sources of disagreement and consensus process

The consensus process, which involved participation of four experts (CAJP, ST, JR, JS), a gastroenterologist (GDH) and a moderator (VJ) was conducted using RAND appropriateness methodology to review items contributing to the greatest variance in an attempt to understand and minimise the sources and improve inter-rater reliability. In addition, nine MRI scans responsible for the greatest disagreement based on their effect on the overall estimation of ICCs were reviewed by a single expert reader (JS) to identify features that may have contributed to disagreement. As a result of this process, modifications were made to four specific items. Subsequently, readers voted on the appropriateness of the modified items (Supplementary **Table 5**). The

recommendations of the experts and the proposed modified items are described in **Table 3** and Supplementary **Table 6**.

Final modified van Assche index based on mixed effects modeling

The mixed effect model building process began with all 9 items described in **Table 1**. Coefficient estimates and associated inferential statistics for these items are shown in **Table 4.** Negligible and statistically insignificant coefficients were observed for number of fistula tracts and location, and presence of a recto/anovaginal tract and hyperintensity on post contrast T1-weighted images were only marginally significant ($p = 0.053$ and 0.051, respectively). The model was then refitted with the remaining five component items, resulting in numerically large and statistically significant coefficients (**Table 4**). The final modified Van Assche index based on mixed effects modeling included “extension,” “hyperintensity on T2-weighted images,” “rectal wall invovlement,” “inflammatory mass,” and “dominant feature of primary tract and extension”. Intra-class correlation coefficients for intra- and inter-rater reliability of the modified Van Assche index based on the model were 0.89 (0.85 to 0.92) and 0.67 (95% CI 0.56 to 0.75), consistent with nearly perfect and “substantial” reliability, respectively. An overall score for this index is calculated as the sum of the products of the observed item scores and the corresponding standardized scores. Total scores range from 0 to 20.

DISCUSSION

Development of novel therapies for patients with perianal fistulising disease is a large unmet need in the management of CD. There are currently no therapies approved by the US Food and Drug Administration, although promising new

approaches to treatment are entering phase 3 programs.[16] In clinical practice, MRI has an essential role in the assessment of perianal disease and thus development of a fully validated MRI-based evaluative instrument with well-defined operating properties is essential for both for the evaluation of clinical disease activity and for the efficient study of new treatments in clinical trials. Besides the Van Assche index, to our knowledge the only other previously described MRI index for assessment of the severity of perianal disease is the index described by Ng and colleagues.[9] In their study, as well as highlighting that the van Assche index was insensitive to change in some patients in whom there was radiological evidence of fistula improvement, Ng et al also semi-quantitatively described within-patient fistula changes in response to therapy as “healed (absence of high-signal tracks on fat saturated T2 sequences)” or “improved,” “unchanged,” or “worse” in cases where fistula tracks remained visible. Both indices have been used to define endpoints in clinical trials despite the lack of a complete characterization of their operating properties or standardized methods for scoring. These limitations informed the need for our study. A critical step in index development is to determine the reliability of index component items, which is defined as the extent to which raters are able to consistently distinguish between study subjects and the degree to which repeated measurements provide similar results.[17]

We found “almost perfect” intra-rater reliability for the original Van Assche index, and “substantial” reliability for all of the individual component items with the exception of fistula location. In contrast, the corresponding inter-rater reliability was lower, and ranged from slight to substantial, with the lowest reliability observed for assessment of fistula extension (0.17, 95% CI 0.06-0.28) and the best reliability observed for assessment of collections (0.61, 95% CI 0.46-0.73). This discordance

between intra- and inter-rater reliability is not unexpected as raters are more likely to agree with themselves than one another, however it provides an opportunity to improve inter-rater item scoring through a systematic consensus process when substantial intra-rater reliability exists. This is an important consideration if multiple readers are required out of logistical considerations in the conduct of a clinical trial or in clinical practice when different radiologists read different scans originating from the same patient. Accordingly, we identified individual items in the original Van Assche index with suboptimal inter-rater performance and attempted to improve them by developing and including standardized scoring conventions and modifications through a systematic consensus process. Specifically, the inter-rater reliability of “location” and “extension,” were substantially improved (see **Tables 1** and **2**). For assessment of fistula extension, the experts felt that definition of the ano-rectal junction was a potential source of disagreement among readers. Two cases with considerable variability for the identification of supra- versus infralevatoric extensions are shown in **Figures 1** and **2**. To provide superior differentiation between these extensions, we defined the ano-rectal junction as the connection of the levator plane to the ano-rectum. This anatomical location is delineated by a two arrows in **Figure 3**.

Despite the application of scoring conventions and inclusion of an additional item (increased signal intensity), ICCs for inter-rater reliability for the assessment of rectal wall involvement decreased marginally in the modified Van Assche index. Poor rectal distension and subjectivity in discerning signal intensity relative to surrounding planes were potential sources of disagreement considered by the experts. Although this item remained significant in the mixed effects modelling approach, it was re-named as “absence or presence of proctitis” and re-defined based on the expert

RAND consensus to improve reliability (Supplementary **Table 6)**. The proctitis item is a composite score of several features that had fair inter-rater reliability in this study. However, improved inter-rater reliability (with the exception of fair-to-moderate inter-rater reliability for perimural enhancement) was observed in a separate recently published study on individual MRI features in Crohn's proctitis that were significantly correlated with endoscopy. As the inter-rater reliability of this item was not tested in this study, it may be worthwhile to further explore a more detailed evaluation of proctitis.[18] No standardized definitions or modifications were initially applied to the item "number of fistula tracts," since this assessment was considered to be self-explanatory. It was thus somewhat unexpected that only "fair" inter-rater reliability was observed for this item. We subsequently modified this item to include the descriptors "single, unbranched" or "complex" based on expert opinion regarding the difficulty associated with differentiating a single tract with multiple braches, from disease with multiple tracts, although this item was not statistically significant in the mixed effects model building, and was not included as an item in the final index.

Inter-rater reliability for the novel items included in the modified Van Assche index ranged from slight to moderate. Modifications to further improve reliability are shown in Supplementary **Table 6**. Although there was no expert consensus on the appropriateness of the inclusion of the absence or presence of an ano/rectovaginal this item was not statistically significant in the mixed effects model building and was excluded from the final index. The relatively poorer inter-rater reliability for this item was thought to be likely due to the difficulty associated with identifying these tracts, which are typically short in length, small in diameter and obscured between vessels. Free drainage to the vagina and rectum, and lack of fluid within the tract were also thought to contribute to difficulty in identification of these tracts.

Limitations of our study should be acknowledged. Firstly, images for the study were acquired from a tertiary center, which is a regional center for complex cases. Accordingly, some of the MRI's reviewed were considered to represent severe disease, where complexity is likely to be higher than the general IBD population. However, the MRI scanners, sequences and protocols used in this study are consistent with those used in most centers. This did not include diffusion weighted imaging and its potential added value to the protocol performed in this study has not yet been clearly demonstrated. Secondly, the scans were carried out for a range of indications and included patients who were receiving various treatments, including anti-TNF agents. It is possible that these factors could have introduced bias (for example, attenuation of inflammatory activity due to treatment effect). Conversely, this heterogeneity may also serve to increase the generalizability of our results as many MRI assessments are performed in patients receiving anti-TNF therapy. Thirdly, whilst the quality of the majority of MRI images was judged to be acceptable, the method by which we evaluated quality (good, adequate or inadequate) was subjective and did not include standardized definitions. Fourthly, readers were highly experienced IBD radiologists and thus the results may not be generalizable to other readers; however since the objective of this exercise was to refine operating characteristics of an existing MRI fistula scoring system for use in clinical trials, this situation mandated the need for specialist radiologists. Fifthly, we acknowledge that the number of scans included was modest (n=50), however, the total number of reads was substantial (n=600), owing to the reading of scans in triplicate by all four radiologists which was, by design, based upon formal statistical methods for a reliability study. Finally, we were unable to test the responsiveness of the new index, which would require comparing baseline and post-treatment MRI's in patients

receiving a treatment of known efficacy, ideally in a randomised, placebo-controlled trial. Once such a dataset becomes available, the responsiveness of the modified van Assche index can be tested and a minimally important score change identified. Furthermore, it has not escaped our attention that MRI-defined disease activity may be an important prognostic indicator since it provides relevant information about deep-tissue inflammation and healing and may provide important information about progressive structural damage. Nevertheless, several questions remain about the optimal timing for MRI assessment, with preliminary data indicating that this may be considerably longer than traditional clinical or endoscopic follow-up periods.[9]

In conclusion, we found “substantial” to “almost perfect” intra-rater reliability amongst radiologists in the assessment of fistulising perianal CD using the modified Van Assche index, but only “slight” to “moderate” inter-rater reliability. Standardized scoring definitions and modifications to the van Assche index developed through a consensus process enabled us to characterize the greatest sources of disagreement, thus generating recommendations that may improve inter-rater reliability. The modified van Assche index derived through a mixed effects modelling approach has the potential for use in routine clinical practice to more consistently assess changes in disease acitivity, as well within clinical trials to evaluate new therapies.

Table 1. The original and modified Van Assche items

Descriptor	Original Index[1]	Item Weight	Modified Index	Item Weight	Definition developed through initial consensus[2]
Number of fistula tracts	○ None	0	○ None	0	
	○ Single, unbranched	1	○ Single, unbranched	1	
	○ Single, branched	2	○ Single, branched	2	
	○ Multiple	3	○ Multiple	3	
Location[3]			○ Submucosal	0	Tract lies superficial to the internal sphincter
			○ Intersphincteric	1	Tract extends through the internal sphincter to the intersphincteric plane then to the perineal skin
	○ Transsphincteric	2	○ Transsphincteric	2	Tract extends via the internal and external anal sphincter (or puborectalis muscle) into the ischioanal fossa then to the perineal skin
	○ Extra- or intersphincteric	1	○ Extrasphincteric	3	Tract extends through the ischioanal fossa, upwards and through the levator ani muscles to the rectal wall completely outside the sphincter mechanism
	○ Suprasphincteric	3	○ Suprasphincteric	4	Tract extends via intersphincteric space, then tracts superiorly to above the puborectalis muscle (ie, above the anorectal junction) before curving downward through the levator muscle lateral to the external anal sphincter and

					puborectalis muscle into the ischioanal fossa then to the perineal skin
Extension[4]			○ Absent	0	
	○ Infralevatoric	1	○ Infralevatoric	1	Extends upward in the ischioanal fossa but remains below the levator ani muscle
			○ Horseshoe configuration	2	Extends into the intersphincteric space on both sides of the midline
	○ Supralevatoric	2	○ Supralevatoric	3	Extends upward in the intersphincteric plane and over the top of the levator ani muscle
Hyperintensity on T2-weighted images[5]	○ Absent	0	○ Absent	0	No hyperintensity visible, only scar tissue
	○ Mild	4	○ Mild	1	Slight increase in signal intensity but less than nearby, in- plane vessels
	○ Pronounced	8	○ Pronounced	2	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels
Collections (cavities > 3 mm diameter)	○ Absent	0			
	○ Present	4			
Rectal wall involvement	○ Normal	0	○ Normal	0	Normal appearance of rectal wall
	○ Thickened	2	○ Thickened	1	Thickened rectal wall (eg, > 3 mm when distended)

	○ Increased signal intensity	2	Hyperintensity of the rectal wall on fat saturated T2-weighted images (compared to nearby, in-plane vessels), mural stratification and/or perimural infiltrate
Presence of a recto/anovaginal tract	○ Absent	0	No recto/anovaginal tract
	○ Rectovaginal tract	1	Fistula arises from rectal mucosa
	○ Anovaginal tract	2	Fistula arises from anal mucosa
Inflammatory mass[4]	○ Absent	0	No inflammatory mass
	○ Diffuse	1	Diffuse inflammation of surrounding tissues
	○ Focal	2	Lesion > 3 mm in diameter on T2-weighted images (but does not include linear tracts with diameter > 3mm) with diffuse enhancement on T1-weighted post contrast images (ie, granulation tissue)
	○ Collection-small	3	Circumscribed cavity 3-10 mm in diameter (but does not include linear tracts with diameter > 3 mm). Hyperintense appearance on fat saturated T2-weightedimages with rim enhancement on T1-weighted post-contrast images
	○ Collection-medium	4	As defined above except diameter measures 11-20 mm

	○ Collection-large	5	As defined above except diameter measures >20 mm
Hyperintensity of primary tract or extensions on post-contrast T1-weighted images[5]	○ Absent	0	No hyperintensity visible
	○ Mild	1	Slight increase in signal intensity but less than nearby, in- plane vessels
	○ Pronounced	2	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels
Dominant feature of primary tract and extensions	○ Predominantly fibrous	0	> 50% of tract has a fibrotic appearance (ie, hypointense on fat saturated T2-weighted images)
	○ Predominantly filled with granulation tissue	1	> 50% of tract is filled with granulation tissue (ie, hyperintense on fat saturated T2-weighted images with enhancement of contents and wall on T1-weighted post-contrast images)
	○ Predominantly filled with fluid or pus	2	> 50% of tract is filled with fluid or pus (ie, hyperintense on fat saturated T2-weighted images with no enhancement of contents on fat saturated T1-weighted post-contrast images [though lining of tract may enhance])

[1]The original van Assche index consists of six anatomical and (weighted) inflammatory disease parameters including assessment of; the numbers of fistula tracts (0 – 3), the location of fistulas (1 – 3), extension of fistulas (1 or 2), hyperintensity on T2-weighted images (0, 4, or 8), presence of collections (defined as cavities >3 mm in diameter) (0 or 4) and rectal wall involvement (score 0 or 2). The total score ranges from 0 to 22; [2]Definitions were applied to modified index only; [3]Adapted from the St. Mark's Classification

in the modified index (the most dominant feature is assessed for both the original and modified indices); [4]For the modified index, all relevant findings are identified; the highest score is chosen; [5]For the modified index, extensions were also assessed and the most severe lesion was rated by comparing signal intensity with nearby, in-plane vessels in the modified index; images were fat saturated

Table 2. Reliability* of the VAS, original and modified van Assche indices.

	Reliability (95% CI)	
	Intra-rater ICC	**Inter-rater ICC**
VAS		
Severity of inflammatory perianal disease activity	0.82 (0.75, 0.86)	0.64 (0.53, 0.72)
Complexity of perianal disease	0.86 (0.81, 0.89)	0.41 (0.28, 0.51)
Global assessment (combined inflammatory activity and complexity)	0.86 (0.82, 0.89)	0.58 (0.47, 0.66)
Original Van Assche Index		
Number of fistula tracts	0.74 (0.65, 0.82)	0.35 (0.20, 0.51)
Location	0.59 (0.48, 0.70)	0.30 (0.18, 0.41)
Extension	0.77 (0.68, 0.85)	0.17 (0.06, 0.28)
Hyperintensity on T2-weighted images	0.70 (0.59, 0.78)	0.54 (0.42, 0.64)
Collections (cavities > 3 mm diameter)	0.80 (0.71, 0.88)	0.61 (0.46, 0.73)
Rectal wall involvement	0.71 (0.63, 0.77)	0.27 (0.17, 0.37)
Total score	0.86 (0.81, 0.90)	0.66 (0.52, 0.76)
Modified Van Assche Index		
Number of fistula tracts	0.74 (0.65, 0.82)	0.35 (0.20, 0.51)
Location (according to St Mark's classification)	0.70 (0.59, 0.79)	0.46 (0.29, 0.59)
Extension	0.77 (0.68, 0.84)	0.48 (0.33, 0.61)
Hyperintensity on T2-weighted images	0.70 (0.58, 0.78)	0.54 (0.42, 0.63)
Rectal wall involvement	0.71 (0.62, 0.78)	0.22 (0.13, 0.31)
Presence of a recto/anovaginal tract	0.85 (0.71, 0.94)	0.15 (0.02, 0.25)
Inflammatory mass	0.84 (0.77, 0.89)	0.59 (0.43, 0.71)
Hyperintensity on post contrast T1-weighted images	0.72 (0.64, 0.79)	0.40 (0.27, 0.50)
Dominant feature of primary tract & extensions	0.74 (0.65, 0.80)	0.37 (0.23, 0.49)
Total score (simple sum)	0.90 (0.86, 0.93)	0.67 (0.55, 0.75)

*****Intra- and inter-rater reliability was interpreted using benchmarks described by Landis and Koch, whereby intraclass correlation coefficients (ICCs) of <0.0, 0.0–0.20, 0.21–0.4, 0.41–0.6, 0.61-0.8, and >0.81 constitute 'poor,' 'slight,' 'fair,' 'moderate,' 'substantial' and 'almost perfect' reliability, respectively.

Table 3. Items and definitions for the modified van Assche index based on the RAND consensus

Descriptor (score the most severe)	Modified Index	Definition developed through RAND consensus
Number of fistula tracts	○ Single, unbranched	
	○ Complex	Single branched tract or multiple tracts
Location	○ Submucosal	Tract lies superficial to the internal sphincter
	○ Intersphincteric	Tract extends through the internal sphincter to the intersphincteric plane then to the perineal skin
	○ Transsphincteric	Tract extends via the internal and external anal sphincter (or puborectalis muscle) into the ischioanal fossa then to the perineal skin
	○ Extrasphincteric	Tract extends through the ischioanal fossa, upwards and through the levator ani muscles to the rectal wall completely outside the sphincter mechanism
	○ Suprasphincteric	Tract extends via intersphincteric space, then tracts superiorly to above the puborectalis muscle (ie, above the anorectal junction) before curving downward through the levator muscle lateral to the external anal sphincter and puborectalis muscle into the ischioanal fossa then to the perineal skin
Extension	○ Absent	No extension
	○ Infralevatoric	Extends upward in the ischioanal fossa but remains below the levator ani muscle
	○ Supralevatoric	Any extension in the supralevatoric space (ie, above where the levator plate is connected to the anorectum)

	○ Horseshoe configuration	Extends into the intersphincteric space on both sides of the midline
Hyperintensity on T2-weighted images[1]	○ Absent	No hyperintensity visible, only scar tissue
	○ Mild	Slight increase in signal intensity but less than nearby, in- plane vessels
	○ Pronounced	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels
Proctitis*	○ Absent	Normal appearance of rectal wall
	○ Present	Increased wall thickness and size of mesorectal lymph nodes (> 5mm), creeping fat, increased perimural T2 signal and enhancement
Presence of a recto/anovaginal tract[2]	○ Absent	No recto/anovaginal tract
	○ Rectovaginal tract	Fistula arises from rectal mucosa
	○ Anovaginal tract	Fistula arises from anal mucosa
Inflammatory mass[2]	○ Absent	No inflammatory mass
	○ Diffuse	Diffuse inflammation of surrounding tissues
	○ Focal	Lesion > 3 mm in diameter on T2-weighted images (but does not include linear tracts with diameter > 3mm) with diffuse enhancement on T1-weighted post contrast images (ie, granulation tissue)
	○ Collection-small	Circumscribed cavity 3-10 mm in diameter (but does not include linear tracts with diameter > 3 mm and if present they should be excluded from the measurement of the size of the infiltrate[3]). Hyperintense appearance on fat saturated T2-weighted images with

		enhancement limited to the rim on T1-weighted post-contrast images
	○ Collection-medium	As defined above except diameter measures 11-20 mm
	○ Collection-large	As defined above except diameter measures >20 mm
Hyperintensity of primary tract or extensions on post-contrast T1-weighted images	○ Absent	No hyperintensity visible
	○ Mild	Slight increase in signal intensity but less than nearby, in- plane vessels
	○ Pronounced	Tract showing equal or greater signal hyperintensity than nearby in-plane vessels
Dominant feature of primary tract and extensions	○ Predominantly fibrous	> 50% of tract has a fibrotic appearance (ie, hypointense on fat saturated T2-weighted images)
	○ Predominantly filled with granulation tissue	> 50% of tract is filled with granulation tissue (ie, hyperintense on fat saturated T2-weighted images with enhancement of contents and wall on T1-weighted post-contrast images)
	○ Predominantly filled with fluid or pus	> 50% of tract is filled with fluid or pus (ie, hyperintense on fat saturated T2-weighted images with no enhancement of contents on fat saturated T1-weighted post-contrast images [though lining of tract may enhance])

*This item and definition should replace rectal wall involvement (see Table 1 for original item and definition).
[1]Primary tract and/or extensions; [2]Mark all that apply and measure fluid collection only (excluding wall) on the shortest axis
[3]Infiltrate describes an inflammatory region as a whole that may or may not include one or more fluid collections.

Table 4. Coefficients (standard error) of the items of the modified Van Assche index based on mixed-effects models

	Full model	P value	Reduced model	P value	Standardized score[1]
Number of fistula tracts	0.096 (0.092)	0.30	N/A		
Location (according to St Mark's classification)	0.032 (0.118)	0.78	N/A		
Extension	0.354 (0.061)	<0.001	0.367 (0.060)	<0.001	1.5
Hyperintensity on T2-weighted images	0.433 (0.155)	0.005	0.566 (0.137)	<0.001	2.3
Rectal wall involvement	0.227 (0.080)	0.005	0.250 (0.079)	0.002	1.0
Presence of a recto/anovaginal tract	0.261 (0.135)	0.053	N/A		
Inflammatory mass	0.289 (0.053)	<0.001	0.291 (0.053)	<0.001	1.2
Hyperintensity on post contrast T1-weighted images	0.308 (0.157)	0.051	N/A		
Dominant feature of primary tract & extension	0.320 (0.145)	0.027	0.284 (0.143)	0.048	1.2
Bayesian Information Criterion[2]	1969.7		1954.0		

[1]The regression coefficient for each item was standardized by dividing by the smallest coefficient and rounding to allow simple calculation.

[2]The Bayesian Information Criterion assesses the quality of fit of the model. Lower values for the reduced model indicate superior fit.

REFERENCES

1. Van Assche G, Dignass A, Reinisch W, et al. The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Special situations. *J Crohns Colitis* 2010;**4**:63-101.
2. Sciaudone G, Di Stazio C, Limongelli P, et al. Treatment of complex perianal fistulas in Crohn disease: infliximab, surgery or combined approach. *Can J Surg* 2010;**53**:299-304.
3. Panes J, Bouhnik Y, Reinisch W, et al. Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. *J Crohns Colitis* 2013;**7**:556-85.
4. Ong EM, Ghazi LJ, Schwartz DA, et al. Crohn's & Colitis Foundation of America, Inc. Guidelines for imaging of Crohn's perianal fistulizing disease. *Inflamm Bowel Dis* 2015;**21**:731-6.
5. Irvine EJ. Usual therapy improves perianal Crohn's disease as measured by a new disease activity index. McMaster IBD Study Group. *J Clin Gastroenterol* 1995;**20**:27-32.
6. Van Assche G, Vanbeckevoort D, Bielen D, et al. Magnetic resonance imaging of the effects of infliximab on perianal fistulizing Crohn's disease. *Am J Gastroenterol* 2003;**98**:332-9.
7. Parks AG, Gordon PH, Hardcastle JD. A classification of fistula-in-ano. *Br J Surg* 1976;**63**:1-12.
8. Horsthuis K, Bipat S, Spijkerboer AM, et al. Evaluation of an MRI-based score of disease activity in perianal fistulizing Crohn’s disease. *Clin Imaging* 2011;**35**:360–5.

9. Ng SC, Plamondon S, Gupta A, et al. Prospective evaluation of anti-tumor necrosis factor therapy guided by magnetic resonance imaging for Crohn's perineal fistulas. *Am J Gastroenterol* 2009;**104**:2973-86.
10. Gecse KB, Bemelman W, Kamm MA, et al. A global consensus on the classification, diagnosis and multidisciplinary treatment of perianal fistulising Crohn's disease. *Gut* 2014;**63**:1381-92.
11. Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, SR Moore SR, Siegel RA, eds. Clinical practice guideline development: methodology perspectives. Rockville, MD: US Department of Health and Human Services, 1994:59–70.
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reproducibility. *Psychol Bull* 1979;**86**:420-8.
13. Eliasziw M, Young SL, Woodbury MG, et al. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Physical Therapy* 1994;**74**:777-8.
14. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge University Press, 1997:100–2.
15. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med* 2012;**31**:3972-81.
16. Panés J, García-Olmo D, Van Assche G, et al. Expanded allogeneic adipose-derived mesenchymal stem cells (Cx601) for complex perianal fistulas in Crohn's disease: a phase 3 randomised, double-blind controlled trial. *Lancet* Published Online First 28 Jul 2016. doi: 10.1016/S0140-6736(16)31203-X.
17. de Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;**59**:1033–9.

18. Tutein Nolthenius CJ, Bipat S, Mearadji B, et al. MRI characteristics of proctitis in Crohn's disease on perianal MRI. *Abdom Radiol (NY)* 2016 **41**:1918–30.

Figure Legends

Figure 1. Transsphincteric track with extensive scar tissue.

Panels A and B represent consecutive coronal oblique T2-weighted images showing a transsphincteric track (arrow) with extensive scar tissue (*). The scar tissue extends high up (open arrow) in the ischioanal space and has a superior extension (open arrow) to the pelvic sidewall. The configuration of the superior extending scar tissue might be misinterpreted as the levator plate while the anorectal junction might be considered to be just below the level of the transsphincteric course of the track. In fact the levator plate is superior to the scar tissue (angulated arrows) as is the anorectal junction (arrow head). This misinterpretation may have led to identification of a supralevatoric extension for this fistula.

Figure 2. Transphinteric track with limited supralevatoric extension

Coronal oblique (panel A) and sagittal (panel B) T2-weighted images showing a transsphincteric track (arrow) extending just above (open arrow) the levator plate (L), the latter resulting in classifying it as (albeit limited) supralevatoric extension of the track. The limited extension above the levator plate and anorectal junction (arrow head) with the track directly adjacent to the levator plate might have led to different interpretations by readers.

Figure 3. The ano-rectal junction.

To provide superior differentiation between supra- and infralevatoric extensions, this junction is now defined as the connection of the levator plane to the ano-rectum and is delineated by two arrows in this coronal oblique T2-weighted image.

Authorship statement

(i) **Guarantor of the work:** VJ

(ii) **Specific author contributions:** MAS, CAJP, BGL, GYZ, MKV, BGF, VJ, JS played a role in planning and/or conducting the study; GYZ, LS played a role in the statistical analysis of the study data; MAS, CAJP, GYZ, LS, SAT, LMS, CS, JR, CYN, VJ, JS played a role in collecting and/or interpreting data, MAS, LMS, VJ played a role in drafting the manuscript. MAS, CAJP, BGL, GYZ, LS, SAT, LMS, MKV, RK, CS, JR, PH, CYN, WJS, GDH, BGF, VJ, JS played a role in reviewing and revising the manuscript for important intellectual content.

(iii) **All authors approved the final version of the manuscript.**

STATEMENT OF INTERESTS

1. Authors' declaration of personal interests

MAS has nothing to disclose.

CAJP has nothing to disclose.

BGL has received consulting fees from AbbVie, Takeda, Nestle Health Sciences, and Prometheus Labs.

GYZ has nothing to disclose.

LS has nothing to disclose.

SAT is an NIHR senior investigator.

LMS has nothing to disclose.

MKV has nothing to disclose.

RK has received consulting fees from AbbVie, Takeda, and Janssen.

CS has nothing to disclose.

JR has nothing to disclose.

PH has received consulting fees from Abbvie and Takeda; speakers fees from Ferring, Falk Pharma, Vifor Pharma, Tillotts Pharma, Chiesi, Takeda and Abbvie.

CYN has nothing to disclose.

WJS has served as a consultant to: AbbVie Inc., ActoGeniX NV, AGI Therapeutics, Inc., Alba Therapeutics Corporation, Albireo, Alfa Wasserman, Amgen, AM-Pharma BV, Anaphore, Astellas Pharma, Athersys, Inc., Atlantic Healthcare Limited, Axcan Pharma (now Aptalis), BioBalance Corporation, Boehringer-Ingelheim Inc, Bristol Meyers Squibb, Celgene, Celek Pharmaceuticals, Cellerix SL, Cerimon Pharmaceuticals, ChemoCentryx, CoMentis, Cosmo Technologies, Coronado Biosciences, Cytokine Pharmasciences, Eagle Pharmaceuticals, Eisai Medical Research Inc., Elan Pharmaceuticals, EnGene, Inc., Eli Lilly, Enteromedics, Exagen

Diagnostics, Inc., Ferring Pharmaceuticals, Flexion Therapeutics, Inc., Funxional Therapeutics Limited, Genzyme Corporation, Genentech (now Roche), Gilead Sciences, Given Imaging, Glaxo Smith Kline, Human Genome Sciences, Ironwood Pharmaceuticals (previously Microbia Inc.), Janssen (previously Centocor), KaloBios Pharmaceuticals, Inc., Lexicon Pharmaceuticals, Lycera Corporation, Meda Pharmaceuticals (previously Alaven Pharmaceuticals), Merck Research Laboratories, MerckSerono, Millennium Pharmaceuticals (subsequently merged with Takeda), Nisshin Kyorin Pharmaceuticals Co., Ltd., Novo Nordisk A/S, NPS Pharmaceuticals, Optimer Pharmaceuticals, Orexigen Therapeutics, Inc., PDL Biopharma, Pfizer, Procter and Gamble, Prometheus Laboratories, ProtAb Limited, Purgenesis Technologies, Inc., Receptos, Relypsa, Inc., Salient Pharmaceuticals, Salix Pharmaceuticals, Inc., Santarus, Schering Plough Corporation (acquired by Merck), Shire Pharmaceuticals, Sigmoid Pharma Limited, Sirtris Pharmaceuticals, Inc. (a GSK company), S.L.A. Pharma (UK) Limited, Targacept, Teva Pharmaceuticals, Therakos, Tillotts Pharma AG (acquired by Zeria Pharmaceutical Co., Ltd), TxCell SA, UCB Pharma, Viamet Pharmaceuticals, Vascular Biogenics Limited (VBL), Warner Chilcott UK Limited; has received speaker's fees from: AbbVie Inc., Bristol Meyers Squibb, and Janssen (previously Centocor); and financial support for research from: AbbVie Inc., Bristol Meyers Squibb, Genentech, Glaxo Smith Kline, Janssen (previously Centocor), Millennium Pharmaceuticals (now Takeda), Novartis, Pfizer, Procter and Gamble Pharmaceuticals, Shire Pharmaceuticals, and UCB Pharma.

GD'H has received consulting and/or lecture fees from AbbVie, ActoGeniX, AIM, Boehringer Ingelheim GmbH, Centocor, Chemo Centryx, Cosmo Technologies, Elan Pharmaceuticals, enGene, Dr Falk Pharma, Ferring, Galapagos, Giuliani SpA, Given

Imaging, GlaxoSmithKline, Janssen Biologics, MSD, Neovacs, Novo Nordisk, Otsuka, PDL BioPharma, Pfizer, Receptos, Salix, SetPoint, Shire Pharmaceuticals, Schering-Plough, Takeda, Tillotts Pharma, UCB Pharma, Versant, and Vifor Pharma; research grants from AbbVie, Janssen, Given Imaging, MSD, Dr Falk Pharma, and PhotoPill; and speaking honoraria from AbbVie, Tillotts, Tramedico, Ferring, MSD, UCB Pharma, Norgine, and Shire.

BGF BGF has received grant/research support from Millennium Pharmaceuticals, Merck, Tillotts Pharma AG, AbbVie, Novartis Pharmaceuticals, Centocor Inc., Elan/Biogen, UCB Pharma, Bristol-Myers Squibb, Genentech, ActoGenix, and Wyeth Pharmaceuticals Inc.; consulting fees from Millennium Pharmaceuticals, Merck, Centocor Inc., Elan/Biogen, Janssen-Ortho, Teva Pharmaceuticals, Bristol-Myers Squibb, Celgene, UCB Pharma, AbbVie, Astra Zeneca, Serono, Genentech, Tillotts Pharma AG, Unity Pharmaceuticals, Albireo Pharma, Given Imaging Inc., Salix Pharmaceuticals, Novonordisk, GSK, Actogenix, Prometheus Therapeutics and Diagnostics, Athersys, Axcan, Gilead, Pfizer, Shire, Wyeth, Zealand Pharma, Zyngenia, GiCare Pharma Inc., and Sigmoid Pharma; and speakers bureaux fees from UCB, AbbVie, and J&J/Janssen.

VJ has received scientific advisory board fees from AbbVie and Sandoz; speakers fees from Takeda and Janssen.

JS has received research-consulting fees from Robarts Clinical Trials.

Robarts Clinical Trials began in 1986 as an academic research unit within the Robarts Research Institute which is affiliated with University Hospital and the University of Western Ontario. A subsequent international (United States of America and Netherlands) expansion in 2012 necessitated establishment of a corporate entity

to meet international federal/taxation regulations. All profits from Robarts Clinical Trials, Inc. are directed towards academic research. The University of Western Ontario is the sole shareholder of Robarts Clinical Trials Inc. None of the authors with affiliation to Robarts Clinical Trials, Inc. have an equity position or any shares in the corporation.

2. Declaration of funding interests: None.