의학석사 학위논문

Interexaminer Reliability of Pharyngeal Injection and Palatine Tonsillar Hypertrophy in a Pediatric Emergency Department

소아응급실에서 목구멍 검진의 평가자간 일치도

2019년 8월

서울대학교 대학원

의학과 응급의학전공

황 소 연

소아응급실에서 목구멍 검진의 평가자간 일치도

지도교수 신 종 환

이 논문을 의학석사 학위논문으로 제출함

2019 년 6 월

서울대학교 대학원

의학과 응급의학전공

황 소 연

황소연의 석사 학위논문을 인준함

2019 년 8 월

위 원 장 곽 영 호 (인)

부 위 원 장 신 종 환 (인)

위 원 서 동 인 (인)

Contents

Abstract

Interexaminer Reliability of Pharyngeal Injection and Palatine Tonsillar Hypertrophy in a Pediatric Emergency Department

Soyun Hwang

Department of Emergency Medicine

The Graduate School

Seoul National University

Background

Interpretation of physical examination findings can differ based on the examiner's characteristics. In previous studies, the interrater reliability of physical examination varies and improvements in standardization of physical examination methods are necessary. In this study, we evaluated the interrater reliability of pharyngeal injection (PI) and palatine tonsillar hypertrophy (PTH) in children according to each examiner's major (emergency medicine (EM) or pediatrics) and training stage using photographs of the throats of children who visited the pediatric ED.

Methods

We performed a prospective observational study of interrater reliability. The

participants included physicians with various amounts of experience and majors who were working in an urban, tertiary hospital. We collected 20 photos of the throats of children who presented to the pediatric emergency department (ED) and performed 2 surveys (with or without medical history). The primary outcome was the interrater agreement for pharyngeal injection (PI) and palatine tonsillar hypertrophy (PTH), and the secondary outcome was the interrater agreement for PI and PTH in subgroups of examiners divided by major and duration of clinical experience.

Results

Thirty-three examiners participated in this study. The overall percent agreement for PI was 0.669, and Fleiss' kappa was 0.296. The interrater reliability was similar before and after providing patients' medical history. The overall percent agreement for PTH was 0.408, and Kendall's W was 0.674. When the patients' medical history was provided, Kendall's W increased (0.692). In the subgroup analysis, Fleiss' kappa for PI ranged from 0.257 to 0.33, and Kendall's W for PTH ranged from 0.593 to 0.711.

Conclusions

Examiners' agreement for PTH was more reliable than that for PI when evaluating children who visited the ED. The interrater reliability did not improve with increased clinical experience. These findings should be considered in the examination of pharyngeal pathology.

* This work is published in The American Journal of Emergency Medicine.

Keywords : interexamininer reliability, emergency department, pharyngeal injection, palatine tonsillar hypertrophy

Student Number : 2017-29165

I. Introduction

In the emergency department (ED), patient flow is mainly determined by physical examination findings, and the reliability and reproducibility of these findings is very important.(1, 2) However, physical examinations can be subjective. In training hospitals in particular, interpretation of physical examination findings can differ based on the examiner's level of clinical experience, which depends on the training stage, so treatment directions may vary by examiner.

Unlike in adult populations, in pediatric patients, physicians tend to minimize unnecessary radiographic imaging and laboratory tests. Radiography tends to be performed as little as possible because of the risk of radiation exposure(3, 4), and blood and urine tests are expensive, time consuming and invasive(5). In short, accurate physical examination is essential for accurate decision making to practice pediatric emergency medicine (EM).

In some previous studies, the interrater reliability of abdominal examinations in children who visited the ED with abdominal pain was evaluated, and the results showed that physical examination findings were different according to the examiner's major and training stage.(6-8) Another study evaluated the inter- and intrarater reliability of the overall clinical appearance of febrile infants and showed modest agreement; the provider's level of experience had little effect on agreement.(9) However, few studies have evaluated the interrater reliability of physical examinations in the pediatric ED, and further studies are necessary to provide quality control for physicians' physical examinations. Thus, improvements in standardization of physical examination methods are necessary.

Fever is the most common chief complaint of infants and children in the

pediatric ED. Every child who presents to the ED with fever undergoes a physical examination to identify the cause of the fever. In particular, pharyngitis and tonsillitis are common causes of fever in children, and blood and urine tests are not required in addition to physical examination in most cases(10, 11). Therefore, the proper diagnosis of pharyngitis and tonsillitis may prevent unnecessary testing. Furthermore, if pharyngitis and tonsillitis are definitely excluded, appropriate screening for the correct diagnosis is indicated.

In this study, we evaluated the interrater reliability of pharyngeal injection (PI) and palatine tonsillar hypertrophy (PTH) in children according to each examiner's major (emergency medicine (EM) or pediatrics) and training stage using photographs of the throats of children who visited the pediatric ED.

II. Methods

2.1 Study design and setting

This was a prospective observational study of interrater reliability. The study was conducted from August 2017 to October 2017. This study was performed at an urban, tertiary teaching hospital ED with residencies and fellowships in EM and pediatrics. This hospital also has a distinctive pediatric EM fellowship and faculty. This study was approved by Institutional Review Board of the Hospital (IRB No. 1706-188-864). Our IRB did not require written consent from the study participants.

2.2 Study participants and sample size

The participants included residents, fellows and faculty in EM and pediatrics. We recruited eleven EM residents, eleven board-certified general and pediatric EM physicians, and eleven pediatrics residents or board-certified pediatricians.

For this interrater reliability study, we assumed that if the relative error was 30%, and the overall agreement was 70%, 20 subjects were needed.(12) Additionally, if the desired coefficient of variation was 20%, then the required number of raters was 10 for each group(13). Considering the subgroup analysis and possible drop-outs, we recruited a total of 33 doctors including eleven physicians in each group.

2.3 Study protocol

We extracted medical photos of the throats of children who presented to the pediatric ED from the electronic medical record (EMR) system. These photos were taken by EM residents with a small endoscope camera and uploaded to the EMR. Twenty photos were selected, and the research personnel developed Google survey forms. The photos were selected to include various ages of patients (from 18-month-old to 18-year-old patients) and clinical severities (from grade 0 to grade 4). The photos were selected by 2 board-certified pediatric emergency physicians, both with more than 10 years of clinical experiences in pediatric EM. These 2 physicians did not participate in the survey. They ensured that photos of various grades of PTH, as assessed by a previously standardized system for evaluating tonsillar size, were included(14). The evaluation system of PTH used in the selection of photos were stated below in table 1.

Table 1. Grading of PTH used in the selection of photos this study.

0: Tonsils are situated in the tonsillar fossa, with no impingement on the oropharyngeal airway
+1: Tonsils sit just outside of the tonsillar fossa with obstruction of less than 25 per cent of the airway
+2: Tonsils are readily seen in the airway-25 to 50 per cent of the airway is obstructed
+3: Tonsils denote a 50 to 75 per cent obstruction of the airway
+4: Tonsils involve a greater than 75 per cent obstruction of the airway

However, since there were no clear criteria for PI, photos were selected based on the medical experience of these two physicians, and the principal investigator determined which photo to include if the two physicians did not agree on a single photo. The photos used in this study are attached to the supplement materials (Supplement Figure 1~20).

There were two sets of Google forms: set A or set B. Set A contained 20 photos of children's throats, and for each photo, the sex and age of the child was provided followed by 2 questions. The first question was regarding the presence of PI, and the answer was either yes or no. The second question was regarding presence and severity of PTH, and the answer ranged from grade 0 to grade 4. Set B was similar to set A, but additional simple clinical informations were also provided, such as the duration of fever and accompanied symptoms such as the presence of the sore throat. Other physical examination findings or laboratory test results were not provided. The order of the photos was not different between the two sets. Other basic characteristics of the examiners were also collected including their majors (EM, pediatrics, or both) and the number of years of clinical experience.

(set A: https://goo.gl/forms/8PkhVtVU1c9TSVHH3;
set B: https://goo.gl/forms/a2h0HDFcYDbjhGBP2)

When eligible participants were identified, the research personnel recruited the participant, and the Google survey form was sent via e-mail. Set A was sent first, and then when the participant answered the survey, set B was sent three days after their response. The participants were encouraged to answer the survey after their duty and not to discuss the answers with other people.

2.4 Measures

Our primary outcome was the interrater agreement of the examiners

regarding PI and PTH. Our secondary outcome was the interrater agreement regarding PI and PTH in subgroups of examiners that were divided according to their majors and duration of clinical experience (residents vs board-certified physicians).

2.5 Data analysis and Statistics

Data were entered into an Excel spreadsheet (Microsoft, 2016), and analysis was performed with STATA (version 14.0, STATA corp., College Station, TX, USA). Proportions were calculated for categorical variables. For interrater agreement regarding PI, Fleiss' kappa coefficient was mainly used because this was a nominal variable in the reproducibility test, and three or more assessors were compared.(15, 16) Percent agreement can be calculated but not recommended because it was not corrected for chance agreement. However, Gwet's first order agreement coefficient (Gwet's AC1) can be used and interpreted as needed.(15, 16)

For interrater agreement regarding PTH, Kendall's coefficient of concordance (Kendall's W) was used because this was an ordinal variable with rankings (grade 0 to 4) in the reproducibility test among three raters or more(17, 18). Percent agreement can be calculated also, but not recommended because of the same reason that stated above. Other chance-corrected agreement coefficients like Pearson’s correlation coefficient was not appropriate, because it measures the extent to which the relationship between two series of score is linear rather than the agreement itself.

Both Fleiss' kappa coefficient and Kendall's W range from -1 to +1, and +1 indicates perfect agreement. Descriptive terms, such as 'poor agreement' and 'moderate agreement', were also used, according to previously published studies.(19, 20) The interpretation of coefficients of interrater reliability used

in this study is shown in Table 2.

Table 2. Interpretation of coefficients of interrater reliability

coefficient	interpretation
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.91-1.00	Almost perfect

III. Results

3.1 Participant characteristics

There were total 33 examiners who participated in this study. The distribution of their majors and duration of clinical experience in years is shown in table 3. There were 20 EM physicians, 11 pediatricians and 2 double-boarded physicians (both in EM and pediatrics; these physicians were considered EM physicians in our analysis because they were currently working in the ED as emergency physicians). The majority of physicians had more than 5 years but less than 10 years of clinical experience. Ten doctors had more than 3 years but less than 5 years of clinical experience, and four doctors had more than 1 year but less than 3 years of clinical experience. Two doctors had more than 10 years of experience. The distribution of clinical experience in years in each subgroup is shown in table 4. No participants in this study were colorblind.

Table 3. Baseline characteristics of study cohort

	N	%
total	33	
Major of the participants		
PED	11	33.33
EM	20	60.61
EM + PED	2	6.06
Clinical experiences		
1~3 years	4	12.12
3~5 years	10	30.3
5~10 years	17	51.52
>10 years	2	6.06

Table 4. Distribution of clinical experience of each subgroup

	EM residents	EM boards	PED
1~3 years	3	0	1
3~5 years	6	0	4
5~10 years	2	9	6
>10 years	0	2	0

3.2 Interrater reliabilities for PI and PTH

The interrater reliabilities for PI and PTH is shown in tables 5, respectively. The overall percent agreement for PI was 0.669 (substantial agreement), and Fleiss' kappa was 0.296, indicating fair agreement. When Gwet's AC1 was calculated, the overall agreement for PI was slightly increased (0.377) but still indicated fair agreement. In cases of PI, the interrater reliability was similar before and after the patients' medical history was provided.

The overall percent agreement for PTH was 0.408 (moderate agreement), which was lower than that for PI. Kendall's W was 0.674 (substantial agreement). When the patients' medical history was provided, Kendall's W increased (0.692).

Table 5. The inter-rater reliability of PI and PTH

	PI		
	Percent Agreement	Fleiss' Kappa	Gwet's AC1
overall	0.669 (0.622-0.717)	0.296 (0.210-0.381)	0.377 (0.252-0.502)
without History	0.666 (0.603-0.729)	0.298 (0.189-0.407)	0.362 (0.191-0.533)
with History	0.673 (0.596-0.750)	0.292 (0.148-0.436)	0.392 (0.194-0.590)
	PTH		
overall	Percent Agreement	Kendall's W	
without History	0.408 (0.384-0.432)	0.674 (N/A)	
with History	0.427 (0.392-0.463)	0.692 (N/A)	

3.3 Subgroup analysis of interrater reliability

We analyzed the interrater reliability in each subgroup (tables 6). For PI, Fleiss' kappa for EM residents, board-certified EM physicians, and pediatricians was 0.289, 0.332 and 0.264, respectively, when the patients' medical history was not provided. These results demonstrated fair agreement, despite small differences among the three subgroups. After the patients' medical history was provided, Fleiss' kappa for EM residents and pediatricians decreased slightly (0.272 and 0.257, respectively) but improved a little in board-certified EM physicians; however, all the groups still had fair agreement.

For PTH, Kendall's W of each subgroup showed substantial agreement (0.690, 0.699 and 0.711) when the patients' medical history was not provided. When the patients' medical history was provided, the agreement improved among EM residents and board-certified EM physicians but decreased among pediatricians from substantial agreement (0.711) to moderate agreement (0.593).

Table 6. Subgroup analysis of inter-rater reliability of PI and PTH

		N	without history	with history
PI	Total	33	0.298 (0.189-0.407)	0.292 (0.148-0.436)
	EM resident	11	0.289 (0.166-0.413)	0.272 (0.112-0.432)
	EM board	11	0.332 (0.174-0.490)	0.370 (0.178-0.563)
	PED	11	0.264 (0.116-0.413)	0.257 (0.116-0.398)
PTH	Total	33	0.692 (N/A)	0.680 (N/A)
	EM resident	11	0.690 (N/A)	0.752 (N/A)
	EM board	11	0.699 (N/A)	0.754 (N/A)
	PED	11	0.711 (N/A)	0.593 (N/A)

IV. Discussion

In this study, we evaluated the interrater reliabilities for PI and PTH in children among EM physicians and pediatricians. The agreement for PI was generally fair, and the agreement for PTH was substantial in our analysis. The agreement for PTH increased slightly when the patients' medical history was provided, and it remained substantial. In the subgroup analyses, the results were generally similar: the agreement for PI was fair, but the agreement for PTH was substantial, with the exception of moderate agreement among pediatricians when the patients' medical history was provided.

In previous studies assessing interrater reliability, the agreement for physical examination was fair to moderate, and slight agreement was reported for pediatric abdominal examination(6, 8, 21). No studies have previously evaluated the interrater reliability for throat examinations in children, but the agreement seems to be similar to that of abdominal examination. In previous studies, some authors mentioned that the low agreement regarding physical examination might be the result of differences in the examiners' training stages and majors(8). However, other studies have showed no improvement in interrater reliability with increasing experience with clinical assessments(9, 22). Those previous studies evaluated gestalt impressions of overall clinical appearance(9) and gut feelings about serious infections(22). There are no clearly predefined guidelines or gold standards for 'overall clinical appearance' or 'gut feelings' in contrast to abdominal examination, which has relatively more established criteria. Our study showed no differences in agreement between examiners regarding throat examination according to their training stages and majors, which is similar to the findings of the abovementioned studies. This result may be because clinicians may not have clear definitions of PI and PTH in their minds; thus, more education about throat

examinations in children is necessary.

The doctors who participated in this study seemed to agree less on PI than on PTH when measured by Fleiss' Kappa and Kendall's W, respectively. However, when calculated with percent agreement, it seemed to be the opposite results. This is because Fleiss' Kappa and Kendall's W are chance-corrected agreement coefficients while percent agreement is not. While there were only two choices in case of PI (yes or no), there were five grades for PTH; thus, the agreement calculated with percent agreement can be exaggerated in measuring of PI while the agreement calculated by chance-corrected agreement coefficients is more reliable.

The possible reason for this result is that there is a well-known and widely used classification system for PTH(14), while there is no established definition for PI. Thus, evaluation of PI is more subjective to each individual's perspectives than evaluation of PTH. Even among people who are not colorblind, perceptions of the degree of 'redness' could be different. It is known that when discrimination is clear, better agreement can be obtained(20).

There are some limitations in this study. First, this study was a single-centered study and included only EM physicians and pediatricians in a training hospital. However, we tried to enroll variety of participants with different majors and durations of clinical experience. Thus, these results may be applicable to other hospitals or medical providers.

Second, this study did not measure a clinical endpoint, such as antibiotic prescriptions. Most cases of acute pharyngitis in children do not requires antibiotics because they are caused by viral organisms(10, 23), but in cases of group A streptococcal pharyngitis, antibiotics are indicated(10); thus, it is important to differentiate streptococcal pharyngitis from benign, self-limiting viral pharyngitis by physical examination. However, differentiation of

streptococcal pharyngitis from viral pharyngitis includes assessment of other physical examination findings, such as cervical lymphadenopathy(11, 24), and our survey did not include this information. Therefore, it would have been difficult and meaningless to have the examiners determine whether to prescribe antibiotics solely based on a single photograph of throat. In addition, our study primarily emphasizes the interrater agreement, rather than the intrarater agreement. Because the internal threshold for prescribing antibiotics may vary from physician to physician depending on the experiences and available resources, measuring the clinical endpoint would be less meaningful. However, the agreement did not improve with the presence of medical history in our study. This may indicate that our study participants' interpretation of physical examination findings were not affected by medical history.

Additionally, unlike appendicitis, throat examination findings cannot be easily confirmed by imaging, laboratory test or histological means. Although computed tomography (CT) scan of neck or rapid antigen detection test (RADT) can be performed in the emergency department where appropriate(25, 26), not every children with fever and sorethroat undergo these studies and whether the result of these tests correlates with physical examination finding was not established. The fact that our study does not have a gold standard may be one of the most important limitation. However, this study did not focus on the correlation of throat examination and specific throat pathology (such as peritonsillar abscess), but on the agreement of the physical examination finding itself. Further study for specific disease and measuring of clinical outcome through cooperation with pediatric outpatient department may be helpful.

Although this study has some limitations, our results indicated that physical exam itself can be subjective depending on the examiner. This problem may be caused by individual differences in both perception of the physical exam

finding and description of it. Furthermore, this low reliability of physical examination cannot be limited to the visual examination alone. Auscultation finding were also inconsistent among physicians(27), the palpable ratings may not be reliable(28), and interrater reliability of olfactory and taste sense was only moderate to good(29).

However, further efforts are needed to overcome this discrepancy of the physical examination. Proper training for each physical examination is important, but the communication of clinical findings also can be improved. One possible way to improve communication is the use of more specific expressions in describing physical examination findings. For example, in one study regarding a new grading scale for gross hematuria, the authors introduced a more specific grading scale using CYMK color codes(30). This study showed excellent agreement among the urologists as well as the laypeople because of an objective and easy-to-use grading tool.

Another breakthrough may come from the evolution of technology. Due to advances in examination room equipment and medical recording systems, communication can be augmented via audio and video media, without relying solely on writing. Although there is no international standard for video and audio media, it will be helpful for more accurate delivery and evaluation of the patient's clinical findings if the medical record technology using supplementary media becomes more generalized in the future.

In conclusion, among children visiting the ED, the interrater reliability for PI was fair, and that for PTH was good. The interrater reliability did not improve with increased clinical experience. These findings should be considered in the examination of pharyngeal pathology.

References

1. McCarthy PL, Lembo RM, Fink HD, Baron MA, Cicchetti DV. Observation, history, and physical examination in diagnosis of serious illnesses in febrile children less than or equal to 24 months. J Pediatr. 1987;110(1):26-30.

2. Reynolds SL, Jaffe DM. Diagnosing abdominal pain in a pediatric emergency department. Pediatr Emerg Care. 1992;8(3):126-8.

3. Kwon H, Jung JY. Effectiveness of a radiation reduction campaign targeting children with gastrointestinal symptoms in a pediatric emergency department. Medicine (Baltimore). 2017;96(3):e5907.

4. Jennings RM, Burtner JJ, Pellicer JF, Nair DK, Bradford MC, Shaffer M, et al. Reducing Head CT Use for Children With Head Injuries in a Community Emergency Department. Pediatrics. 2017;139(4).

5. Ouellet-Pelletier J, Guimont C, Gauthier M, Gravel J. Adverse Events Following Diagnostic Urethral Catheterization in the Pediatric Emergency Department. CJEM. 2016;18(6):437-42.

6. Kharbanda AB, Fishman SJ, Bachur RG. Comparison of pediatric emergency physicians' and surgeons' evaluation and diagnosis of appendicitis. Acad Emerg Med. 2008;15(2):119-25.

7. Hunter BR, Seupaul RA. Interrater reliability of history and physical examination is limited among children with possible appendicitis. J Pediatr. 2012;161(3):566.

8. Kharbanda AB, Stevenson MD, Macias CG, Sinclair K, Dudley NC, Bennett J, et al. Interrater reliability of clinical findings in children with

possible appendicitis. Pediatrics. 2012;129(4):695-700.

9. Walsh P, Thornton J, Asato J, Walker N, McCoy G, Baal J, et al. Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. PeerJ. 2014;2:e651.

10. Bisno AL. Acute pharyngitis. N Engl J Med. 2001;344(3):205-11.

11. Bisno AL, Gerber MA, Gwaltney JM, Jr., Kaplan EL, Schwartz RH, Infectious Diseases Society of A. Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. Infectious Diseases Society of America. Clin Infect Dis. 2002;35(2):113-25.

12. Cantor AB. Sample-size calculations for Cohen's kappa. Psychological Methods. 1996;1(2):150.

13. Gwet KL. HANDBOOK OF INTER-RATER RELIABILITY. 3rd ed. Maryland, USA: Advanced Analytics, LLC; 2012.

14. Brodsky L. Modern assessment of tonsils and adenoids. Pediatr Clin North Am. 1989;36(6):1551-69.

15. Cicchetti DV, Volkmar F, Sparrow SS, Cohen D, Fermanian J, Rourke BP. Assessing the reliability of clinical scales when the data have both nominal and ordinal features: proposed guidelines for neuropsychological assessments. J Clin Exp Neuropsychol. 1992;14(5):673-86.

16. Rucker G, Schimek-Jasch T, Nestle U. Measuring inter-observer agreement in contour delineation of medical imaging in a dummy run using Fleiss' kappa. Methods Inf Med. 2012;51(6):489-94.

17. Bottcher HF, Posthoff C. Mathematical treatment of rank correlation--a comparison of Spearman's and Kendall's coefficients. Z Psychol Z Angew Psychol. 1975;183(2):201-17.

18. Baumgartner R, Somorjai R, Summers R, Richter W. Assessment of cluster homogeneity in fMRI data using Kendall's coefficient of concordance. Magn Reson Imaging. 1999;17(10):1525-32.

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.

20. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.

21. Yen K, Karpas A, Pinkerton HJ, Gorelick MH. Interexaminer reliability in physical examination of pediatric patients with abdominal pain. Arch Pediatr Adolesc Med. 2005;159(4):373-6.

22. Van den Bruel A, Thompson M, Buntinx F, Mant D. Clinicians' gut feeling about serious infections in children: observational study. BMJ. 2012;345:e6144.

23. Putto A. Febrile exudative tonsillitis: viral or streptococcal? Pediatrics. 1987;80(1):6-12.

24. McIsaac WJ, Kellner JD, Aufricht P, Vanjaka A, Low DE. Empirical validation of guidelines for the management of pharyngitis in children and adults. JAMA. 2004;291(13):1587-95.

25. Tanz RR, Gerber MA, Kabat WK et al. Performance of a Rapid Antigen-Detection Test and Throat Culture in Community Pediatric Offices: Implications for Management of Pharyngitis. Pediatrics. 2009;123(2):437-444.

26. Baker KA, Stuart, JD, Sykes, KJ, et al. Use of Computed Tomography in the Emergency Department for the Diagnosis of Pediatric Peritonsillar Abscess. Pediatric Emergency Care. 2012;28(10):962-965.

27. Florin TA, Ambroggio L, Brokamp C, Rattan MS, Crotty EJ, Kachelmeyer A, et al. Reliability of Examination Findings in Suspected

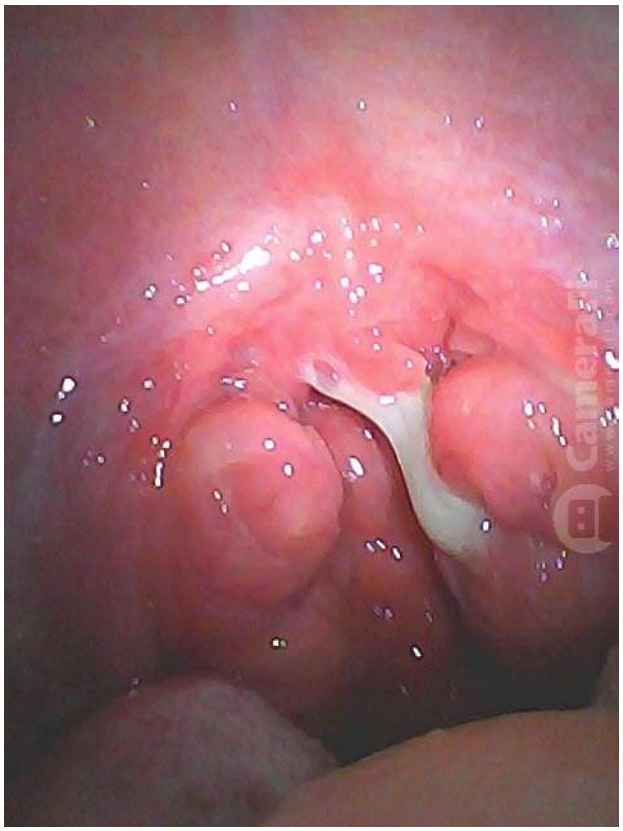
Community-Acquired Pneumonia. 2017;140(3):e20170310.

28. Gordon JK, Girish G, Berrocal VJ, Zhang M, Hatzis C, Assassi S, et al. Reliability and Validity of the Tender and Swollen Joint Counts and the Modified Rodnan Skin Score in Early Diffuse Cutaneous Systemic Sclerosis: Analysis from the Prospective Registry of Early Systemic Sclerosis Cohort. 2017:jrheum.160654.

29. Rawal S, Hoffman HJ, Honda M, Huedo-Medina TB, Duffy VBJCP. The Taste and Smell Protocol in the 2011-2014 US National Health and Nutrition Examination Survey (NHANES): Test-Retest Reliability and Validity Testing. 2015;8(3):138-48.

30. Lee JY, Chang JS, Koo KC, Lee SW, Choi YD, Cho KS. Hematuria Grading Scale: A New Tool for Gross Hematuria. Urology. 2013;82(2):284-9.

Supplement Materials



Supplement Figure 1. Photo of children's throat used in the study (1)

Supplement Figure 2. Photo of children's throat used in the study (2)

Supplement Figure 3. Photo of children's throat used in the study (3)

Supplement Figure 4. Photo of children's throat used in the study (4)

Supplement Figure 5. Photo of children's throat used in the study (5)

Supplement Figure 6. Photo of children's throat used in the study (6)

Supplement Figure 7. Photo of children's throat used in the study (7)

Supplement Figure 8. Photo of children's throat used in the study (8)

Supplement Figure 9. Photo of children's throat used in the study (9)

Supplement Figure 10. Photo of children's throat used in the study (10)

Supplement Figure 11. Photo of children's throat used in the study (11)

Supplement Figure 12. Photo of children's throat used in the study (12)

Supplement Figure 13. Photo of children's throat used in the study (13)

Supplement Figure 14. Photo of children's throat used in the study (14)

Supplement Figure 15. Photo of children's throat used in the study (15)

Supplement Figure 16. Photo of children's throat used in the study (16)

Supplement Figure 17. Photo of children's throat used in the study (17)

Supplement Figure 18. Photo of children's throat used in the study (18)

Supplement Figure 19. Photo of children's throat used in the study (19)

Supplement Figure 20. Photo of children's throat used in the study (20)

요약(국문초록)

서론 : 신체 검사 소견의 해석은 평가자의 특성에 따라 다를 수 있으며, 이전의 연구에서는 복부 검진에서 평가자의 전공과 경력에 따라 평가자 간의 해석이 다르게 나타났다. 본 연구에서는 소아응급실을 방문한 어린이의 목구멍 사진을 이용하여 평가자의 임상 경험과 전공에 따른 어린이의 인두 발적 및 구개편도비대의 평가자 간 일치도를 평가 하였다.

방법 : 본 연구는 평가자 간 일치도에 대한 전향적 관측 연구였으며, 연구 대상자들은 3차 병원에서 일하는 소아과 및 응급의학과 의사로, 전공의와 전공의를 두루 포함하였다. 소아 응급실에 방문한 어린이들의 목구멍 사진 20 장을 추출하여 웹 기반 설문지를 작성 후 연구 대상자들에게 답변하게 하여 자료를 수집하였다. 1차 연구 결과는 인두 빌적과 구개편도비대에 대한 평가자 간 일치도였고, 2차 연구 결과는 임상 경험 기간 및 전공 별 소그룹로 나누어 평가자 간 일치도를 측정하였다.

결과 : 모집 결과 소아과 의사 11명과 응급의학과 의사 22명, 총 33명의 평가자가 연구에 참여하였다. 인두 발적의 경우 전반적으로 퍼센트 동의 0.669, Fleiss' kappa 0.296의 일치도를 보였으며, 환아의 병력을 제공하기 전화 후의 일치도 변화는 없었다. 구개편도비대의 경우 전반적으로 퍼센트 동의 0.408, Kendall's W 0.679의 일치도를 보였으며 환아의 병력이 제공된 경우 Kendall's W가 소폭 증가 (0.692) 하였다. 임상 경험 기간 및 전공 별 소그룹 분석에서 인두 발적에 대한 Fleiss' kappa는 0.257에서 0.33의 범위를 보였으며 구개편도비대에 대한 Kendall's W는 0.593에서 0.711의 범위를 보였다.

결론 : 소아응급실을 방문한 어린이들의 구개편도비대에 대한 평가자간 일치도는 인두 발적의 평가자간 일치도보다 더 높은 경향이 있었다. 평가자간 일치도는 임상 경험이 증가함에 따라 향상되지 않았다. 이러한 결과를 소아의 목구멍 검진 시 고려할 필요가 있다.

* 본 내용은 The American Journal of Emergency Medicine에 출판 완료된 내용임

주요어 : 신체검진, 평가자간 일치도, 응급실, 소아

학번 : 2017-29165