Original Investigation

Semiautomatic Assessment of the Terminal Ileum and Colon in Patients with Crohn Disease Using MRI (the VIGOR++ Project)

Carl A.J. Puylaert, MSc[1], Peter J. Schüffler, PhD[1], Robiel E. Naziroglu, PhD, Jeroen A. W. Tielbeek, MD, PhD, Zhang Li, PhD, Jesica C. Makanyanga, MD, Charlotte J. Tutein Nolthenius, MD, PhD, C. Yung Nio, MD, Douglas A. Pendsé, MD, PhD, Alex Menys, PhD, Cyriel Y. Ponsioen, MD, PhD, David Atkinson, PhD, Alastair Forbes, MD, PhD, Joachim M. Buhmann, PhD, Thomas J. Fuchs, PhD, Haralambos Hatzakis, Lucas J. van Vliet, PhD, Jaap Stoker, MD, PhD, Stuart A. Taylor, MD, PhD, Frans M. Vos, PhD

Rationale and Objectives: The objective of this study was to develop and validate a predictive magnetic resonance imaging (MRI) activity score for ileocolonic Crohn disease activity based on both subjective and semiautomatic MRI features.

Materials and Methods: An MRI activity score (the "virtual gastrointestinal tract [VIGOR]" score) was developed from 27 validated magnetic resonance enterography datasets, including subjective radiologist observation of mural T2 signal and semiautomatic measurements of bowel wall thickness, excess volume, and dynamic contrast enhancement (initial slope of increase). A second subjective score was developed based on only radiologist observations. For validation, two observers applied both scores and three existing scores to a prospective dataset of 106 patients (59 women, median age 33) with known Crohn disease, using the endoscopic Crohn's Disease Endoscopic Index of Severity (CDEIS) as a reference standard.

Results: The VIGOR score (17.1 × initial slope of increase + 0.2 × excess volume + 2.3 × mural T2) and other activity scores all had comparable correlation to the CDEIS scores (observer 1: $r = 0.58$ and 0.59, and observer 2: $r = 0.34$–0.40 and 0.43–0.51, respectively). The VIGOR score, however, improved interobserver agreement compared to the other activity scores (intraclass correlation coefficient = 0.81 vs 0.44–0.59). A diagnostic accuracy of 80%–81% was seen for the VIGOR score, similar to the other scores.

Conclusions: The VIGOR score achieves comparable accuracy to conventional MRI activity scores, but with significantly improved reproducibility, favoring its use for disease monitoring and therapy evaluation.

Key Words: Crohn disease; Magnetic resonance imaging; Image interpretation, computer-assisted; Ileum; Colon.

INTRODUCTION

Crohn disease (CD) is an inflammatory bowel disease, which can present throughout the gastrointestinal tract, particularly affecting the small bowel and the colon. Magnetic resonance imaging (MRI) is increasingly used for diagnosis and phenotyping of CD because it is safe, noninvasive, and has high accuracy for evaluating enteric disease and extramural complications (1). MRI features such as wall thickness and T1 and T2 bowel wall signals have been validated as biomarkers of CD activity, demonstrating good correlation with endoscopic and histopathologic grading of inflammation (2–4). Recent years have seen several MRI disease

Acad Radiol 2018; ■:■■–■■

From the Department of Radiology and Nuclear Medicine, Academic Medical Centre, Meibergdreef 9, P.O 22660, 1100DD, Amsterdam, The Netherlands (C.AJ.P., J.AW.T., C.J.T.N., C.Y.N., J.S., F.M.V.); Department of Computer Sciences, Eidgenössische Technische Hochschule Zurich, Zurich, Switzerland (P.J.S., J.M.B.); Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, New York (P.J.S., T.J.F.); Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands (R.E.N., Z.L., L.J..V., F.M.V.); College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China (Z.L.); Center for Medical Imaging, University College London Hospitals National Health Service Foundation Trust, London, UK (J.C.M., D.A.P., A.M., D.A., S.A.T.); Department of Gastroenterology, Academic Medical Centre, Amsterdam, The Netherlands (C.Y.P.); Norwich Medical School, University of East Anglia, Norwich (A.F.); Biotronics3D Ltd, London, UK (H.H.). Received October 24, 2017; revised December 20, 2017; accepted December 25, 2017.

Sources of Funding: The VIGOR++ project was funded through a research grant from the European Union's Seventh Framework Programme (project number 270379). The project was supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre. Thomas Fuchs and Peter Schüffler were funded in part through the National Institutes of Health and National Cancer Institute Cancer Center Support Grant P30 CA008748.

[1]These authors contributed equally to this article.**Address correspondence to:** C.A.J.P. e-mail: c.a.puylaert@amc.uva.nl

https://doi.org/10.1016/j.acra.2017.12.024

activity scores being developed and externally validated, combining multiple MRI features to predict overall disease activity (3–6). These scores are gradually disseminating into clinical practice, although at present, they are predominantly employed as research tools. The magnetic resonance index of activity (MaRIA), for example, has been developed using the Crohn's Disease Endoscopic Index of Severity (CDEIS) as a reference standard. The MaRIA is based on quantitative measurement of bowel wall relative contrast enhancement, along with subjective evaluation of mural ulceration and abnormal T2 signal (3). Other indices, such as the London score and the Crohn disease MRI index (CDMI), rely on qualitative grading of various features by reporting radiologists (4,6). Such activity scores can be applied to individual bowel segments, as well as to the patient as a whole, as both are important to clinical management. Before MRI scores can be widely adopted for evaluating disease activity and therapeutic monitoring, high accuracy across the spectrum of disease severity and good reproducibility among radiologists must be proven. The current literature, however, reports variable reproducibility for many features used in MRI activity scores (6,7).

One potential solution to the current limitations of MRI activity scoring is to incorporate novel software solutions, which can automatically extract relevant features from MRI data. Such software could reduce both interobserver variability and the risk of observer bias inherent to subjective evaluation (8). New MRI image processing methods are available, which give semiautomatic measurements of bowel wall thickness, providing superior reproducibility over manual measurement (9). Further techniques have been developed that automatically extract perfusion parameters from motion corrected free-breathing dynamic contrast-enhanced (DCE) MRI (10). Although several studies have shown the potential of semiautomatic MRI assessment of CD (9–11), none of those have examined clinical practicability or validated their results using a large, independent cohort.

We hypothesize that a scoring system combining semiautomatic software measurements with conventional subjective radiologist scoring of MRI features can improve accuracy and reproducibility in comparison to existing MRI scores. Accordingly, our aim was to develop and validate a predictive MRI score for ileocolonic CD activity incorporating novel software-assisted semiautomatic measurement of MRI features using an ileocolonoscopic standard of reference, and to compare its performance with existing MRI activity scores.

MATERIALS AND METHODS

The study was divided in two phases. Firstly, a detailed modeling process was undertaken to derive two new MRI activity scores. Secondly, these new scores were validated and compared to existing scores regarding accuracy for diagnosis and grading of disease, as well as score reproducibility. Ethical permission was obtained from both institutions' medical ethics committee, and written informed consent was obtained from all patients.

Phase 1—Model Development

The modeling process employed a previously described cohort of 27 patients with known CD (6). The first developed score specifically incorporated semiautomatic measurements of bowel wall thickness and enhancement (described in more detail further in phase 2) and was termed the "virtual gastrointestinal tract (VIGOR) score." The second score incorporated only the best performing combination of a number of subjective evaluations made by radiologists (termed the "subjective score"). A full description of the model development is given in Appendix A.

Phase 2—Prospective Activity Score Testing and Model Comparison

The validation and comparison of the newly developed and existing activity scores were performed using an independent prospective cohort. Between October 2011 and September 2014, consecutive patients aged ≥18 years with suspected or known CD and scheduled for ileocolonoscopy were recruited from two European tertiary referral centres for inflammatory bowel disease (1. Academic Medical Center (AMC), Amsterdam, the Netherlands, and 2. University College London Hospital (UCLH), London, United Kingdom). All included patients underwent MRI and ileocolonoscopy within 2 weeks. The Harvey-Bradshaw Index (HBI) was collected at the time of MRI (12).

Patient exclusion criteria were contraindications to MRI (eg, pacemakers and claustrophobia), a final diagnosis other than CD, failure to comply with the oral contrast protocol, >2 weeks between MRI and ileocolonoscopy, and incomplete MRI protocol (eg, missing sequences or incomplete imaging), or insufficient bowel cleansing precluding accurate mucosal assessment, as determined by the endoscopist.

Reference Standard

Ileocolonoscopy was performed within 2 weeks of MRI using a standard endoscope (model CF-160L, Olympus) by either a gastroenterologist or a senior resident in gastroenterology under direct supervision of a gastroenterologist. The endoscopist applied the CDEIS to evaluate endoscopic disease (13). The endoscopist was blinded to findings on MRI, except for cases where a balloon-dilatation procedure was indicated. In these cases, the length of stenosis on MRI was used to determine the feasibility of balloon dilatation.

MRI Protocol

Patients fasted for at least 4 hours before the examination and were instructed to drink a total of 2400 mL 2.5% mannitol solution (Baxter, Utrecht, The Netherlands) split in two doses, 800 mL (3 hours before MRI) and 1600 mL (1 hour before MRI), to achieve distension of both colonic and small bowel segments. MRI examinations were performed on a 3-T MRI unit (Ingenia and Achieva; Philips, Best, The Netherlands)