Constipation in (early) infancy and childhood : pathogenesis and diagnostic procedures

de Lorijn, F.

Chapter 8

The Leech method for diagnosing constipation; intra-, inter-observer variability and accuracy

Fleur de Lorijn, Rick R. van Rijn, Jarom Heijmans, Johannes B. Reitsma, Wieger P. Voskuijl, Onno D.F. Henneman, Jan A. Taminiau, and Marc A. Benninga

submitted

Abstract

Background

Inconsistent data exist concerning the value of a plain abdominal radiograph in childhood constipation. Recently, positive results were reported of a new scoring system "the Leech-method" to assess faecal loading on a plain abdominal radiograph.

Aim

To assess intra- and inter-observer variability of the Leech method and to determine the diagnostic accuracy of this score in identifying children with functional constipation.

Methods

89 consecutive children (58% boys, median age 9.8 yrs) with functional gastro-intestinal disorders (Rome II criteria) were included in the study. Based on clinical parameters, 51 fulfilled the criteria for functional constipation (FC), 7 and 31 children fulfilled the criteria for functional abdominal pain (FAP), and functional non-retentive faecal soiling (FNRFS), respectively, and were classified as controls. Intra- and inter-observer variability of the Leech method were assessed by scoring the same abdominal radiograph twice by three observers, using an interval of 4 weeks between scoring of the same radiograph. A Leech score of ≥9 (0-15) was considered as suggestive for (the diagnosis) of constipation. ROC-analysis was used to determine the diagnostic accuracy of the Leech method in separating patients with functional constipation from patients of the "control" group.

Results

The Leech-score was significantly higher in children with FC (mean: 10.1) compared to the score in patients of the "control" group (mean: 8.5; p=0.002). The intra-observer variability was high, whereas the scores between observers (inter-observer variability) differed both systematically and displayed large variability. The Leech score had poor diagnostic accuracy for diagnosing functional constipation, as indicated by the area under the ROC curve of 0.68. The cut-off value of the Leech score that leads to the lowest sum of false-positive and false-negative test results was 9 and leads to a sensitivity of 75% and a specificity of 59%.

Conclusion

The Leech scoring method to assess faecal loading on a plain abdominal radiograph proves to be of limited value in the diagnosis of functional constipation in children.

Introduction

Constipation is a common, if not universal phenomenon in childhood. The symptoms vary from a relative mild, short-lived bowel problem to severe chronic constipation with faecal impaction and the involuntary loss of faeces (encopresis). Key elements in the diagnostic process are a careful medical history and physical examination.

In addition to history and physical examination, a plain abdominal radiograph is frequently used to assess the presence of retained stool or enlargement of the colon or rectum in both the diagnostic process and treatment evaluation. Several scoring systems exist to assess the severity of constipation in children on abdominal radiographs[1,2]. However, inconsistent data are reported using the different scoring methods. The Barr and Blethyn scoring methods showed low intra- and inter-observer variability. Whereas others found both high intra- and inter-observer variability re-evaluating the Barr-score or their own scoring system[3,4].

Recently, a new scoring system was introduced to appraise faecal retention on a single abdominal radiograph[5]. The authors claimed that this method is clinically useful and reproducible in assessing childhood constipation. It is well-known however, that evaluation of an abdominal radiograph is highly subjective. Therefore, the aim of our present study was to reassess the intra- and inter-observer variability of the Leech method and to determine its accuracy in the diagnosis of functional constipation among patients with functional gastrointestinal disorders.

Methods

Study population

Between 2001 and 2003, 89 patients were enrolled in the present study. All children were referred to our tertiary paediatric gastro-intestinal outpatient clinic for the evaluation of abdominal pain, constipation or encopresis by family practitioners, paediatricians, psychiatrists and school doctors.

Clinical diagnosis

A diagnosis of functional constipation (FC) was made if at least two of the following four criteria were present: 1) defecation frequency of less than three times per week, 2) two or more episodes of encopresis per week, 3) production of large amounts of stool once per 7-30 days, and 4) the presence of a palpable abdominal or rectal mass[6].

The diagnosis of functional non-retentive faecal soiling (FNRFS) was based on the following criteria: 1) two or more encopresis episodes per week with no signs of constipation, 2) a defecation frequency of ≥3 per week, 3) no periodic passage of very large amounts of stool at least once every 7 – 30 days and 4) no palpable abdominal or rectal mass on physical examination, for a period of at least one week in the preceding 12 weeks[7]. Encopresis was defined as the voluntary or involuntary loss of loose stool in the underwear after the age of 4 years[8].

Functional abdominal pain (FAP) was defined as; at least 12 weeks of: 1) Continuous or nearly continuous abdominal pain in a school-aged child or adolescent; and 2) no or only occasional relation of pain with physiological events; and 3) some loss of daily functioning; and 4) the pain is not feigned; and 5) the patient has insufficient criteria for other functional gastrointestinal disorders that would explain the abdominal pain[7].

Children with clinical characteristics of FNRFS and FAP have little or no faecal loading on an abdominal radiograph and were therefore classified as the "control" group.

Abdominal radiograph procedure and scoring methods

Treatment with oral or rectal laxatives was discontinued in each patient for at least 4 days. Thereafter, the patient ingested 1 capsule with 10 small radiograph opaque markers on 6 consecutive days in order to determine the colonic transit time (CTT). Subsequently, a plain abdominal radiograph was taken at day 7. This radiograph was used both in the Leech method and in the CTT measurement.

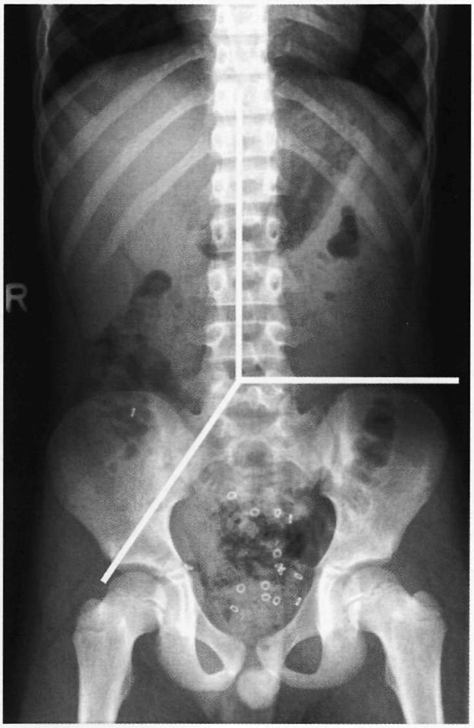

Figure 1. Plain abdominal radiograph of a child with a functional gastrointestinal disorder showing the division of the abdomen into three segments for the purpose of scoring according to the Leech-method.

Leech-score

The Leech scoring method divides[5] the colon in three segments; the right colon, the left colon and the rectosigmoid segment (figure 1). Each colonic segment is provided with a score from 0 to 5: 0 = no faeces visible, 1 = scanty faeces visible, 2 = mild faecal loading, 3 = moderate faecal loading, 4 = severe faecal loading, and 5 = severe faecal loading with bowel dilatation, resulting in a score from 0 to a maximum of 15. A total score of ≥9 is considered as suggestive for (the diagnosis) of constipation.

Colonic transit time (CTT)

We used the Bouchoucha method[9] to determine the CTT. The radiograph of day 7 was used to count the number of markers visible in the colon. The number of markers multiplied by 2.4 produced the total CTT in hours. Localization of markers and CTT were calculated according to a previously described formula[10]. The normal range for total transit time was based on the upper limits (mean+2SD) from a study in healthy children[11]. Based on this study, a CTT of more than 62 hours was considered delayed.

Observers

Using the Leech method, the same radiograph was scored independently by three experienced doctors: 1 fifth year radiology resident (ODFH), 1 paediatric radiologist (RRvR) and 1 senior paediatric gastroenterologist (JAT). No clinical information about the patient was made available to the observers. After 4 weeks, the same radiograph was scored a second time by the three observers to assess intra-observer variability. Before the first scoring of the radiographs and before the second scoring after 4 weeks, the Leech method was discussed intensively among the 3 observers.

CTT times were assessed once by a single observer (FdL). It was assumed that the counting of radio-opaque markers would not lead to intra- or inter-observer variation.

Statistical analyses

Parametric and non-parametric tests were used to compare general characteristics, Leech scores, and CTT values between patients diagnosed with FC and the control group (FNRFS and FAP).

Intra-observer variability was assessed by calculating the mean in differences between the first and second score for each observer and standard deviation (SD). This SD was used to calculate the limits of agreement defined as the mean of the difference in scores +/-1.96*SD[12]. 95% of the differences in repeated scores are expected to lie within these limits.

A receiver operating characteristic (ROC) plot was made showing how pairs of sensitivity and specificity vary using the full range of cut-off values of the Leech and CTT method. The area under the ROC curve (AUC) was used as a single indicator of diagnostic accuracy. The AUC can be interpreted as the probability that a randomly chosen case (FC) has a higher Leech score or CTT value than a randomly chosen control (FNRFS or FAP). The perfect test has an AUC of 1. A non-parametric test was used to compare the AUC of the Leech method to the AUC of the CTT method, acknowledging that we had paired observations, e.g. CTT and Leech values were measured on the same patient[13]. Furthermore, we calculated the sensitivity and specificity belonging to the cut-off value of the Leech score that would lead to the lowest number of misclassifications (sum of false-positives and false-negatives), also known as the Youden index[14].

Results

Baseline characteristics

We included a series of 89 patients (58% boys, median age 9.8 years) with functional gastrointestinal disorders in this study. Within this cohort, a total of 52 children fulfilled the clinical criteria of functional constipation (FC), 31 children had functional non-retentive faecal soiling (FNRFS), and 6 children were diagnosed having functional abdominal pain (FAP). For the remaining analyses, FNRFS and FAP patients were grouped together as the "control" group.

The baseline characteristics of the study population are presented in table 1. There was no significant difference in age and gender between the children with functional constipation (FC) and the control group. As expected, the median defecation frequency at intake was lower in children with FC compared to controls (p=0.0002). A palpable abdominal (15%) or rectal mass (36%) was a common finding at physical examination in the FC group. Encopresis was common in both the constipation (81%) and the control group (84%).

Figures 2 and 3 show the distribution of Leech-scores using the first score and CTT measurements in children with FC and in the control group. A higher mean Leech-score (10.1 versus 8.5, p=0.002, t-test) and CTT (92 versus 37 hours, p<0.0001, t-test) was found in children with FC compared to controls.

Table 1 Baseline characteristics.

	Constipation (n=52)	Control group (n=37) (FNFRS and FAP)
Number of boys (%)	28 (54%)	24 (65%)
Mean age at intake in years (SD)	10.0 (2.8)	10.3 (2.7)
Median age of onset of symptoms (P25-P75)	4.0(2.4-5.0)	4.0(4.0-6.0)
Median defecation frequency/week (P25-P75)	2.0(0.5-3.0)*	7.0(4.0-10.0)
Encopresis	81%	84%
Large amounts of stool	65% **	0%
Palpable abdominal mass	15%	0%
Palpable rectal mass	36% ***	0%

FNFRS=functional non-retentive faecal soiling; FAP= functional abdominal pain
*p=0.0001, Mann-Whitney U test; **p=0.0005 and ***p=0.0002, chi-square

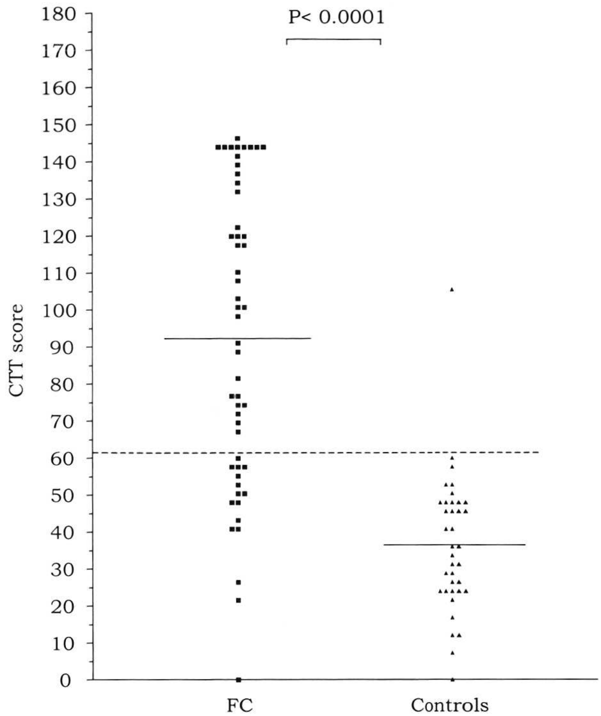

Figure 3. Scatter plot of CTT in children with functional constipation and in the control group (functional non-retentive faecal soiling and functional abdominal pain). Horizontal solid lines indicate the mean value. Dashed horizontal line bar indicates the CTT of 62 above which a patient is considered constipated.

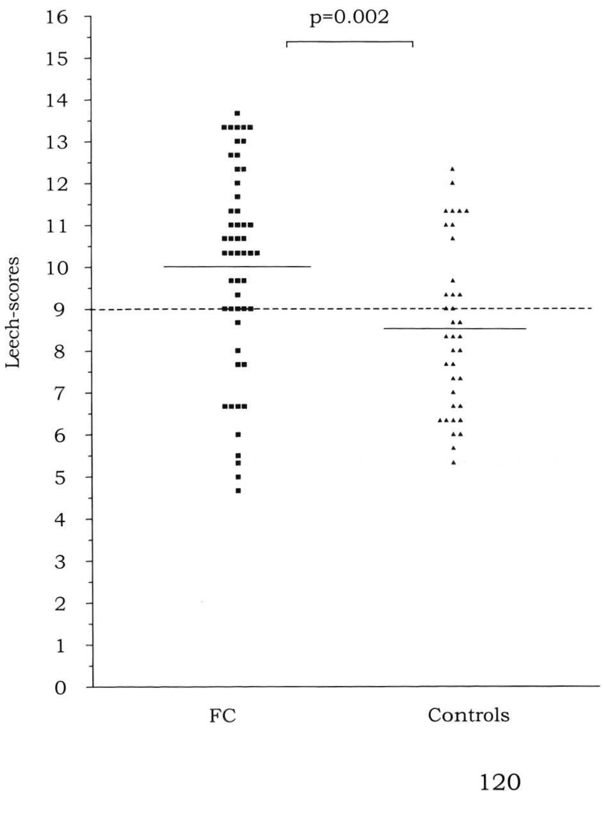

Figure 2. Scatter plot of Leech-scores using the first score in children with functional constipation and in the control group (functional non-retentive faecal soiling and functional abdominal pain). Horizontal solid lines indicate the mean value. Dashed horizontal line indicates a Leech-score of 9 above which a patient is considered constipated.

Diagnostic accuracy of Leech-method versus CTT-method

Figure 4 shows the receiver operating characteristic curve (ROC curve) for both the Leech and CTT-method. These ROC curves illustrate how sensitivity and specificity vary using different cut-off values. CTT has a better overall accuracy because its ROC curve lies above and to the left of the Leech curve. The AUC for the Leech method was lower than the AUC for the CTT method (0.68 versus 0.90), indicating to a poorer accuracy for the Leech method. This difference was highly significant (p=0.00015).

The cut-off value for the Leech method in our sample, producing the highest Youden index, was comparable to 9 which is the cut-off value used in the literature and leads to a sensitivity of 75% and specificity of 59%. For CTT, this optimal cut-off point was 54 hours in our sample, leading to a sensitivity of 79% and specificity of 92%. The most frequently used cut-off value for CTT in the literature is 62 hours leading to a sensitivity of 71% and specificity of 95%.

Using the cut-off values of the literature, the Leech method showed a positive predictive value (PPV) and a negative predictive value (NPV) of 72% and 63% respectively. On the other hand, CTT showed a comparable PPV of 69% and a higher NPV of 97%.

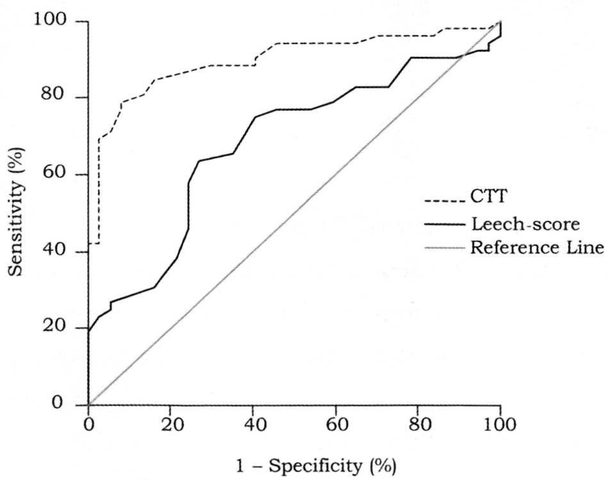

Figure 4. ROC curves for the Leech-method and CTT-method for diagnosing functional constipation among children with functional gastrointestinal disorders. The Leech score used in this figure is the mean score of the three observers.

Intra-observer variability of Leech method

The same abdominal radiograph was scored twice by all three observers. Two observers had significantly higher or lower scores (bias) in their repeat scoring of the same radiograph. The third observer had the largest difference with the second Leech score being on average 1.6 points higher than the first ($p<0.0001$), while for observer 2 the second score was on average 0.7 points lower ($p=0.005$). The first observer had no systematic difference between the two scores ($p=0.89$) (table 2).

In addition, differences in repeated scores from the same observer showed large variability (measurement error) even after accounting for a systematic difference.

The standard deviation of differences between the first and second score for all three observers is shown in table 2. These standard deviations indicate that 5% of second scores by the same observer could be more than 4.4 points higher or lower than the first score. These "limits of agreement" are large in comparison to the scale on which the Leech score is measured (minimum score is 0, maximum is 15).

Table 2 Intra-observer variability of all three observers.

Difference between first and second score	Observer 1	Observer 2	Observer 3
Systematic difference:			
mean value	0.7	0.03	-1.6
(95% CI)	(0.2 to 1.2)	(-0.4 to 0.5)	(-2.0 to-1.3)
(p-value)	(0.89)	(0.0005)	(<0.0001)
Variability:			
standard deviation	2.2	2.2	1.5
limits of agreement	-6.0 to +5.0	-7.0 to +7.0	-5.0 to +3.0

Inter-observer variability of Leech method

Figures 5-7 show the differences in Leech scores and limits of agreement between pairs of observers using the first score. Observer 3 scored consistently lower than observer 1 (mean of differences 2.7, $p<0.0001$) and observer 2 (mean of differences 2.9, $p<0.0001$). No systematic difference was found between observer 1 and 2. Limits of agreement (1.96*SD) were wide for all pairs of observers. In 5% of the cases, the Leech score of the same patient could differ 4 points or more by using a different observer.

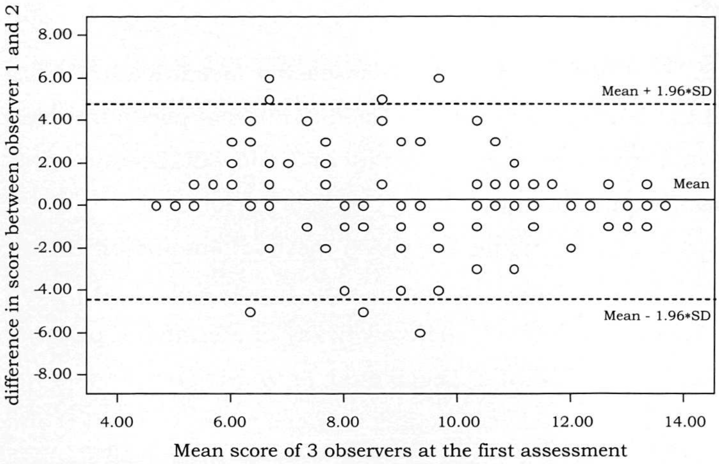

Figure 5. Mean difference and limits of agreement between scores from observer 1 and 2.
95% of the differences in scores are expected to lie within this interval.

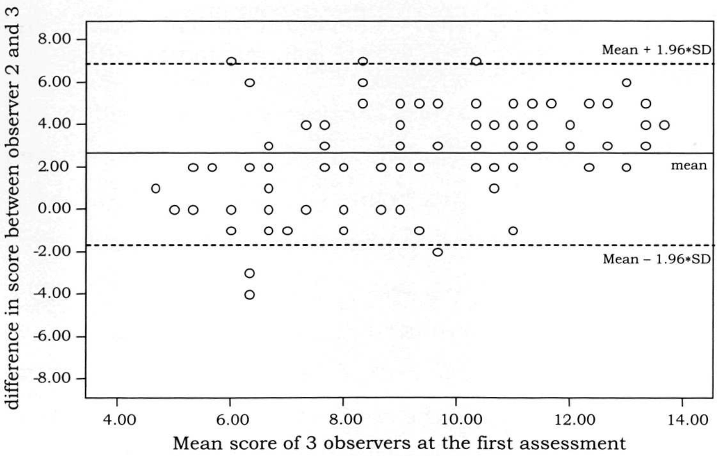

Figure 6. Mean difference and limits of agreement between scores from observer 2 and 3.

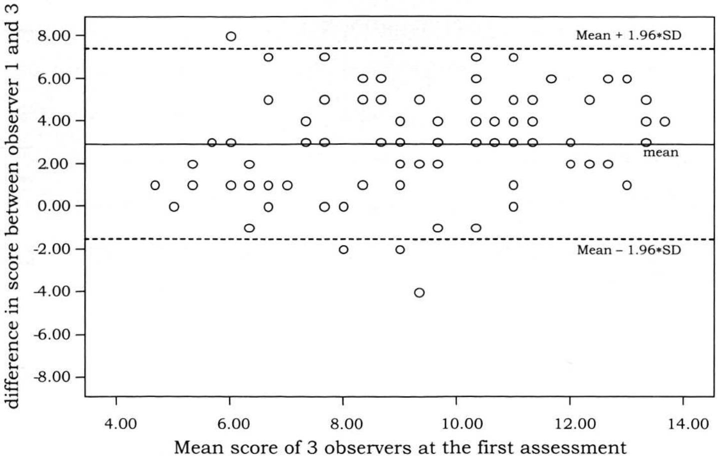

Figure 7. Mean difference and limits of agreement between scores from observer 1 and 3.

Discussion

This study clearly shows that the Leech method assessing faecal loading on a plain abdominal radiograph is of limited value in the diagnosis of paediatric constipation. A high intra- and inter-observer variability and a low diagnostic accuracy (sensitivity and specificity) were found using this scoring method.

In accordance with earlier studies using the Leech method, significant higher scores of the radiographs were found in patients with functional constipation compared to controls (FNRFS and FAP)[5,15]. The accuracy of identifying patients with FC using the proposed cut-off point of Leech et al. (≥ 9) was poor, indicating that there was considerable overlap in Leech scores between patients with FC and patients from the control group (figure 1 and 2). Only 72% of all children with a Leech-score ≥ 9 were diagnosed as FC (PPV). While 37% of the children with a Leech-score <9 were also diagnosed as FC (NPV). There are several reasons for the poor diagnostic accuracy of the Leech-scoring method. Scoring a single plain abdominal radiograph is a 'one-shot observation'. The degree of faecal loading has daily fluctuation and depends on food intake, will of the child to defecate and whether or not the child went to the toilet just before the abdominal radiograph was taken. An earlier study showed no correlation in the degree of faecal loading when comparing two abdominal radiographs, taken 4 days apart, in the same child with functional constipation (while not receiving laxative treatment)[3].

At intake, a palpable rectal mass was found in 36% of the patients with functional constipation and in 0% of the patients out of the control group. Because the time between rectal examination and the abdominal radiography was at least 7 days, we were unable to reliably correlate these findings. A significant association, likelihood ratio (LR) of 1.6 (95% c.i. 1.2 to 2.0), between stool on rectal examination and radiological constipation was reported in a study evaluating 251 children presenting with abdominal pain to a paediatric emergency department[16]. This finding was however, not confirmed by Rockney et al in a retrospective case study in 60 children with encopresis, LR of 1.5 (95% c.i. 0.8 to 2.3)[17]. We strongly believe however, that a positive rectal examination is associated with functional constipation, in which case an abdominal radiograph is not necessary to make that diagnosis.

Another determinant of diagnostic accuracy is observer agreement. If the reproducibility of a test method is poor within or between observers, it will negatively affect diagnostic accuracy[18]. The inter- and intra-reliability of scoring systems for

faecal loading on abdominal radiography varies among studies from poor to excellent[1-4]. We found large intra- and inter-observer variability for the Leech method. The original report by Leech et al. also found a large inter-observer variability, but in contrast to our finding, a low intra-observer variability. It has been suggested that experienced radiologists score more reliable than less experienced colleagues. In our study, however all observers were skilled and experienced in reading plain abdominal radiographs. Quality of the radiograph, reporting facilities, time and possibly personality type are other factors that may influence the observer's interpretation of a radiograph. On the other hand, adequate clinical information might improve reproducibility and accuracy when reporting a radiograph[4]. Others suggested the use of two observers to reduce interpretation error[19]. This however, is in light of the increasing workload of clinical radiologists a non-viable option in daily routine.

Assessing CTT with radio-opaque markers is another method to score abdominal radiographs and prolongs the observation period to a week. In accordance with earlier findings in both adults and children, normal CTTs were found in 31% of children with functional constipation[20-22]. De Lorijn et al showed that a delay of colonic transit less than 100 hours was not predictive of treatment outcome[21]. Only patients with CTT >100 hours had a less favourable outcome at 1 year[21]. Others have found a significant correlation between the severity of paediatric constipation and the delay in colonic transit time[20,23]. CTTs were significantly prolonged in children with a defecation frequency less than 1 per week and /or an encopresis frequency more than twice daily.

These findings stress that an abdominal radiograph with or without markers is of limited clinical value in children with functional constipation. A bowel diary, in which defecation frequency and encopresis frequency are reported by the child and the parents, and physical examination, is sufficient to make the diagnosis. Furthermore, defecation- and encopresis frequency are ideal parameters to evaluate success of treatment.

An abdominal radiograph might only be helpful in obese patients, if a child refuses a rectal examination, or if there is suspicion of sexual abuse[25]. In combination with radio-opaque markers, an abdominal radiograph is also useful to differentiate between children with functional constipation and functional non-retentive faecal soiling. In this latter group, medical history, physical examination in combination

with normal CTT confirms the diagnosis FNRFS which needs a different treatment approach[25,26].

In conclusion, assessing faecal loading on a plain abdominal radiograph using the Leech scoring method proves to be of limited value in the diagnosis of constipation because of its high observer variability and low diagnostic accuracy. History and physical examination remain the cornerstone in the diagnostic work-up of children with functional gastrointestinal disorders.

References

1. 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 1979;18:674-9.

2. Blethyn AJ, Verrier JK, Newcombe R et al. Radiological assessment of constipation. Arch Dis Child 1995;73:532-3.

3. 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.

4. Tudor GR, Finlay D, Taub N. An assessment of inter-observer agreement and accuracy when reporting plain radiographs. Clin Radiol 1997;52:235-8.

5. 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.

6. Loening-Baucke V. Biofeedback therapy for fecal incontinence. Dig Dis 1990;8:112-24.

7. Rasquin-Weber A, Hyman PE, Cucchiara S et al. Childhood functional gastrointestinal disorders. Gut 1999;45:60-8.

8. van der Plas RN, Benninga MA, Redekop WK et al. Randomised trial of biofeedback training for encopresis. Arch Dis Child 1996;75:367-74.

9. Bouchoucha M, Devroede G, Arhan P et al. What is the meaning of colorectal transit time measurement? Dis Col Rect 1992;35:773-82.

10. Metcalf AM, Phillips SF, Zinsmeister AR et al. Simplified assessment of segmental colonic transit. Gastroenterology 1987;92:40-7.

11. Arhan P, Devroede G, Jehannin B et al. Segmental colonic transit time. Dis Col Rect 1981;24:625-9.

12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.

14. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.

15. Afzal N, Murch S, Thirrupathy K et al. Constipation with acquired megarectum in children with autism. Pediatrics 2003;112:939-42.

16. Beckmann KR, Hennes H, Sty JR et al. Accuracy of clinical variables in the identification of radiographically proven constipation in children. WMJ 2001;100:33-6.

17. Rockney RM, McQuade WH, Days AL. The plain abdominal roentgenogram in the management of encopresis. Arch Pediatr Adolesc Med 1995;149:623-7.

18. Quinn MF. Relation of observer agreement to accuracy according to a two-receiver signal detection model of diagnosis. Med Decis Making 1989;9:196-206.

19. Boyd NF, Wolfson C, Moskowitz M et al. Observer variation in the interpretation of xeromammograms. J Natl Cancer Inst 1982;68:357-63.

20. 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.

21. 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.

22. 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.

23. Wald A. Colonic transit and anorectal manometry in chronic idiopathic constipation. Arch Int Med 1986;146:1713-6.

24. Eidlitz-Markus T, Mimouni M, Zeharia A et al. Occult constipation: a common cause of recurrent abdominal pain in childhood. Isr Med Assoc J 2004;6:677-80.

25. Benninga MA, Buller HA, Heymans HS et al. Is encopresis always the result of constipation? Arch Dis Child 1994;71:186-93.

26. Voskuijl WP, Heijmans J, Heijmans HS et al. Use of Rome II criteria in childhood defecation disorders: applicability in clinical and research practice. J Pediatr 2004;145:213-7.