Abdominal Radiology (2019) 44:821–827
https://doi.org/10.1007/s00261-018-1869-5

HOLLOW ORGAN GI

Radiographic stool quantification: an equivalence study of 484 symptomatic and asymptomatic subjects

Omar Khan[1,2] · Prasad R. Shankar[1,2,7] · Adish D. Parikh[3] · Richard H. Cohan[1] · Nahid Keshavarzi[4] · Shokoufeh Khalatbari[4] · Richard J. Saad[5] · Matthew S. Davenport[1,2,6]

Published online: 14 December 2018

Abstract

Purpose To determine if symptomatic patients referred for radiographic stool quantification have equivalent stool burden to asymptomatic patients.

Method This was an IRB-approved HIPAA-compliant retrospective equivalence cohort study. An a priori equivalence power calculation was performed. Consecutive abdominal radiographs performed in adult outpatients with bloating, constipation, diarrhea, or abdominal pain to assess "fecal loading" [$n=242$ (fecal cohort)] were compared to those performed in asymptomatic adult outpatients to assess "renal stones" [$n=242$ (renal cohort)]. Radiographs were randomized and reviewed by two blinded independent abdominal radiologists. Exclusion criteria, designed to avoid unblinding, included urinary tract calculi ≥ 0.5 cm, multiple urinary tract calculi, and ureteral stent(s). Readers scored all radiographs ($n=484$) for stool burden using validated Leech criteria [scale: 0 (none) to 15 (extreme diffuse)]. Mean Leech scores and 95% confidence intervals were calculated. Multivariable generalized linear modeling was performed to adjust for baseline medication use, age, and gender. The adjusted parameter estimate was used to test for equivalence in the mean difference between cohorts using Schuirmann's method of two one-sided t-tests. Inter-reader agreement was assessed with intraclass correlation coefficients.

Results Overall mean Leech scores for fecal [6.9 (95% CI 6.7, 7.2)] and renal [7.3 (95% CI 7.1, 7.5)] cohorts were equivalent within a margin of 0.75 (adjusted mean difference: -0.4 [90% CI -0.7, -0.04]; p value $=0.02$). Inter-reader agreement was good [ICC: 0.62 (95% CI 0.56, 0.68)].

Conclusion Radiographic stool quantification produces equivalent results in symptomatic and asymptomatic adults and is of uncertain value.

Keywords Constipation · Fecal loading · Utilization · Abdominal radiograph · Value

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00261-018-1869-5) contains supplementary material, which is available to authorized users.

✉ Prasad R. Shankar
pshankar@med.umich.edu

[1] Department of Radiology, Michigan Medicine, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

[2] Michigan Radiology Quality Collaborative, Michigan Medicine, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

[3] University of Michigan Medical School, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

[4] Michigan Institute for Clinical and Health Research (MICHR), University of Michigan, 1600 Huron Parkway, Building 400, Ann Arbor, MI 48105, USA

[5] Department of Gastroenterology, Michigan Medicine, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

[6] Department of Urology, Michigan Medicine, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

[7] Department of Radiology, University Hospital, University of Michigan, 1500 E. Medical Center Drive, B1 D530H, Ann Arbor, MI 48109, USA

Introduction

An estimated 15% (range 2–27%) of Americans suffer from constipation [1–3]. Functional constipation generally is diagnosed by objective and subjective clinical factors as defined by the Rome criteria, now in its fourth iteration (Rome IV) [4, 5]. The role of imaging in assisting with the diagnosis and quantification of constipation is a question of active interest for both radiologists and gastroenterologists [6–8]. Although commonly performed [8], radiographic stool quantification is not a component of the Rome IV criteria [4].

Radiographic stool quantification has been studied extensively in pediatric patients, with overall questionable-to-doubtful benefit [6, 7, 9]. It has been studied sparsely in the adult population [8]. Despite the lack of clear evidence supporting its use, in 2018, Reber et al. [8] found a steady increase in utilization of radiographs for adult stool volume quantification at Mayo Clinic from 2004 to 2014. Given increased utilization and absence of supporting data, there is a need to test the validity of this practice. One way to do so would be to compare the stool burden in symptomatic patients to those of a control cohort undergoing radiography for a non-gastrointestinal indication. Based on anecdotal clinical experience, we hypothesized that the stool burden in these cohorts would be equivalent. If so, it would challenge the validity of using these examinations in the diagnosis and management of constipation. The purpose of our study was to determine if symptomatic patients referred for radiographic stool quantification have equivalent stool burden to asymptomatic patients.

Methods

Institutional review board (IRB) approval was obtained for this Health Insurance Portability and Accountability Act-compliant retrospective observational 2-arm equivalence cohort study. The requirement for written informed consent was waived. STROBE criteria (i.e., STrengthening the Reporting of OBservational studies in Epidemiology) were followed.

Study population

The study population flow diagram is provided in Fig. 1. Ordering clinical indications were manually identified in the

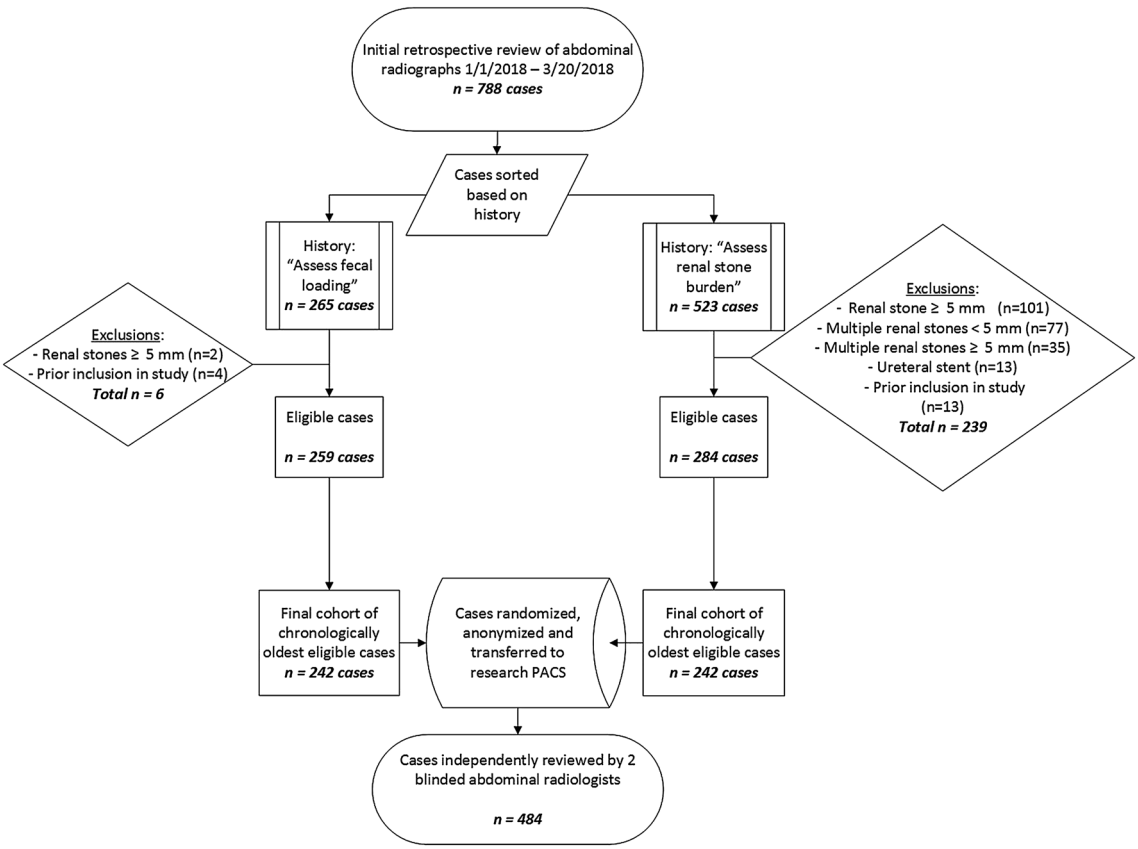

Fig. 1 Study flow diagram

radiology information system for consecutive adult patients (≥ 18 years old) who underwent outpatient abdominal radiography at a quaternary academic medical center from 1/1/2018 to 3/20/2018. Inclusion criteria were radiography performed in symptomatic outpatients (bloating, constipation, diarrhea, or abdominal pain [not mutually exclusive]) for "fecal loading" ($n = 265$) or in asymptomatic outpatients (from a gastrointestinal perspective) undergoing radiography for "renal stones" ($n = 523$). One study team member manually reviewed all imaging indications, imaging data, and finalized radiology reports. Exclusion criteria included (Fig. 1) prior inclusion in the study, renal stone(s) ≥ 5 mm, multiple renal stones, and ureteral stent(s). Exclusion criteria were designed to prevent potential inadvertent unblinding of the study cohorts during image review. Following exclusion criteria, there were 584 eligible subjects (fecal cohort: $n = 259$, renal cohort: $n = 284$). Based on an a priori equivalence power calculation, the 242 chronologically oldest consecutive eligible examinations in each cohort were selected (Fig. 1).

Radiography protocol

All radiographic examinations ($n = 484$) were performed at the study institution using one of the following X-ray machines as part of a routine outpatient clinical practice: GE Definium 8000 (GE Healthcare, Chicago, IL), GE Discovery 656 (GE Healthcare, Chicago, IL), Siemens Axiom Aristos MX (Siemens Healthineers, Erlangen, Germany), Siemens Luminos TF (Siemens Healthineers, Erlangen, Germany).

All examinations were performed in the supine projection with imaging from diaphragm through urinary bladder. No protocol alterations were made based on the ordering indication. Both cohorts were imaged with the same imaging parameters. A single radiograph was obtained in most cases. If the patient was too large to be imaged with a single radiograph, additional radiographs were obtained in the supine projection to cover the relevant field of view.

Data collection and image transfer

Gender, age, and relevant medication history were obtained for the full study population through a retrospective review of the electronic medical record by one study team member. Specific symptoms preceding radiography and any change in laxative regimen within 30 days following radiography were determined for the fecal cohort. The Electronic Medical Research Engine (EMERSE) tool was used to ensure fidelity and improve efficiency of data collection [10]. For medications prescribed on an "as needed" basis, charts were manually reviewed to evaluate for evidence of medication use within the 30-day window (e.g., evidence of refills, clinical notes, telephone encounters).

Medication use prior to imaging was grouped by opioid use, anti-cholinergic medication use, and laxative use (Appendix A). Medication use prior to imaging was considered positive if consumed within 30 days prior to radiography. The 30-day time period was selected to mirror institutional clinical practice standards. Search terms for opioid medications were based on guidelines from the American Addiction Centers [11]. Commonly encountered anti-cholinergic and laxative medication lists were created through focused chart audit and consultation with an attending gastroenterologist (13 years of experience) study team member with expertise in functional bowel disorders and constipation.

The final study group of 484 abdominal radiographs were de-identified, randomized, assigned designated study ID numbers, and sent to a dedicated research Picture Archiving and Communications (PACS) workstation for review. Anonymization and assignment of study ID numbers on the research PACS was automated by using the Radiological Society of North America's (RSNA) Medical Imaging Resource Community (MIRC) software [12].

Image review

The Leech scale, developed in the pediatric population [13], divides the colon into three zones: right hemicolon, left hemicolon, and rectosigmoid colon. Each of these zones is scored on a scale of 0–5 based on the following criteria: 0 = no feces, 1 = scant feces, 2 = mild fecal load, 3 = moderate fecal load, 4 = severe fecal load, 5 = severe fecal load and distension. The scores of all three zones are summed to obtain the Leech score (scale 0–15), with scores ≥ 9 representing constipation [13].

Images were reviewed independently by two fellowship-trained abdominal radiologists with 4 (Reader 1) and 33 (Reader 2) years of faculty experience who were blinded to exam indication, patient-level data, and each other. Prior to participating in the study, each reading radiologist was familiarized with the Leech scale and given the opportunity to ask questions about how it compared to the qualitative scoring used in clinical practice (e.g., mild fecal loading). The reading radiologists were accustomed to interpreting approximately 80 abdominal radiographs on their daily gastrointestinal radiology rotation, of which approximately 15–25 were ordered for stool quantification. Each radiologist quantified the radiographic stool burden on each examination using the validated Leech scale [13–15]. Both readers completed image review within a 2-month period.

Primary outcome and sample size

The primary outcome was difference in mean Leech score between cohorts. An a priori equivalence power analysis

was performed. Based on exploratory data and the work of Lorijn et al. [16], the anticipated mean Leech score for the fecal cohort was 10.1 with a standard deviation of 2.5. An equivalence margin (d) of 0.75 was selected because it represented a margin of 5% on the 0–15 Leech scale. Clinically, this margin was considered equivalent because it was less than 1 scoring unit difference between cohorts. Based on the above assumptions, alpha of 0.05, and power of 90%, this rendered a sample size requirement of 242 subjects per cohort (484 total).

Data analysis

Continuous data were presented as median and interquartile range (IQR) while categorical data were presented as counts and percentages. Mann–Whitney *U* test was performed to assess the age differences between cohorts. The distributions of categorical variables (gender and medication use prior to imaging) were assessed and compared between cohorts using Chi-squared (χ^2) tests.

Mean Leech scores and 95% confidence intervals for fecal and renal cohorts were determined per reader and overall. Adjusted mean differences in Leech scores between cohorts were determined through a multivariable linear regression model controlling for cohort differences in age, gender, opioid use prior to imaging, anti-cholinergic medication use prior to imaging, and laxative use prior to imaging (5 variables). Each of these variables was independently tested using repeated measure multivariable regression models for any potential significant association with Leech scores.

Equivalence testing of adjusted mean Leech score difference between fecal and renal cohorts was performed using Schuirmann's method of two one-sided tests *t* tests (TOST) [17], with an equivalence margin (d) of 0.75. A result was considered statistically significant for equivalence if the 90% confidence interval of the parameter estimate excluded the equivalence margin (0.75), yielding a $p < 0.05$ for both tails of the 2 one-sided *t* tests.

As a secondary outcome, single-measure two-way random intraclass correlation coefficients (ICC) for absolute agreement were performed to assess inter-reader agreement within the study [18–20]. ICC results were stratified qualitatively and quantitatively (poor: < 0.40, fair: 0.40–0.59, good: 0.60–0.74, excellent: 0.75–1.00) [21].

Power analysis was performed with PASS 15 Power Analysis and Sample Size Software (2017; NCSS, LLC., Kaysville, UT). Statistical analysis was performed with SPSS software (version 25; IBM Corporation, Armonk, NY) and SAS (version 9.4; SAS Institute, Inc., Cary NC).

Results

A total of 484 exams were reviewed by each reader (fecal cohort: 242, renal cohort: 242). Study population details are summarized in Table 1. For the fecal cohort, the most common indications for imaging were "constipation" (89.5% [214/239]) and "abdominal pain" [67.8% (162/239)] (Table 1).

Both cohorts had similar median age (fecal cohort: 49.0 years [IQR: 31], renal cohort: 49.5 years [IQR:22], $p = 0.914$). There were statistically significant differences between cohorts for gender and medication use (Table 1). Age was the only covariable with a predictive relationship with Leech score on multivariable regression analysis ($\beta = 0.0153$, $p = 0.003$). Gender ($p = 0.413$), opioid use ($p = 0.511$), anti-cholinergic use ($p = 0.133$), and laxative use ($p = 0.830$) were not significantly associated with Leech score. Following radiography, 67.8% (162/239) in the fecal cohort had their laxative regimen changed (3 were lost to follow-up).

Table 1 Study population details

Variable	Fecal cohort	Renal cohort	*p* value
Median age (25th–75th percentile range)	49 (34–65)	49.5 (38–60)	0.914
Gender, % Female	80.6% (195/242)	54.5% (132/242)	< 0.0001
% taking opioids (prior to radiograph)	24.4% (59/242)	42.6% (103/242)	< 0.0001
% taking anti-cholinergic (prior to radiograph)	28.1% (68/242)	15.7% (38/242)	0.001
% taking laxatives (prior to radiograph)	47.5% (115/242)	9.5% (23/242)	< 0.0001
Indication for stool quantification radiography[a]			
Bloating	40.6% (97/239)	N/A	
Constipation	89.5% (214/239)	N/A	
Diarrhea	45.2% (108/239)	N/A	
Abdominal pain	63.6% (152/239)	N/A	
% change in laxative regimen (after radiograph)	67.8% (162/239)	N/A	

p values refer to comparisons of the two cohorts

[a]History was unavailable in 3 cases, indications are not mutually exclusive

Aggregate Leech scores (scale 0–15) were equivalent within a margin of 0.75 controlling for age, gender, opioid use, anti-cholinergic use, and laxative use [fecal cohort: 6.9 (95% CI 6.7, 7.2), renal cohort: 7.3 (95% CI 7.1, 7.5), adjusted difference: −0.4, $p=0.021$] (Table 2). Leech score adjusted differences (controlling for covariables) for Reader 1 alone were equivalent [adjusted difference: −0.3 (90% CL −0.6, −0.05) $p=0.006$). Leech scores for Reader 2 slightly exceeded the 0.75 equivalence threshold (adjusted difference: −0.4 [90% CI −0.8, 0.02), $p=0.091$]. Post hoc equivalence calculations showed that Leech scores for Reader 2 were equivalent to a threshold of 0.85 ($p=0.04$).

Reader scoring distributions are summarized in Fig. 2. Inter-reader agreement was good overall [ICC: 0.62 (95% CI 0.56, 0.68)] and within cohorts (fecal cohort ICC: 0.62 [95% CI 0.5, −0.69], renal cohort ICC: 0.63 [95% CI 0.54, 0.70) (Table 3).

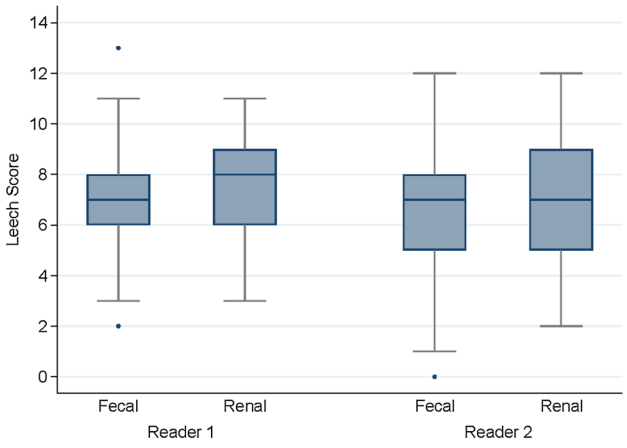

Fig. 2 Box and whisker plots of Leech score distributions for fecal loading and renal indications by reader

Discussion

We found that stool quantification produces equivalent results in symptomatic adult outpatients imaged for "fecal loading" and asymptomatic (from a gastrointestinal perspective) adult outpatients imaged for a non-gastrointestinal indication. For both readers in our study, validated Leech score measurements for stool volume between cohorts were within 1 Leech scale point (scale: 0–15), and there was good inter-reader agreement (ICC: 0.62). These findings support the hypothesis that there are similar baseline radiographic stool volumes in symptomatic and asymptomatic adult patients, and the results challenge the utility of radiography as an effective tool in the diagnosis and management of patients with suspected constipation-related symptoms.

Prior studies evaluating radiographic stool content have been performed predominantly in the pediatric population and with smaller cohorts. Lorijn et al. studied 89 children (52 with functional constipation, 6 with functional abdominal pain, 31 with functional non-retentive fecal incontinence) and found that the Leech score had poor diagnostic accuracy for predicting constipation (area under ROC curve: 0.68 [95% CI 0.58–0.80]) [16]. The authors [16] concluded that the Leech scoring system had limited utility in the diagnosis of functional constipation in children. In a systematic review of 6 publications, Reuchlin-Vroklage et al. found conflicting and insufficient evidence to support radiography for use in the diagnosis of constipation in children [9]. Cowlam et al. performed a prospective study of 100 adult patients comparing fecal loading on plain radiography and radio-opaque colonic marker transit time and deduced poor correlation between the two (Spearman's Rho rank correlation 0.26–0.31) [7]. Our results suggest similar findings that radiographic stool volume may be independent of clinical symptoms.

Despite the limited data supporting the practice of stool quantification, radiographs are commonly ordered for this indication. During the study period at our institution (a little less than 3 months), 265 such examinations were performed. At the Mayo Clinic, Reber et al. found a 56% growth in constipation-related radiography volume from 2004 to 2014, with most imaging performed in female patients [8]. Regardless of conflicting and negative results in prior studies [7, 9, 16], ordering providers at their [8] institution "somewhat agreed" that radiographs were "helpful in determining

Table 2 Radiographic stool quantification by Leech scoring

Reader	Fecal	Renal	Adjusted difference	*p* value
Both	6.9 (6.7, 7.2)	7.3 (7.1, 7.5)	−0.4 (−0.7, −0.04)	0.021*
Reader1	7.1 (6.9, 7.3)	7.5 (7.3, 7.7)	−0.3 (−0.6, −0.05)	0.006*
Reader2	6.7 (6.4, 7.1)	7.1 (6.8, 7.4)	−0.4 (−0.8, 0.02)	0.091

Data are aggregate and per-reader mean Leech scores with 95% confidence intervals. Adjusted differences are provided with 90% confidence intervals. p value < 0.05 refers to equivalence between fecal and renal cohorts with an equivalence margin of 0.75

*Significant p value (< 0.05) denoting equivalence by TOST with equivalence margin of 0.75

Table 3 Inter-rater agreement for radiographic stool quantification by Leech scoring

Cohort	ICC
All cases ($n=484$)	0.624 (0.561, 0.679)
Fecal cohort ($n=242$)	0.615 (0.527, 0.690)
Renal cohort ($n=242$)	0.629 (0.543, 0.702)

Data are intraclass correlation coefficients and 95% confidence intervals

management" and found "quantitation of stool burden within the radiology report helpful." The same ordering providers [8] also "tend[ed] to agree that the increasing use of radiographs for constipation [was] appropriate." Based on the results of our study, it would appear that radiographs performed in symptomatic patients with the explicit purpose of assessing stool burden will reveal an equivalent stool burden as a cohort imaged for an unrelated indication. Therefore, the utility of this practice is questionable.

Our retrospectively derived cohorts had significant differences in gender and medication use prior to imaging that may have affected our results. Specifically, the gender distribution of the renal cohort was dissimilar to the typical population seeking care for constipation at our institution. However, medication use and gender did not have significant effects on Leech scores in either cohort. While medication use is a commonly attributed cause of clinically symptomatic constipation, to our knowledge, the impact of medication on radiographic fecal volume in adults has not been previously described. Additionally, medication differences, age, and gender were controlled in the multivariable models used to calculate cohort-specific Leech scores. One of the readers did not reach the 0.75 equivalence threshold; however, post hoc equivalence was shown for this reader at a 0.85 level (still within a clinically negligible margin) and absolute differences for this reader indicated less stool burden in the fecal cohort compared to the renal cohort. Few adult radiologists report Leech scores when evaluating abdominal radiographs in clinical practice. However, this scale has previously been used in the literature as a way to quantify stool burden and had good inter-rater agreement in our study population.

In summary, radiographic stool quantification produces equivalent results in symptomatic and asymptomatic adults, questioning the value of these examinations in the evaluation of constipation. Future prospective studies are needed to assess the role and necessity of imaging in clinical management and patient-reported outcomes. Additional efforts are needed to optimize decision support and guide appropriate imaging (if any) in patients with suspected constipation.

Funding Unfunded study.

References

1. Lacy BE, Levenick JM, Crowell M (2012) Chronic constipation: new diagnostic and treatment approaches. Therapeutic advances in gastroenterology 5 (4):233-247. https://doi.org/10.1177/1756283x12443093
2. Vazquez Roque M, Bouras EP (2015) Epidemiology and management of chronic constipation in elderly patients. Clinical interventions in aging 10:919-930. https://doi.org/10.2147/cia.s54304
3. Higgins PD, Johanson JF (2004) Epidemiology of constipation in North America: a systematic review. The American journal of gastroenterology 99 (4):750-759. https://doi.org/10.1111/j.1572-0241.2004.04114.x
4. Lacy BE, Mearin F, Chang L, Chey WD, Lembo AJ, Simren M, Spiller R (2016) Bowel Disorders. Gastroenterology 150 (6):1393-1407.e1395. https://doi.org/10.1053/j.gastro.2016.02.031
5. Simren M, Palsson OS, Whitehead WE (2017) Update on Rome IV Criteria for Colorectal Disorders: Implications for Clinical Practice. Current gastroenterology reports 19 (4):15-15. https://doi.org/10.1007/s11894-017-0554-0
6. National Collaborating Centre for Ws, Children's H (2010) National Institute for Health and Clinical Excellence: Guidance. In: Constipation in Children and Young People: Diagnosis and Management of Idiopathic Childhood Constipation in Primary and Secondary Care. RCOG Press. National Collaborating Centre for Women's and Children's Health., London
7. Cowlam S, Vinayagam R, Khan U, Marsden S, Minty I, Moncur P, Bain I, Yiannakou YJ (2008) Blinded comparison of faecal loading on plain radiography versus radio-opaque marker transit studies in the assessment of constipation. Clinical radiology 63 (12):1326-1331. https://doi.org/10.1016/j.crad.2008.06.011
8. Reber J, McGauvran A, Froemming A (2018) Abdominal radiograph usage trends in the setting of constipation: a 10-year experience. Abdominal radiology (New York) 43 (9):2231-2238. https://doi.org/10.1007/s00261-018-1466-7
9. Reuchlin-Vroklage LM, Bierma-Zeinstra S, Benninga MA, Berger MY (2005) Diagnostic value of abdominal radiography in constipated children: a systematic review. Archives of pediatrics & adolescent medicine 159 (7):671-678. https://doi.org/10.1001/archpedi.159.7.671
10. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K (2015) Supporting information retrieval from electronic health records. J of Biomedical Informatics 55 (C):290-300. https://doi.org/10.1016/j.jbi.2015.05.003
11. The Big List of Narcotic Drugs. (2018) American Addiction Centers. https://americanaddictioncenters.org/the-big-list-of-narcotic-drugs/. Accessed October 23 2018
12. Medical Imaging Resource Community (MIRC). (2018) Radiological Society of North America. http://mirc.rsna.org/query. Accessed 10/23/18 2018
13. Leech SC, McHugh K, Sullivan PB (1999) Evaluation of a method of assessing faecal loading on plain abdominal radiographs in children. Pediatric radiology 29 (4):255-258. https://doi.org/10.1007/s002470050583
14. van den Bosch M, Graafmans D, Nievelstein R, Beek E (2006) Systematic assessment of constipation on plain abdominal radiographs in children. Pediatric radiology 36 (3):224-226. https://doi.org/10.1007/s00247-005-0065-2
15. Rezazadeh A, Javaherizadeh H, Chahardahcherik F, Yavarahmadi MH, Sadjadei N, Tahmasebi M (2016) Reliability of Barr, Leech, and Blethyn score in using of plain radiography in determining fecal impaction in children with and without constipation. Arquivos de gastroenterologia 53 (3):141-145. https://doi.org/10.1590/s0004-28032016000300004
16. de Lorijn F, van Rijn RR, Heijmans J, Reitsma JB, Voskuijl WP, Henneman OD, Taminiau JA, Benninga MA (2006) The Leech method for diagnosing constipation: intra- and interobserver variability and accuracy. Pediatric radiology 36 (1):43-49. https://doi.org/10.1007/s00247-005-0031-z
17. Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of pharmacokinetics and biopharmaceutics 15 (6):657-680
18. Shankar PR, Kaza RK, Al-Hawary MM, Masch WR, Curci NE, Mendiratta-Lala M, Sakala MD, Johnson TD, Davenport MS (2018) Impact of Clinical History on Maximum PI-RADS Version 2 Score: A Six-Reader 120-Case Sham History Retrospective

Evaluation. Radiology 288 (1):158-163. https://doi.org/10.1148/radiol.2018172619
19. Landers R (2015) Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS. The Winnower . https://doi.org/10.15200/winn.143518.81744
20. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychological bulletin 86 (2):420-428.
21. Cicchetti D (1994) Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology, vol 6. https://doi.org/10.1037/1040-3590.6.4.284