Pediatr Gastroenterol Hepatol Nutr. 2019 May;22(3):225-232
https://doi.org/10.5223/pghn.2019.22.3.225
pISSN 2234-8646·eISSN 2234-8840

Pediatric Gastroenterology, Hepatology & Nutrition pghn

Original Article

Large Variation in Clinical Practice amongst Pediatricians in Treating Children with Recurrent Abdominal Pain

Michael W. van Kalleveen, Elise J. Noordhuis, Carole Lasham, and Frans B. Plötz

Department of Pediatrics, Tergooi Hospital, Blaricum, the Netherlands

Received: Aug 17, 2018
Revised: Oct 1, 2018
Accepted: Oct 5, 2018

Correspondence to
Frans B. Plötz
Department of Pediatrics, Tergooi Hospital, Rijksstraatweg 1, 1261AN Blaricum, the Netherlands.
E-mail: fbplotz@tergooi.nl

Conflict of Interest
The authors have no financial conflicts of interest.

ABSTRACT

Purpose: To evaluate intra- and inter-observer variability and guideline adherence amongst pediatricians in treating children aged between 4 and 18 years referred with recurrent abdominal pain (RAP) without red flags.
Methods: The first part of the study is a retrospective single-center cohort study. The diagnostic work-ups of eight pediatricians were compared to the national guidelines. Intra- and inter-observer variability were examined by Cramer's V test. Intra-observer variability was defined as the amount of variation within a pediatrician and inter-observer variability as the amount of variation between pediatricians in the application of diagnostic work-up in children with RAP. Prospectively, the same pediatricians were requested to provide a report on their management strategy with a fictitious case to prove similarities in retrospective diagnostic work-up.
Results: A total of 10 patients per pediatrician were analyzed. Retrospectively, a (very) weak association between pediatricians' diagnostic work-ups was found (0.22), which implies high inter-observer variability. The association between intra-observer diagnostic was moderate (range, 0.35–0.46). The Cramer's V of 0.60 in diagnostic work-up between pediatricians in the fictitious case implied the presence of a moderately strong association and lower inter-observer variability than in the retrospective study. Adherence to the guideline was 66.8%.
Conclusion: We found a high intra- and inter-observer variability and moderate guideline adherence in daily clinical practice amongst pediatricians in treating children with RAP in a teaching hospital.

Keywords: Guideline adherence; Interobserver variability; Intra-observer variability; Pediatricians; Abdominal pain

INTRODUCTION

Recurrent abdominal pain (RAP) is common amongst children and adolescents and often a reason for referral to a pediatrician. In approximately 10% of cases with RAP, an identifiable somatic cause for RAP is identified. As a consequence, RAP is often classified, according to the pediatric Rome III criteria, as an abdominal pain-related-functional gastrointestinal disorder (AP-FGID). AP-FGIDs include 5 Rome III criteria conditions: functional abdominal pain, functional abdominal pain syndrome, irritable bowel syndrome, functional dyspepsia, and abdominal migraine [1,2].

Clinical decision guidelines have been developed to guide clinicians in making a decision to perform or to omit additional investigations in children with RAP. The Dutch evidence-based guidelines advise detailed examination of the history and physical examination to detect alarm symptoms, the so-called "red flags," in children with RAP. In the absence of these "red flags," explanation and reassurance are justified. According to the guidelines, extensive diagnostic tests are not recommended in view of a low pre-test probability of finding a somatic cause [3]. In addition, this policy may serve to reduce financial costs, minimize nonspecific findings, and remove fear for painful diagnostic testing [4-6]. Despite well-defined guidelines, it is unknown whether pediatricians adhere to the guidelines during daily clinical practice.

The present study was undertaken to evaluate current clinical practice amongst pediatricians in treating children aged 4 years and older, referred with RAP without alarm symptoms in a large teaching hospital. We retrospectively studied adherence to the Dutch guidelines and prospectively studied adherence in a fictitious case, and especially investigated guideline adherence and intra- and inter-observer variability.

MATERIALS AND METHODS

Study design

This single-center study was conducted at the Tergooi Hospital in Blaricum, the Netherlands between August 2016 and December 2016. Tergooi Hospital is a 496-bed teaching hospital and serves a population of approximately 250,000 habitants. The first part of the study is a retrospective single-center cohort study. The second part is a prospective survey study amongst pediatricians working at the Tergooi Hospital.

The Scientific Review Committee of Tergooi Hospital approved the study in 2016 (reference letter: kv/16.090, CTS-nr. 16.66 studie). The committee judged that the study did not fall under the Medical Research Involving Human Subjects Act so that informed consent from patients and caregivers was not required.

Part 1. Retrospective cohort study

1. Participants

The pediatricians working in Tergooi Hospital since January 2013 and with at least 10 patients with RAP between 2013 and 2015 in their care were included for intra- and inter-observer variability analysis. The number of included pediatricians was based on the maximum attainable number of patients with a minimum of 10 patients per pediatrician. Intra-observer variability was defined as the amount of variation within a pediatrician and inter-observer variability as the amount of variation between pediatricians in the application of diagnostic work-up in children with RAP.

2. Study protocol

Patients were eligible for inclusion if they were between 4 and 18 years old and attended the outpatient department of Tergooi Hospital between January 1st, 2013 and December 31st, 2015 with RAP. Included patients were referred to a pediatrician by a general practitioner. Diagnostic work-up and follow-up were performed by the same pediatrician (with the exception of medical students and pediatric trainees under supervision). RAP was expected as the major symptom and was required to be present at least during 3 episodes in three months (severe enough to affect daily activities). Patients were excluded from this study

if "red flags" in the medical history were present, which were defined as unintentional weight loss, gastrointestinal blood loss, vomiting (prolonged, bilious, or projectile), chronic diarrhoea (≥3 watery stools per day, longer than 2 weeks), arthralgia, unexplained fever and/or positive family history for inflammatory bowel disease (IBD), celiac disease or familial Mediterranean fever. Patients were also excluded if abnormalities during the physical examination were found (i.e. abnormal growth, fever, uveitis, mouth ulcers, erythema nodosum, arthritis, icterus, suspected anemia, persistent abdominal pain localized in the right upper or lower quadrant, and/or perianal abnormalities). Finally, patients under 4 years of age were excluded due to a higher pre-test probability of underlying somatic causes [4,5].

3. Data extraction

In the Netherlands, diseases or symptoms are classified using 'diagnostic treatment combinations (DBC)'. RAP is as such classified as a DBC, in contrast to a Rome III or IV diagnosis. The patient care administration department at the Tergooi Hospital provided a list of children classified with 'RAP' and their pediatricians during the study period. The medical records of included patients were reviewed in a reverse chronicle order to represent the most recent population. The following data were obtained from the medical records: demographic characteristics, diagnosis according to Rome III criteria [1], characteristics of outpatient visits, and performed diagnostic work-up by the pediatrician.

4. Guidelines for RAP

Guideline adherence was studied by comparing diagnostic work-ups with the national guidelines. The guidelines recommended investigations were: complete blood count (CBC), C-reactive protein (CRP), and celiac serology. In patients who suffered from diarrhea, additional feces for Giardia Lamblia was advised and in patients who were suspected for IBD, a fecal calprotectin was advised.

Part 2. Prospective survey

The pediatricians who were included in the first part of the study were invited to complete a questionnaire. Briefly, the questionnaire consisted of several items, namely demographic characteristics, diagnostic work-up in a fictitious case of a child with RAP without red flags, reasons and considerations to perform diagnostic tests in children with RAP, and questions and reasons about the use of guidelines (awareness, application, individual preferences, and reasons to deviate).

Statistical analysis

We used the SPSS version 22.0 (SPSS Inc, Chicago, IL, USA) for the statistical analyses. A Kruskal-Wallis test was performed to analyze differences between pediatricians on several domains (i.e. patients' characteristics and clinical work-up). The level of significance was set at $p<0.05$. Intra- and interobserver variability was studied by means of a Cramer's V-test. The result of a Cramer's V-test lies between 0 and 1 and is interpreted as following: 0, no association; 0.01–0.3, very weak to the weak association; 0.3–0.5, moderate association; and >0.5, moderately strong to a very strong association.

RESULTS

Part 1. Retrospective cohort study

1. Participants and patients

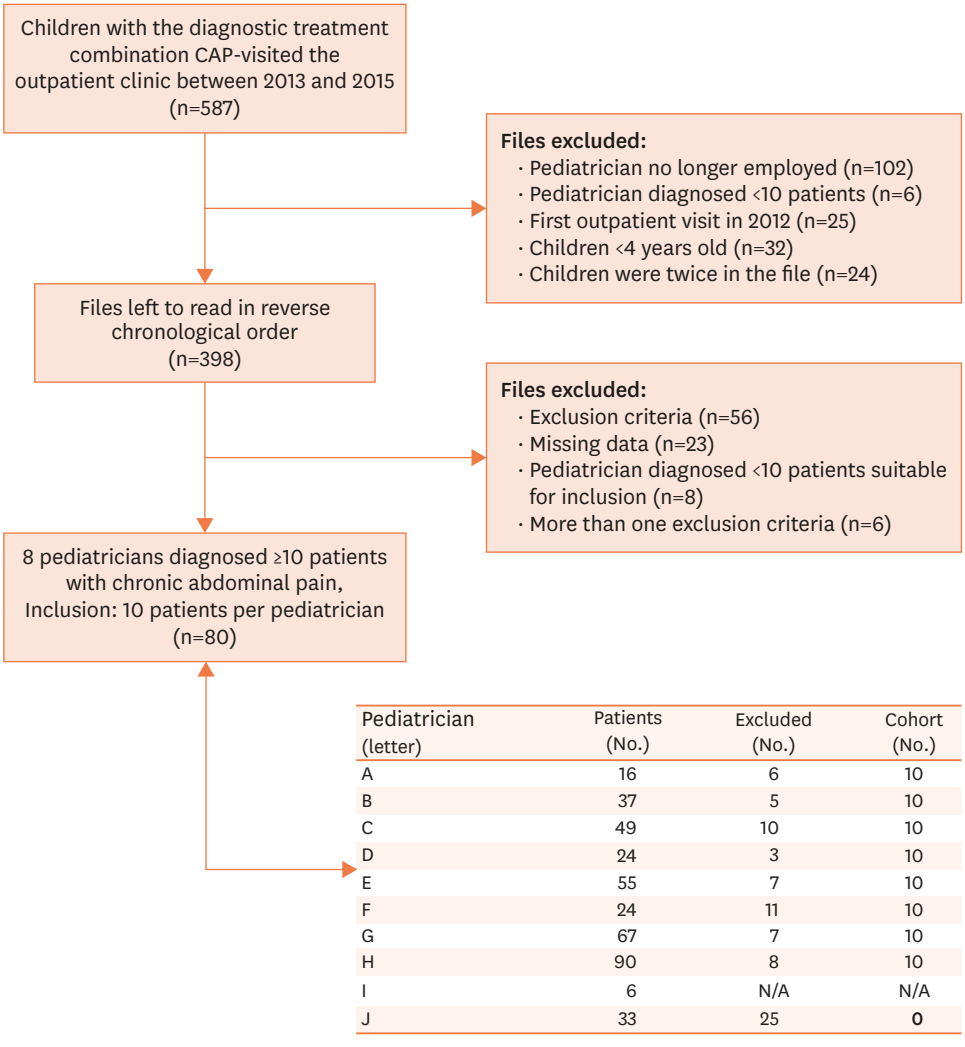
https://doi.org/10.5223/pghn.2019.22.3.225

During the study period, 587 children visited the outpatient department of Tergooi Hospital with RAP (**Fig. 1**). After the first review of 587 children, 189 children were excluded and 398 records remained for analysis in reverse chronicle order. After reviewing the 398 files, 8 of 10 pediatricians met the inclusion criteria (≥10 patients with RAP). Included pediatricians were anonymously categorized (A–H). Ten most recently diagnosed patients with RAP were categorized per pediatrician.

The clinical characteristics of patients are summarized in **Table 1**. We found that 70% (n=56) of patients with RAP were classified as AP-FGID according to Rome III criteria. In 26% (n=21) cases, the organic cause was found, and in 4% (n=3) cases, a combination of the organic and functional cause was found, respectively. There were no statistically significant differences between the included pediatricians except for the total number of outpatient visits (p=0.045).

2. Guidelines adherence

The clinical work-up per pediatrician in patients with RAP are shown in **Table 2**. Guidelines adherence was defined as: 0%, no adherence; 1%–49%, very weak to weak adherence; 50%–80%, moderate adherence; and >80%, moderately strong to very strong adherence.



Pediatrician (letter)	Patients (No.)	Excluded (No.)	Cohort (No.)
A	16	6	10
B	37	5	10
C	49	10	10
D	24	3	10
E	55	7	10
F	24	11	10
G	67	7	10
H	90	8	10
I	6	N/A	N/A
J	33	25	**0**

Fig. 1. Study flow chart.

https://doi.org/10.5223/pghn.2019.22.3.225

Table 1. Clinical characteristics of patients per pediatrician

Characteristics	Total (n=80)	A (n=10)	B (n=10)	C (n=10)	D (n=10)	E (n=10)	F (n=10)	G (n=10)	H (n=10)	*p*-value
Age (y)	9.7 (4–17)	10.2 (7–17)	7.1 (4–14)	11.2 (4–17)	8.1 (4–16)	10.1 (8–13)	10.1 (4–16)	10.3 (5–15)	10.4 (7–15)	0.173
Sex (male)	53.75	60	60	30	60	50	60	50	60	0.873
Outpatient visits	2.1 (1–7)	1.6 (1–2)	2.2 (1–4)	3.3 (2–7)	2.5 (1–4)	2 (1–4)	1.4 (1–2)	1.8 (1–3)	1.9 (1–3)	0.045
Telephone consultations	1.1 (0–4)	1.2 (0–3)	1.1 (0–3)	1.1 (0–3)	1.1 (0–4)	1 (0–2)	1 (0–2)	1.3 (0–3)	0.8 (0–3)	0.961
Duration of symptoms (mo)	9.3 (0.5–72)	10.1 (1–24)	11.4 (2–48)	31 (4.5–72)	22.2 (1.5–42)	10.6 (3–24)	9.7 (1–18)	7 (0.5–12)	11.3 (0.75–36)	0.155
Follow-up (wk)	8.8 (0–74)	8.3 (0–20)	8 (0–27)	20.6 (2–74)	10.5 (2–42)	7 (2–13)	5 (1–9)	6.1 (1–21)	5 (0–12)	0.356
Time to diagnosis (wk)	5.6 (0–74)	4.7 (0–20)	5.8 (0–27)	15.1 (0–74)	5.8 (0–39)	4.8 (2–13)	2.8 (1–5)	3.6 (0–8)	2.8 (0–11)	0.645

Data are presented as median (interquartile range) or number (%).

Table 2. Diagnostic tests performed in retrospect (n=10) and in a fictitious case (n=1)

Per pediatrician	Total retro	Total pros	AR	AP	BR	BP	CR	CP	DR	DP	ER	EP	FR	FP	GR	GP	HR	HP
Hematology panel, Celiac serology, CRP	67%	13%	70%		47%		77%		67%		53%		67%	X	80%		73%	
Hematology panel	83%	63%	80%		50%		100%		90%	X	80%	X	90%	X	80%	X	90%	X
Celiac serology	77%	63%	50%		60%		100%		90%	X	80%	X	70%	X	100%	X	60%	X
CRP	41%	13%	80%		30%		30%		20%		-		40%	X	60%		70%	
ESR	69%	13%	80%		40%		50%		90%		60%		90%		60%		80%	X
LBP panel	40%	13%	-		30%		70%		20%		70%		60%		30%		40%	X
Kidney panel	41%	13%	-		30%		80%		30%		80%		60%		30%		20%	X
Electrolytes	6%	-	-		-		20%		-		10%		20%		-		-	
Thyroid panel	25%	-	-		-		60%		50%		50%		40%		-		-	
Iron levels	11%	13%	-		-		10%		10%		50%	X	-		-		20%	
Allergy panel	5%	-	20%		-		10%		10%		-		-		-		-	
Urine	16%	-	20%		20%		20%		-		40%		-		10%		20%	
Parasites	74%	50%	90%	X	90%	X	90%		90%		70%	X	70%		10%		80%	X
Calprotectin	46%	25%	70%	X	30%		80%	X	30%		30%		50%		30%		50%	
SSYC	13%	-	30%		-		20%		20%		-		-		-		30%	
H. pylori screening	21%	13%	40%	X	30%		30%		-		30%		20%		20%		-	
Abdominal US	19%	-	-		-		70%		-		30%		30%		-		20%	
Immunoglobulins	5%	-	-		-		-		30%		-		10%		-		-	
No tests performed	10%	-	30%		30%		-		-		10%		-		-		10%	

A–H are the anonymized pediatricians. Every first column represents the retrospective part of the study, every second column represents the result from the prospective survey.
CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, LBP: liver, biliary, and pancreatic, SSYC: *Shigella*, *Salmonella*, *Yersinia*, and *Campylobacter*, *H. pylori*: *Helicobacter pylori*, Abdominal US: abdominal ultrasound.

None of the pediatricians strictly followed the guidelines proposed combination of CBC, CRP, and celiac serology (67%; range, 47%–80%), which represents a moderate adherence. The adherence to the guidelines was moderately strong for performing a CBC (83%; range, 50%–100%), moderate for celiac serology (77%; range, 50%–100%), and weak for CRP (41%; range, 0%–80%).

3. Intra- and inter-observer variability

A very weak association (Cramer's V value 0.22) between pediatricians' diagnostic work-up was found, which implies a high inter-observer variability. In terms of intra-observer variability, a moderate association (mean Cramer's V value, 0.40; range, 0.35–0.46) was found for all pediatricians.

Part 2. Prospective survey

The survey was completed by 8 pediatricians who participated in the retrospective study. Diagnostic tests performed by pediatricians in retrospective (R, n=10 patients) and in prospective fictitious cases (P, n=1 patient) are shown in **Table 2**. Only 1 pediatrician performed the proposed combination of CBC, CRP, and celiac serology. The Cramer's V of 0.60 in diagnostic work-up between pediatricians in the fictitious case implies the

presence of moderately strong association and lower inter-observer variability compared to a retrospective study. The reasons to deviate from the guidelines included feelings of being insufficiently informed about the guidelines, disagreement with the guidelines, and not being convinced with the added value.

DISCUSSION

The aim of this study was to objectify intra- and inter-observer variability and the degree of guidelines adherence in diagnostic work-up among pediatricians in treating children with RAP without red flags. We observed that guidelines adherence was moderate and inconsistent and the inter- and intra-variability in diagnostic work-up in children with RAP was large. To the best of our knowledge, until date, simultaneous reporting of intra- and inter-observer variability in diagnostic work-up has not been reported. The results of the study demonstrate that pediatricians follow their clinical experience rather than practicing evidence-based guidelines. The consequences of such an approach to patient-related outcomes were not examined.

The results of our study are in line with various other pediatric studies that evaluated diagnostic work-up in the treatment of common pediatric disorders [7-10]. For example, in a multi-center retrospective cohort study in 30 large pediatric centers in the United States, a variation of 38%–89% in performing chest X-rays in hospitalized infants <1-year-old with bronchiolitis was reported [7]. Also, the performance of diagnostic tests, for example, CBC's, blood cultures, blood chemistries, viral studies, inflammatory markers, and chest radiographs in children with community-acquired pneumonia across emergency departments in the United States showed a large variation [8]. Low adherence has also been reported in the presence of well-defined evidence-based guidelines [9,10]. Niele et al. reported that almost 50% of clinicians in the Netherlands managing children with minor traumatic brain injury often deviate from the evidence-based guidelines [10]. Urkin et al. [9] reported non-adherence in 50% of pediatricians managing children with acute pharyngitis.

Apparently, it is indicated that clinical decision making is based on the combination of evidence-based guidelines and clinical experience. The main reason for the guideline deviation in our study was disagreement with the guidelines. Other reasons for guideline deviation include lack of knowledge regarding the guidelines and unclear recommendations by the guidelines [11]. The degree of adherence to evidence-based guidelines is also partly influenced by the clinicians' number of years in practice [9,10,12]. Experienced clinicians are more likely to deviate from guidelines [10,12]. This may also be true for the pediatricians involved in our study as all the pediatricians possessed experience of at least 10 years of clinical practice. Adherence has been reported to be higher if guidelines recommendations are based on stronger evidence compared to the guidelines based on lower evidence [13]. Finally, it is hypothesized that various cultural, psychological, and local factors may also influence the behavior of clinicians, which cannot be easily incorporated into the guidelines [9].

Although non-adherence to evidence-based guidelines may have clinical consequences, the reverse is also true in case of strict adherence to guidelines. Both the approaches may result in the execution of unnecessary diagnostic tests leading to possible false-positive or false-negative results, generation of financial costs, and unnecessary anxiety. For example, strict

adherence to mild traumatic head injury increases computed tomography scan use, which may increase the risk of non-specific findings [14].

Based on our results, we hypothesize that it is extremely imperative to frequently evaluate the guidelines, in particular, the diagnostic performance and the extent of applicability in clinical practice. Such an evaluation would increase the awareness about the existence of differences in diagnostic approach within a team of pediatricians as well as in individuals, as can be seen in the present study. Furthermore, patient-related outcome measures should also be taken into account while evaluating diagnostic work-up or treatment of common pediatric disorders. Finally, it is important to explore the factors influencing the pediatricians for not adhering to evidence-based guidelines. Various studies have been conducted to identify possible factors required for improving pediatricians adherence to guidelines, monthly educational meetings, hospital-wide internet-based learning modules, creating pocket cards, and by a survey regarding barriers to adherence in order to make local modifications for the guidelines [15-17].

This study has some limitations. First, the use of a fictitious case has limitations. A paper case does not offer the opportunity to detect subtle signs during the clinical presentation and therefore does not completely represent the real situation. Second, we have lower percentages of children diagnosed with functional RAP (70%) than the reported literature (90%) [18]. This can be explained that we classified a positive stool for parasites as an organic cause although it remains very questionable whether a positive stool for parasites fully explains the symptoms of organic RAP. Therefore, the percentage of children identified with functional RAP might be an underestimate. However, our sample of patients represents a general pediatric population and the patient characteristics per pediatrician were comparable. The major strength of our study is the examination of both retrospective and prospective intra- and inter-observer variability. We were able to illustrate differences and similarities in retrospective clinical work-up and in a fictitious case in children with RAP. To the best of our knowledge, there exists little information regarding intra-observer variability among pediatricians.

In conclusion, we observed a high intra- and inter-observer variability and moderate guideline adherence amongst pediatricians in the treatment of children with RAP in daily clinical practice. We recommend evaluation of guidelines performance with respect to diagnostic efficiency and the extent of applicability in clinical practice.

REFERENCES

1. Baber KF, Anderson J, Puzanovova M, Walker LS. Rome II versus Rome III classification of functional gastrointestinal disorders in pediatric chronic abdominal pain. J Pediatr Gastroenterol Nutr 2008;47:299-302.
PUBMED | CROSSREF
2. Hyams JS, Di Lorenzo C, Saps M, Shulman RJ, Staiano A, van Tilburg M. Childhood functional gastrointestinal disorders: child/adolescent. Gastroenterology 2016;150:1456-1468.e2.
PUBMED | CROSSREF
3. Di Lorenzo C, Colletti RB, Lehmann HP, Boyle JT, Gerson WT, Hyams JS, et al. Chronic abdominal pain in children: a technical report of the American Academy of Pediatrics and the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. J Pediatr Gastroenterol Nutr 2005;40:249-61.
PUBMED | CROSSREF

4. Wright NJ, Hammond PJ, Curry JI. Chronic abdominal pain in children: help in spotting the organic diagnosis. Arch Dis Child Educ Pract Ed 2013;98:32-9.
PUBMED | CROSSREF
5. Rajindrajith S, Zeevenhooven J, Devanarayana NM, Perera BJ, Benninga MA. Functional abdominal pain disorders in children. Expert Rev Gastroenterol Hepatol 2018;12:369-90.
PUBMED | CROSSREF
6. Yacob D, Di Lorenzo C. How to deal with pediatric functional gastrointestinal disorders. Curr Pediatr Rep 2013;1:198-205.
CROSSREF
7. Christakis DA, Cowan CA, Garrison MM, Molteni R, Marcuse E, Zerr DM. Variation in inpatient diagnostic testing and management of bronchiolitis. Pediatrics 2005;115:878-84.
PUBMED | CROSSREF
8. Florin TA, French B, Zorc JJ, Alpern ER, Shah SS. Variation in emergency department diagnostic testing and disposition outcomes in pneumonia. Pediatrics 2013;132:237-44.
PUBMED | CROSSREF
9. Urkin J, Allenbogen M, Friger M, Vinker S, Reuveni H, Elahayani A. Acute pharyngitis: low adherence to guidelines highlights need for greater flexibility in managing paediatric cases. Acta Paediatr 2013;02:1075-80.
PUBMED
10. Niele N, Willemars L, van Houten M, Plötz FB. National survey on managing minor childhood traumatic head injuries in the Netherlands shows low guideline adherence and large interhospital variations. Acta Paediatr 2018;107:168-9.
PUBMED | CROSSREF
11. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implement Sci 2009;4:54.
PUBMED | CROSSREF
12. Halm EA, Atlas SJ, Borowsky LH, Benzer TI, Metlay JP, Chang YC, et al. Understanding physician adherence with a pneumonia practice guideline: effects of patient, system, and physician factors. Arch Intern Med 2000;160:98-104.
PUBMED | CROSSREF
13. Cnossen MC, Scholten AC, Lingsma HF, Synnot A, Tavender E, Gantner D, et al. Adherence to guidelines in adult patients with traumatic brain injury: a living systematic review. J Neurotrauma 2016;33:1-14.
PUBMED | CROSSREF
14. Jansen PR, Dremmen M, van den Berg A, Dekkers IA, Blanken LM, Muetzel RL, et al. Incidental findings on brain imaging in the general pediatric population. N Engl J Med 2017;377:1593-5.
PUBMED | CROSSREF
15. Paul R, Melendez E, Stack A, Capraro A, Monuteaux M, Neuman MI. Improving adherence to PALS septic shock guidelines. Pediatrics 2014;133:e1358-66.
PUBMED | CROSSREF
16. Nicastro E, Lo Vecchio A, Liguoro I, Chmielewska A, De Bruyn C, Dolinsek J, et al. The impact of e-learning on adherence to guidelines for acute gastroenteritis: a single-arm intervention study. PLoS One 2015;10:e0132213.
PUBMED | CROSSREF
17. Redaèlli M, Vollmar HC, Simic D, Maly-Schürer C, Löscher S, Koneczny N. Guideline implementation study on asthma: results of a pragmatic implementation approach. Z Evid Fortbild Qual Gesundhwes 2015;109:124-31.
PUBMED
18. Maas L, Dorigo-Zetsma JW, de Groot CJ, Bouter S, Plötz FB, van Ewijk BE. Detection of intestinal protozoa in paediatric patients with gastrointestinal symptoms by multiplex real-time PCR. Clin Microbiol Infect 2014;20:545-50.
PUBMED | CROSSREF