

Lack of Utility of Abdominal X-rays in the Evaluation of Children With Constipation: Comparison of Different Scoring Methods

Licia Pensabene, Carlo Buonomo, Laurie Fishman, Denesh Chitkara, and Samuel Nurko

ABSTRACT

Background and Aim: Abdominal x-rays are used diagnostically in the evaluation of children with constipation. However, their clinical utility has not been established. The aim of the study was to assess the accuracy of different methods in identifying children with functional constipation (FC) or nonretentive fecal incontinence (NRFI).

Patients and Methods: Retrospective review of abdominal x-rays in which colonic transit (CT), Barr, Leech, and fecal loading (FL) scores were blindly measured by blinded pediatric gastroenterologists and a radiologist. Children were classified a priori as FC or NRFI.

Results: One hundred sixty patients (125 FC, 35 NRFI) were studied. There were significant differences ($P < 0.05$) when comparing those with FC and those with NRFI: CT: 51 ± 18 vs 40 ± 21 hours; Barr: 14 ± 5 vs 11 ± 4 ; Leech: 10 ± 2 vs 8 ± 2 ; FL: 2 ± 0.5 vs 1.7 ± 0.4 . More than 20% of FC had normal Barr and Leech scores, whereas $>50\%$ of NRFI had abnormal scores. CT discriminated better between FC and NRFI. There was a significant correlation ($P < 0.05$) between CT and Barr (0.45), Leech (0.41) and FL scores (0.36), and between Barr and Leech scores ($r = 0.94$). There was good intraobserver correlation between Barr, Leech, and FL scores but poor interobserver reproducibility.

Conclusions: Although significant differences in overall FC and NRFI scores exist, the discriminative value is low for all scores. There is poor interobserver reproducibility of the Barr, Leech, and FL scores. These findings confirm the limited value of the plain abdominal x-ray in the evaluation of children with constipation.

(JPGN 2010;51: 155–159)

One of the most common functional gastrointestinal disorders (FGIDs) in childhood is chronic constipation (1–4). The Rome III FGID classification (5–7) and the Paris Consensus on Childhood Constipation Terminology (PACCT) (8) state that the identification and classification of functional constipation (FC) should be based on symptoms and clinical history. However, a

plain abdominal radiograph, in which fecal loading (FL) is determined, is still frequently used in the evaluation of children with constipation, either to support the diagnosis or to suggest the type or response to treatment (9). Plain visual inspection is highly subjective, and its clinical utility has not been established, as suggested in the evidence-based guidelines published by the North American Society of Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) (9). Limited studies and a recently published systematic review have indicated that the plain abdominal radiograph to assess FL has high inter- and intraobserver variability and poor diagnostic accuracy (9–15). Therefore, to make the evaluation of FL in abdominal radiograph less subjective, different scoring systems have been developed (11,16). However, there is limited information on the use and utility of these different scoring methods during daily clinical work (9,13,14,17–20).

The aim of this study was to assess and compare the accuracy of different radiographic scoring methods for the identification of children with FC.

PATIENTS AND METHODS

Study Population

All of the children with defecation complaints who were referred to the Pediatric Gastrointestinal Motility Unit at Children's Hospital Boston from 2000 to 2004, who had undergone a colonic transit (CT) study as part of their clinical evaluation, were eligible for this study. Patients with organic disease, known causes of constipation (Hirschsprung disease, hypothyroidism, myelomeningocele metabolic, or renal diseases), or fecal impaction as defined by the PACCT criteria (8) were excluded. The study was approved by the institutional review board.

Clinical Diagnosis

Children were classified a priori, without any knowledge of the CT time, by 2 blinded investigators, as having either FC or nonretentive fecal incontinence (NRFI) using the PACCT (8) and Rome III criteria (6,7). The diagnosis of FC was based on the following criteria.

The occurrence of 2 or more of the following characteristics for at least 8 weeks:

1. Frequency of bowel movements, less than 3 per week
2. At least 1 episode of fecal incontinence per week
3. Large stools in the rectum or palpable on abdominal examination
4. Passing of stools so large that they may obstruct the toilet
5. Display of retentive posturing and withholding behaviors
6. Painful defecation

Received September 7, 2009; accepted November 8, 2009.

From the Center for Motility and Functional Gastrointestinal Disorders, Children's Hospital, Boston, MA.

Address correspondence and reprint requests to Samuel Nurko, MD, Children's Hospital, 300 Longwood Ave, Boston, MA 02115 (e-mail: Samuel.nurko@childrens.harvard.edu).

Dr Pensabene is presently at the Department of Pediatrics, University Magna Graecia, Catanzaro, Italy.

This study was supported in part by NIH grant K24DK82792-1 (Dr Nurko). The authors report no conflicts of interest.

Copyright © 2010 by European Society for Pediatric Gastroenterology, Hepatology, and Nutrition and North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition

DOI: 10.1097/MPG.0b013e3181cb4309

Patients with FC were further divided into 2 groups according to the presence or absence of fecal incontinence. The diagnosis of NRFI was based on the passage of stools in an inappropriate place for at least 8 weeks, occurring in children with a mental age of 4 years and older, with no evidence of constipation based on history and/or examination.

Abdominal Radiography and Scoring Methods

Deidentified abdominal x-rays performed for CT measurement were retrospectively reviewed by 3 blinded pediatric gastroenterologists (S.N., L.F., and D.C.) and a pediatric radiologist (C.B.), and scores for CT, Barr, Leech, and a subjective amount of FL were measured. The x-rays were shown to each investigator in a random fashion. Before scoring any films, the methods for the Barr and Leech scores were reviewed and discussed until both investigators agreed on the characteristics of each scoring element (see below). Then the Barr and Leech scores were independently calculated by 1 pediatric gastroenterologist (S.N.) and the radiologist (C.B.). Three pediatric gastroenterologists (S.N., L.F., and D.C.) independently scored the amount of FL present in each radiograph. No attempt to standardize this subjective score was undertaken. The CT time was assessed by a single independent scorer (L.P.) in duplicate to check for accuracy because it was assumed that counting radiopaque markers would not lead to intra- or interobserver variability.

CT

All laxatives were stopped at least 48 hours before the CT was done. Therefore, the CT was performed while patients were not taking medications. All of the patients had a physical examination before the CT was requested to ensure that there was no fecal impaction (as defined by the PACCT criteria) (8). Therefore, no patient with a fecal impaction was included.

The CT study was done following a modification of the technique reported by Metcalf et al (21): patients ingested 1 capsule containing 24 radiopaque markers (Sitzmarks; Konsyl Pharmaceuticals, Inc, Fort Worth, TX) at a similar time each morning for 3 consecutive days; on the fourth day a plain abdominal radiograph was obtained at the same time of day (21). The abdomen was then divided into 3 segments as outlined by Arhan et al (22). Markers were counted in each segment. Total transit time was calculated using the formula described by Metcalf et al, adjusting for the fact that 24 markers were ingested each day instead of 20 as in the original formula. This adjustment allows the calculation of transit times (in hours) by simply counting the markers present in the abdominal radiograph taken on the fourth day. The reference values used were those of Arhan et al (22) in healthy children. The presence of >62 markers was considered abnormal.

Leech Score

The Leech score divides colon into 3 segments: right colon, left colon, and rectosigmoid segment. Each segment is provided with a score from 0 to 5; 0 = no feces visible, 1 = scanty feces, 2+ = mild FL; 3 = moderate FL; 4 = severe FL; 5 = severe FL with bowel dilatation (11). Scores range from 0 to 15. A score of 9 or more was considered abnormal (11).

Barr Score

The Barr score quantifies the amount of feces in 4 bowel segments (ascending colon, transverse colon, descending colon, and

rectum), and also the consistency of the feces (10). A score of >10 was considered abnormal.

FL

FL was a subjective score in which the investigators classified the amount of stool in the radiograph as mild (1 point), moderate (2 points), or severe (3 points), in accordance with their clinical practice. A score of >2 was considered abnormal.

Statistical Analyses

All of the results are expressed as mean \pm SE. Parametric and nonparametric tests were used to compare the different patient characteristics and the scores between the 2 groups. The reported scores represent the averages obtained from the individual scores. A subgroup comparison was performed in the group with FC according to the presence or absence of fecal incontinence. Sensitivity and specificity were calculated; κ coefficients were calculated as indicators of inter- and intraobserver agreement for categorical outcome variables. Receiver operator characteristic (ROC) curves were created, and the area under the curve (AUC) and 95% confidence intervals calculated. From the AUC, the diagnostic accuracy was graded as low (0.5–0.7), moderate (0.7–0.9), or high (>0.9) (23).

RESULTS

There were 160 consecutive patients with functional defecation disorders who underwent abdominal x-rays for CT. Within this cohort, a total of 125 children fulfilled the clinical criteria for FC, and 35 were diagnosed as having NRFI. Table 1 shows baseline characteristics of the studied patients.

Comparison of Scores Between Children With FC or NRFI

There were significant differences ($P < 0.05$) among different scores when comparing children with FC to those with NRFI (Fig. 1). Distribution of normal and abnormal scores in the 2 patient groups is shown in Table 2. Although Barr and Leech scores were abnormal more often in children with FC, it should be noted that >20% of FC had normal scores and >50% of NRFI had abnormal scores.

TABLE 1. General characteristics of the patients

	FC	NRFI
n	125	35
Age, y	8.7 \pm 0.4	9.2 \pm 0.5
Sex (male), %	58%	80%*
Duration of problem, y	4.0 \pm 0.3	4.5 \pm 0.6
BM, per week	4.5 \pm 0.4	7.6 \pm 0.7*
Fecal incontinence, %	62.4	100*
Accidents, per week	16.0 \pm 0.4	12.8 \pm 2.2
Abdominal pain, %	33	9*
Rectal bleeding, %	7	0
Toilet trained, %	70	68

BM = bowel movement; FC = functional constipation; NRFI = nonretentive fecal incontinence.

* $P < 0.01$.

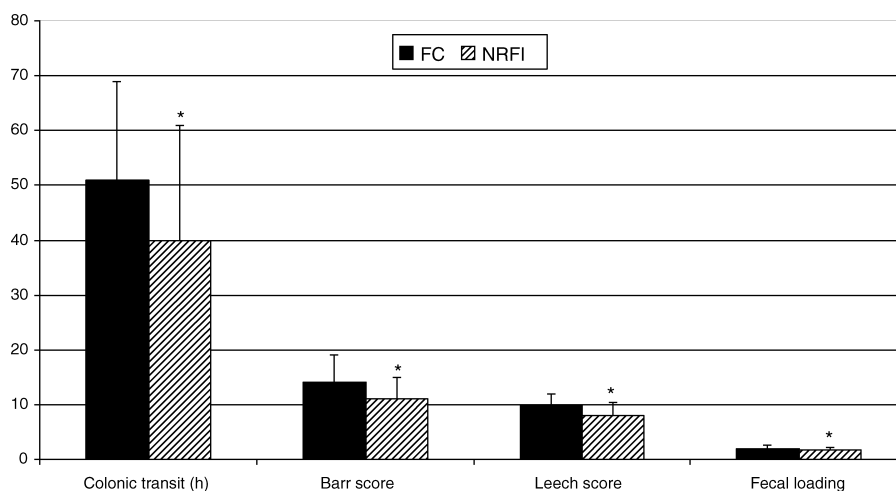


FIGURE 1. Comparison among children with FC and those with NRFI. There were significant differences among different scores when comparing children with FC to those with NRFI (* $P < 0.05$). FC = functional constipation; NRFI = nonretentive fecal incontinence.

Comparison of Scores in Children With FC Between Those With or Without Fecal Incontinence

Scores in patients with FC with and without fecal incontinence are shown in Figure 2. Patients with constipation and fecal incontinence had significantly higher scores as compared with those with constipation alone ($P < 0.01$).

Characteristics of the Scores

Correlation between the different scoring methods is shown in Table 3. There was a good correlation between the different methods. There was also a good intraobserver correlation between the Barr, Leech, and FL scores. However, there was a poor interobserver reproducibility of the Barr (κ 0.3), Leech (κ 0.25), and FL

scores (κ 0.3). Figure 3 shows ROC curves obtained by the different methods. The AUC were as follows: Barr score 0.672 (95% confidence interval 0.575–0.770); Leech score 0.661 (0.564–0.759), FL 0.626 (0.522–0.730) and CT 0.643 (0.535–0.751).

DISCUSSION

FC and NRFI are symptom-based diagnoses (6,7,9); however, in clinical practice the existence of fecal retention is frequently evaluated on plain abdominal radiographs and used in determination of diagnosis and therapy (15). We found that even though there were significant differences in colonic transit time and all visual scores when comparing patients with FC to those with NRFI, all of the scoring methods had low sensitivity, low interobserver reproducibility, and a low discriminative value to differentiate patients with FC from those with NRFI. In our study the subjective assessment of FL by experienced gastroenterologists had the worst

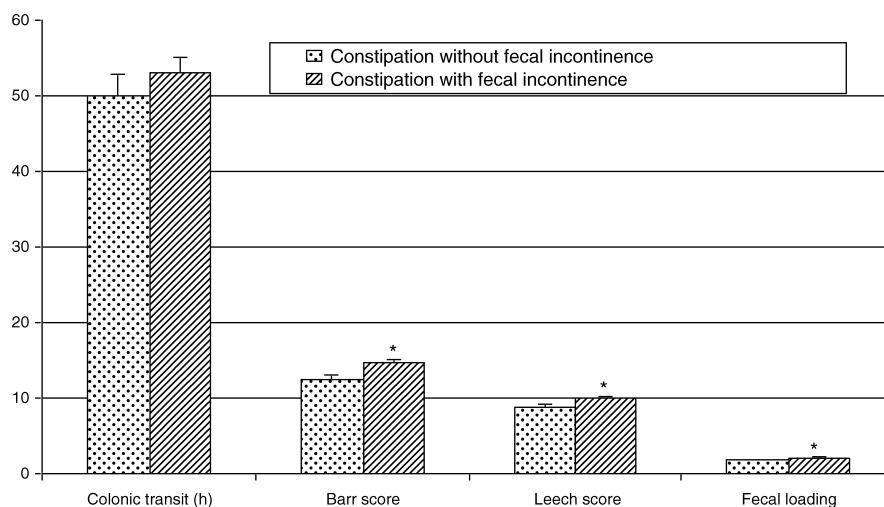


FIGURE 2. Comparison of scores in patients with functional constipation with or without fecal incontinence. Patients with constipation and fecal incontinence had significantly higher scores as compared with those with constipation alone (* $P < 0.01$).

TABLE 2. Distribution of normal and abnormal scores among the different patient populations

	CT	Barr	Leech	FL
Abnormal in FC, %	40	79	72	63
Abnormal in NRFI, %	23	59	48	41
Sensitivity	0.40	0.75	0.57	0.37
Specificity	0.77	0.48	0.74	0.80

Although Barr and Leech scores were abnormal more often in children with FC, it should be noted that >20% of FC had normal scores and >50% of NRFI had abnormal scores.

CT = colonic transit; FC = functional constipation; FL = fecal loading; NRFI = nonretentive fecal incontinence.

performance with respect to its ability to differentiate between groups. This is an important previously unpublished observation, because in clinical practice this subjective nonstandardized assessment is the method most commonly used to evaluate children with defecation abnormalities. Our findings suggest that if an abdominal x-ray is used in the evaluation of children with defecation problems, either the Barr or the Leech score will provide more reliable information, although all of the scores have a low diagnostic accuracy (23).

There was good correlation between the different scoring methods and a good intraobserver correlation, but a low interobserver reproducibility. Although experienced radiologists score more reliably than less experienced colleagues (24), high inter- and intraobserver variability and poor diagnostic accuracy of assessing FL on an abdominal radiograph have also been reported previously (10–14). In the present study, all of the observers were skilled and experienced in reading plain abdominal radiographs, yet showed the same poor interobserver reproducibility. However, the fact that there is a good intraobserver correlation suggests that there may be some value in the longitudinal follow-up and evaluation of the patient, as long as the same person is interpreting the x-ray every time.

The poor diagnostic accuracy of scoring a single abdominal radiograph may be due to the fact that it represents a single observation in time. The amount of FL is subject to daily variation, depending on food intake and timing of the last defecation. Another possibility for the low sensitivity of the different scoring methods may be that the Rome III and PACCT criteria are too broad for describing fecal disorders in children. Future research performed in unselected populations would be useful in addressing this question.

Because an abdominal radiograph has low sensitivity and requires the exposure of a child to a small dose of ionizing radiation, we do not advocate its use in the routine evaluation of children with constipation and other defecation abnormalities. An abdominal radiograph may be necessary in patients with obesity (9), if the child refuses a rectal examination (9), if there is suspicion of sexual abuse (25), or if there are other psychological factors that would

TABLE 3. Correlation between the different methods

	Barr score	Leech score	FL	CT
Barr score	1	0.937**	0.768**	0.450**
Leech score	0.937**	1	0.776**	0.414**
FL	0.768**	0.776**	1	0.361**
CT	0.450**	0.414**	0.361**	1

CT = colonic transit; FL = fecal loading.

** $P < 0.01$.

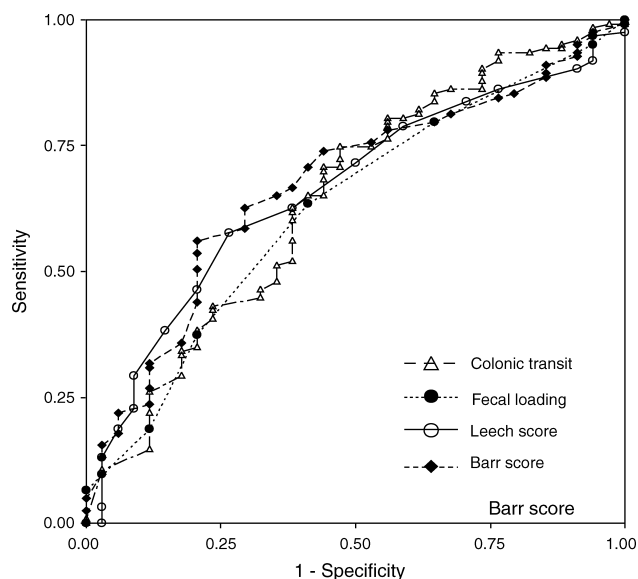


FIGURE 3. ROC curves for the different methods. There were no differences observed, and in all, the area under the curve was similar.

make rectal examination inappropriate or too traumatic (26). However, the findings of the radiograph need to be interpreted with caution.

There are some limitations to our study. We do not have concurrent physical examination data to correlate with the findings of the x-ray. However, a recently published systematic review (15) reported that most studies that have tried to assess whether there is an association between clinical symptoms and FL on radiographs in children have shown a lack of correlation. Another limitation is that we do not know the clinical impact that the interpretation of the x-ray had on the patient. We also do not have follow-up information to know whether our clinical diagnosis remained the same or changed over time. Prospective studies to validate the Rome III criteria and correlate clinical findings with findings on x-ray, as well as outcome studies based on x-ray findings, are needed.

In conclusion, the discriminative value to detect FC was low for all of the scoring methods, and there was poor interobserver reproducibility of the Barr, Leech, and FL scores. These findings confirm the limited value of the plain abdominal x-ray in the evaluation of children with constipation: use of radiographs is not a substitute for careful history taking and physical examination, which remain the cornerstone in the diagnostic workup of children with functional defecation disorders.

REFERENCES

- Loening-Baucke V. Chronic constipation in children. *Gastroenterology* 1993;105:1557–64.
- Partin JC, Hamill SK, Fischel JE, et al. Painful defecation and fecal soiling in children. *Pediatrics* 1992;89(6 Pt 1):1007–9.
- Loening-Baucke V. Prevalence, symptoms and outcome of constipation in infants and toddlers. *J Pediatr* 2005;146:359–63.
- Boccia G, Manguso F, Coccorullo P, et al. Functional defecation disorders in children: PACCT criteria versus Rome II criteria. *J Pediatr* 2007;151:394–8.
- Rasquin-Weber A, Hyman PE, Cucchiara S, et al. Childhood functional gastrointestinal disorders. *Gut* 1999;45(suppl 2):II60–8.
- Hyman PE, Milla PJ, Benninga MA, et al. Childhood functional gastrointestinal disorders: neonate/toddler. *Gastroenterology* 2006;130:1519–26.

7. Rasquin A, Di Lorenzo C, Forbes D, et al. Childhood functional gastrointestinal disorders: child/adolescent. *Gastroenterology* 2006;130:1527–37.
8. Benninga M, Candy DC, Catto-Smith AG, et al. The Paris Consensus on Childhood Constipation Terminology (PACCT) Group. *J Pediatr Gastroenterol Nutr* 2005;40:273–5.
9. Baker S, Liptak G, Colletti R, et al. Constipation in infants and children: evaluation and treatment. A medical position statement of the North American Society for Pediatric Gastroenterology Hepatology and Nutrition. *J Pediatr Gastroenterol Nutr* 2006;43:e1–3.
10. Barr RG, Levine MD, Wilkinson RH, et al. Chronic and occult stool retention: a clinical tool for its evaluation in school-aged children. *Clin Pediatr (Phila)* 1979;18:674. 676, 377–679, passim.
11. Leech SC, McHugh K, Sullivan PB. Evaluation of a method of assessing faecal loading on plain abdominal radiographs in children. *Pediatr Radiol* 1999;29:255–8.
12. Blethyn AJ, Verrier Jones K, Newcombe R, et al. Radiological assessment of constipation. *Arch Dis Child* 1995;73:532–3.
13. Benninga MA, Buller HA, Staalman CR, et al. Defaecation disorders in children, colonic transit time versus the Barr score. *Eur J Pediatr* 1995;154:277–84.
14. Jackson CR, Lee RE, Wylie AB, et al. Diagnostic accuracy of the Barr and Blethyn radiological scoring systems for childhood constipation assessed using colonic transit time as the gold standard. *Pediatr Radiol* 2009;39:664–7.
15. Reuchlin-Vroklage LM, Bierma-Zeinstra S, Benninga MA, et al. Diagnostic value of abdominal radiography in constipated children: a systematic review. *Arch Pediatr Adolesc Med* 2005;159:671–8.
16. Blethyn AJ, Jenkins HR, Roberts R, et al. Radiological evidence of constipation in urinary tract infection. *Arch Dis Child* 1995;73:534–5.
17. Zaslavsky C, da Silveira TR, Maguilnik I. Total and segmental colonic transit time with radio-opaque markers in adolescents with functional constipation. *J Pediatr Gastroenterol Nutr* 1998;27:138–42.
18. Bouchoucha M, Devroede G, Arhan P, et al. What is the meaning of colorectal transit time measurement? *Dis Colon Rectum* 1992;35:773–82.
19. Papadopoulou A, Clayden GS, Booth IW. The clinical value of solid marker transit studies in childhood constipation and soiling. *Eur J Pediatr* 1994;153:560–4.
20. de Lorijn F, van Wijk MP, Reitsma JB, et al. Prognosis of constipation: clinical factors and colonic transit time. *Arch Dis Child* 2004; 89:723–7.
21. Metcalf AM, Phillips SF, Zinsmeister AR, et al. Simplified assessment of segmental colonic transit. *Gastroenterology* 1987;92:40–7.
22. Arhan P, Devroede G, Jehannin B, et al. Segmental colonic transit time. *Dis Colon Rectum* 1981;24:625–9.
23. Collinson P. Of bombers, radiologists, and cardiologists: time to ROC. *Heart* 1998;80:215–7.
24. Gutierrez C, Marco A, Nogales A, et al. Total and segmental colonic transit time and anorectal manometry in children with chronic idiopathic constipation. *J Pediatr Gastroenterol Nutr* 2002;35:31–8.
25. Benninga MA, Buller HA, Heymans HS, et al. Is encopresis always the result of constipation? *Arch Dis Child* 1994;71:186–93.
26. Bongers ME, Voskuil WP, van Rijn RR, et al. The value of the abdominal radiograph in children with functional gastrointestinal disorders. *Eur J Radiol* 2006;59:8–13.