

Maurice van den Bosch · Doortje Graafmans ·
Rutger Nievelstein · Erik Beek

Systematic assessment of constipation on plain abdominal radiographs in children

Received: 8 September 2005 / Accepted: 1 November 2005 / Published online: 18 January 2006
© Springer-Verlag 2006

Abstract *Background:* Constipation in childhood is common and its clinical assessment is often difficult. Plain abdominal radiography is simple and used to quantify constipation. Three scoring systems, those of Barr et al., Leech et al. and Blethyn et al., have been developed to quantify fecal loading on the abdominal radiograph. *Objective:* In order to determine which method is the most useful in clinical practice, we assessed the reproducibility of the three scoring systems. *Materials and methods:* Plain abdominal radiographs from 40 clinically constipated children were retrospectively reviewed by two paediatric radiologists on two separate occasions. The radiographs were scored according to three different systems developed by Barr et al., Leech et al., Blethyn et al. Intraobserver variability and interobserver reproducibility were determined for each system. Kappa coefficients were calculated as indicators of inter- and intraobserver agreement for categorical outcome variables. *Results:* The Leech score showed the highest reproducibility: the intraobserver agreement was high for both observers (κ values of 0.88 and 1.00, respectively, $P < 0.05$). Furthermore, the interobserver agreement was also high: κ 0.91 in the first round and 0.84 in the second. *Conclusions:* The Leech score proved to be a highly reproducible tool for assessment of

childhood constipation and is of value in clinical practice for systematic assessment of constipation on plain abdominal radiographs in children.

Keywords Constipation · Radiography · Comparison · Children

Introduction

Constipation is a frequently encountered problem in children and may cause considerable distress to the child and family [1]. There is no generally accepted definition of constipation [2]. Subjective history criteria, such as defecation frequency, straining on defecation and stool consistency are used most often. Objective criteria, such as whole gut transit time and mean daily stool weight have also been used, but are cumbersome in that they require stool collection or daily abdominal radiographs for 4–5 days [3, 4].

Plain abdominal radiography is frequently used to complement clinical history and examination. An abdominal radiograph is simple and cheap, and does not expose the patient to a high radiation dose [1]. However, in the clinical setting there is no common policy on how to review the plain abdominal film. Over previous decades, three different scoring systems for assessment of fecal retention on a single radiograph of the abdomen have been proposed [5–7]. The earliest was reported by Barr et al. [5] in 1979, followed by Blethyn et al. [6] in 1995 and Leech et al. [7] in 1999. The systems have in common that they score both the amount of feces and the localization of it in different colon segments. Previous studies have assessed the diagnostic accuracy of these systems and independently concluded that all three methods are reliable, with both sensitivity and specificity up to 90% [7].

No studies have investigated and compared the reproducibility of these systems. To assess which scoring system is the most useful in the clinical setting we de-

M. van den Bosch (✉) · D. Graafmans
Department of Radiology (Room E.01.132),
University Medical Center Utrecht,
Heidelberglaan 100,
Utrecht, 3584CX, The Netherlands
e-mail: m.a.vandenbosch@azu.nl
Tel.: +31-30-2506688
Fax: +31-30-2581098

R. Nievelstein · E. Beek
Department of Paediatric Radiology,
Wilhelmina Children's Hospital,
Utrecht, The Netherlands

terminated intra- and interobserver variability of these three scoring systems.

Materials and methods

We retrospectively reviewed 40 plain abdominal radiographs of children consecutively referred to our hospital for assessment of constipation between January and December 2001. The radiographs were taken at presentation. Patients complained of infrequent defecation, soiling, encopresis, or abdominal pain.

Masked plain abdominal radiographs of the 40 children were independently evaluated by two observers, both experienced paediatric radiologists. The observers assessed each abdominal radiograph on two separate occasions, 6 weeks apart. Each abdominal radiograph was scored according to the three different scoring systems.

The first method described by Barr et al. [5] quantifies the amount of feces in four different bowel segments (ascending colon, transverse colon, descending colon and rectum) and also the consistency of the feces, i.e. granular or rocky stools. Constipation is defined as a Barr score >10. The second method described by Blethyn et al. [6] is a rough scoring system used to assess the amount of feces in the large bowel. In this system the degree of constipation is graded as follows: *grade 0* normal, feces in rectum and caecum only; *grade 1* mild constipation, feces in rectum and caecum and discontinuous elsewhere; *grade 2* moderate constipation, feces in rectum and caecum, continuous and affecting all segments; *grade 3* severe constipation, feces in rectum and caecum, continuous elsewhere and dilated colon and impacted rectum. The third method was described more recently by Leech et al. [7]. In this system the colon is divided into three colonic segments (Fig. 1) (1 ascending colon and proximal transverse colon; 2 distal transverse colon and descending colon; 3 rectosigmoid) and the amount of feces in each segment is scored from 0 to 5, 0 indicating no feces and 5 indicating severe fecal loading and bowel dilatation. With a possible score of 0–15, a score of >8 is considered to indicate constipation.

Intraobserver variability was determined for each scoring system by comparing data from the same observer at two different reading sessions. Interobserver reproducibility was determined by comparing data from the two observers on one occasion. Thus two intraobserver and two interobserver variabilities could be derived for each parameter. Kappa coefficients were calculated as indicators of intra- and interobserver variability, and coefficients <0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 were considered to indicate poor, fair, moderate, good, and very good agreement, respectively [8].

Results

Table 1 shows the age and sex distribution of the total study population. The mean age of the patients was 7 years (range 3–12 years); 18 patients (45%) were girls and 22

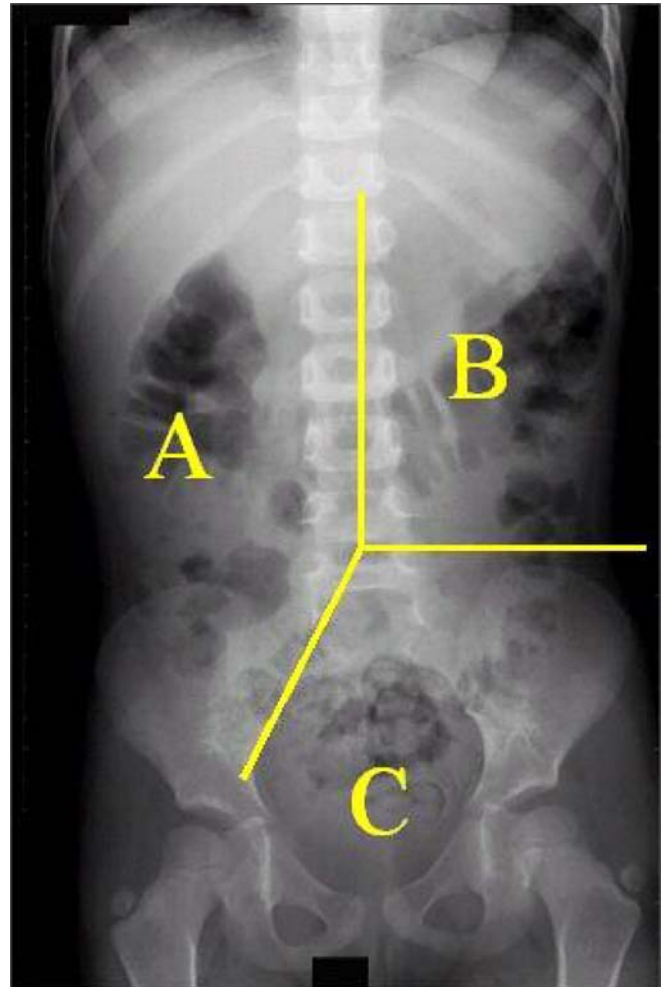


Fig. 1 Systematic review of an abdominal radiograph according to the method developed by Leech et al. [7]

(55%) were boys. Most of these patients (38%) presented with diffuse abdominal pain. Other symptoms included encopresis/soiling (23%), low defecation frequency (27%) and diarrhoea (10%).

The intra- and interobserver coefficients according to the three different scoring systems are presented in Table 2. The Leech score showed the highest reproduc-

Table 1 Baseline characteristics of the study population

Age (years)	
Mean	7
Range	3–12
Boys, <i>n</i> (%)	22 (55%)
Symptoms, <i>n</i> (%)	
Diffuse abdominal pain	15 (38%)
Encopresis/soiling	9 (23%)
Defecation fewer than three times per week	11 (27%)
Diarrhoea	4 (10%)
Other ^a	

^aBehavioural problems in three patients and eating problems in one patient

Table 2 Interobserver and intraobserver variability (κ values) according to different scoring systems

Scoring system	Intraobserver variability		Interobserver variability period 1	
	Observer 1	Observer 2	Period 1	Period 2
Barr et al.	0.75	0.66	0.45	0.71
Blethyn et al.	0.61	0.65	0.43	0.31
Leech et al.	0.88	1.00	0.91	0.84

All κ values are statistically significant ($P < 0.05$)

ibility with high intraobserver agreement for both observers (κ 0.88 and 1.00, respectively), and high interobserver agreement (κ 0.91 in the first round and κ 0.84 in the second round).

Agreement on the existence of constipation as measured by the Barr score between the first and second scoring round was fair for both observers with an intraobserver agreement (κ) of 0.75 and 0.66, respectively ($P < 0.05$). However, the interobserver agreement was moderate in the first round reflected by a κ value of 0.45. The reproducibility of the Blethyn score was the lowest of the three scoring systems with κ values ranging from 0.61 to 0.65 for intraobserver agreement and 0.31 to 0.43 for interobserver variability.

Discussion

Constipation is a common clinical problem and frequently dominates the life of a child and his or her family. There is no consensus on the criteria to be used in the diagnosis of constipation [2]. Usually, an arbitrary combination of subjective indicators (history data) and objective indicators is used [9]. Various objective methods have been used to assess constipation, including rectal examination after defecation, rectal manometry with balloon insufflation and anal sphincter electromyography, but all are invasive. Others have used radiopaque markers for the assessment of constipation. This type of study produces semiquantitative information on the colorectal transit time, but requires up to 6 days to complete [10].

The plain abdominal radiograph is another diagnostic tool that can be used to quantify constipation [3–5]. The test is convenient for both patient and physician, takes little time and is less cumbersome than, for instance, measurement of the whole gut pellet transit time. Furthermore, previous studies have shown that a plain abdominal radiograph is as reliable as measurement of fecal weight or marker transit and can thus be advocated as one of the first diagnostic procedures.

During the past decades three different scoring systems for systematic review of a plain abdominal film, used to

estimate the amount of fecal retention, have been described [5–7]. The first one was developed by Barr and colleagues in 1976 [5]. They stated that in the absence of a clear diagnosis of constipation, only one abdominal film is necessary for detecting and grading occult stool retention in children with symptoms of constipation. More recently, Blethyn et al. [6] devised a simple scoring system and reported a strong correlation between their score and a symptom score based on bowel frequency [6]. Their findings were supported by the study of Leech et al. [7] who developed another scoring system, and concluded it to be a clinically useful tool with sensitivities and specificities up to 91% and 85%, respectively [7].

Although the diagnostic accuracy of these methods has been evaluated in several studies, their reproducibility has never been properly determined. To our knowledge, our study is the first to assess which scoring system is the most reproducible in a clinical setting by calculating and comparing the intra- and interobserver variability of the three different scoring systems. We found that the Leech score is the most reproducible grading system for assessment of childhood constipation on a plain abdominal radiograph with high intra- and interobserver agreement with κ values ranging from 0.84 to 1.00. We therefore conclude that this system could be of value in clinical practice for systematic quantification of constipation on plain abdominal radiographs in children.

References

- Halligan S, Bartram CI (1995) The radiological investigation of constipation. *Clin Radiol* 50:429–435
- Hinton JM, Lennard-Jones JE (1968) Constipation: definition and classification. *Postgrad Med J* 40:720
- Benninga MA, Buller HA, Staalman CR, et al (1995) Defaecation disorders in children: colonic transit time versus the Barr-score. *Eur J Pediatr* 154:277–284
- Starreveld JS, Pols MA, Van Wijk HJ, et al (1990) The plain abdominal radiograph in the assessment of constipation. *Z Gastroenterol* 28:335–338
- Barr RG, Levine MD, Wilkinson RH, et al (1979) Chronic and occult stool retention; a clinical tool for its evaluation in school-aged children. *Clin Pediatr* 18:674–679
- Blethyn AJ, Verrier Jones K, Newcombe R, et al (1995) Radiological assessment of constipation. *Arch Dis Child* 73:532–533
- Leech SC, McHugh K, Sullivan PB (1999) Evaluation of a method of assessing faecal loading on plain abdominal radiographs in children. *Pediatr Radiol* 29: 255–258
- Altman DG (1991) *Practical statistics for medical research*. Chapman and Hall, London
- Manning G, Wyman JB, Heaton KW (1976) How trustworthy are bowel histories? Comparison of recalled and recorded information. *BMJ* 2:213–214
- Metcalf AM, Phillips SF, Zinsmeister AR, et al (1987) Simplified assessment of segmental colonic transit. *Gastroenterology* 92:40–47