

Determining the clinical significance of errors in pediatric radiograph interpretation between emergency physicians and radiologists

Jonathan Taves, MD*; Steve Skitch, PhD*; Rahim Valani, M Med ED, MBA*

ABSTRACT

Objectives: Emergency physicians (EPs) interpret plain radiographs for management and disposition of patients. Radiologists subsequently conduct their own interpretations, which may differ. The purposes of this study were to review the rate and nature of discrepancies between radiographs interpreted by EPs and those of radiologists in the pediatric emergency department, and to determine their clinical significance.

Methods: We conducted a retrospective review of discrepant radiology reports from a single-site pediatric emergency department from October 2012 to December 2014. All radiographs were interpreted first by the staff EP, then by a radiologist. The report was identified as a “discrepancy” if these reports differed. Radiographs were categorized by body part and discrepancies classified as false positive, false negative, or not a discrepancy. Clinically significant errors that required a change in management were tracked.

Results: There were 25,304 plain radiographs completed during the study period, of which 252 (1.00%) were identified as discrepant. The most common were chest radiographs (41.7%) due to missed pneumonia, followed by upper and lower extremities (26.2% and 17.5%, respectively) due to missed fractures. Of the 252 discrepancies, 207 (82.1%) were false negatives and 45 (17.9%) were false positives. In total, 105 (0.41% of all radiographs) were clinically significant.

Conclusion: There is a low rate of discrepancy in the interpretation of pediatric emergency radiographs between emergency department physicians and radiologists. The majority of errors occur with radiographs of the chest and upper extremities. The low rate of clinically significant discrepancy allows safe management based on EP interpretation.

RÉSUMÉ

Objectifs: Les urgentologues interprètent des radiogrammes simples afin de traiter les patients et de déterminer les suites

à donner. Les radiologistes font, par la suite, leur propre interprétation, qui peut être différente de celle des urgentologues. L'étude avait donc pour buts d'examiner le taux de divergence et la nature des différences entre l'interprétation des radiogrammes par les urgentologues et celle des radiogrammes par les radiologistes au service des urgences (SU) pédiatriques, et de déterminer leur portée clinique.

Méthode: Il s'agit d'un examen rétrospectif de rapports divergents en radiologie provenant d'un seul SU pédiatriques, couvrant la période d'octobre 2012 à décembre 2014. Tous les radiogrammes ont d'abord été interprétés par un urgentologue, puis par un radiologiste. Les rapports portaient la mention « Divergence » si le contenu différait. Les radiogrammes ont été catégorisés selon les parties du corps, et les divergences, classées en faux positif, en faux négatif ou en aucune divergence. Les erreurs d'interprétation cliniquement importantes qui ont nécessité une modification de la prise en charge ont fait l'objet de suivi.

Résultats: Au total, 25 304 radiographies simples ont été réalisées durant la période à l'étude et, sur ce nombre, 252 (1,00 %) ont donné lieu à des résultats divergents. La plupart concernaient des radiogrammes de la poitrine (41,7 %) dans lesquels une pneumonie était passée inaperçue, et des radiogrammes des membres supérieurs et inférieurs (26,2 % et 17,5 % respectivement) dans lesquels des fractures étaient passées inaperçues. Sur les 252 cas de divergence, 207 (82,1 %) consistaient en de faux négatifs, et 45 (17,9 %), en de faux positifs. En tout, 105 (0,41 %) interprétations erronées de radiogrammes étaient cliniquement importantes.

Conclusion: Le taux de divergence d'interprétation des radiogrammes au SU pédiatriques entre urgentologues et radiologistes est faible. La plupart des erreurs concernaient des radiogrammes de la poitrine et des membres supérieurs. Compte tenu du faible taux de divergence cliniquement importante, il est permis de croire en une prise en charge sûre des cas, reposant sur l'interprétation des radiogrammes par les urgentologues.

Keywords: pediatric, X-ray, discrepancy, quality improvement

From the Department of Emergency Medicine, Hamilton General Hospital, McMaster University, Hamilton, ON.

Correspondence to: Rahim Valani, Department of Emergency Medicine, Hamilton General Hospital, McMaster Clinic 2nd Floor, 237 Barton Street East, Hamilton, ON L8L 2X2; Email: valanir@mcmaster.ca

© Canadian Association of Emergency Physicians

CJEM 2018;20(3):420-424

DOI 10.1017/cem.2017.34

INTRODUCTION

Radiographs are a key diagnostic tool used by the emergency physician (EP). During their emergency department (ED) shift, EPs interpret radiographs to aid or confirm a diagnosis that affects both treatment plan and disposition of the patient. In most EDs, radiographs are subsequently interpreted by a radiologist, who issues an official report. Most EDs use a quality assurance (QA) system that ensures that discrepancies between the initial ED and final radiologist interpretations are addressed and changes in management made as required. Despite this retrospective system, the accuracy of the initial, real-time radiograph interpretation is important, given the risk of errors in the ED that could have consequences for morbidity and mortality.

Studies in pediatric EDs have examined the accuracy of EPs' interpretation of radiographs and the impact of discrepancies on patients. Reported discrepancy rates between EP and radiologist readings have ranged between 1% and 28%.¹⁻¹³ However, clinically significant discrepancies that lead to a change in patient management are relatively uncommon; reported rates ranged from 0% to 9%, according to the studies cited previously. Studies in adult EDs have generally reported lower imaging discrepancy rates than those found in pediatric EDs, ranging from 0.8% to 3.7%.^{5,12} Similarly, rates of clinically significant discrepancies in adult studies are also lower, ranging from 0% to 2.8%.^{1,11}

The purpose of this study was to examine discrepancies interpreted between EPs and radiologists and determine whether they were clinically significant. Changes in anatomy and ossification centres make pediatric radiographs more challenging to interpret. Earlier studies have either combined adult and pediatric patients, thereby limiting their study to imaging of fewer body parts, or were conducted when there were few trained pediatric EPs.

METHODS

The Hamilton Integrated Research Ethics Board granted approval for this study (REB #14-667-C). It consisted of a retrospective 18-month review of discrepant radiology reports from McMaster Children's Hospital, a tertiary-care academic pediatric ED, from October 2012 to December 2014. The department has a census of more than 40,000 patient visits a year. The current practice is for all radiographs to be interpreted

first by the staff EP and then reviewed by a staff radiologist within 24 hours. If this final impression differs from the EP's interpretation, the report is flagged as a "discrepancy" according to ED protocol. These films are reviewed as part of a QA process where the EP on duty is given a list of discrepant reports, reconciles the findings with the original patient chart, and provides the family with a new treatment plan, if indicated. For the purposes of comparison, we took the staff radiologist's report as the gold standard.

This chart review followed the Gilbert criteria.¹⁴ Inclusion criteria consisted of all plain X-rays completed during this period and classified as discrepant. A process was created to save and index all such imaging results for further review. Other imaging modalities (computed tomography, magnetic resonance imaging, ultrasound) and trauma imaging were excluded. If there was a question in the interpretation of the radiograph in real time and the radiologist's interpretation provided a preliminary report, then the subject was also excluded. Radiographs were categorized as chest, abdomen, axial skeleton, upper extremity, lower extremity, soft-tissue neck, and other. On the basis of the radiologist's final interpretation, discrepancies were classified as false positive (FP, abnormality noted by the EP but deemed normal by the radiologist), false negative (FN, abnormality missed by the EP), or not a discrepancy (abnormality noted by the EP elsewhere in the patient chart).

All charts were reviewed and data abstracted by one investigator (JT). If the correct abnormality was identified in a false-negative report, the case was marked as a correct diagnosis. Clinically significant errors that required a change in management of the patient were identified on the basis of medical records. Clinically significant was defined as a discrepancy requiring a change in patient management, including a new prescription, a return to the ED, or follow-up in a specialized clinic.

A total of 77 discrepancies and associated charts (26%) were randomly selected and reviewed by a second investigator (SS) to determine inter-rater reliability. Consistency between reviewers was assessed by the Kappa statistic. Data were collected using a standardized abstraction form and documented in Microsoft Excel.

RESULTS

A total of 25,304 plain radiographs were completed during this period. They included 7,939 chest (CXR),

Table 1. Patient demographics of discrepant images based on body image

Body part	Male	Female	Average age (yr)	Maximum age (yr)	Minimum age (yr)	SD
Abdomen	6	5	7.89	16.25	0.01	5.58
Axial skeleton	11	9	12.74	17.64	2.23	4.85
Chest	58	47	4.04	17.39	0.06	4.28
Upper extremity	29	37	8.78	15.89	0.59	4.45
Lower extremity	21	23	10.23	16.28	1.02	4.26
Other	2	1	15.24	16.28	13.68	1.37
Soft tissue	2	1	4.87	7.80	1.21	3.36

SD = standard deviation; yr = year.

2,914 abdomen (AXR), 6,407 upper extremity (UE), 4,396 lower extremity (LE), 2,336 axial skeleton (AS), 407 soft-tissue neck (STN), and 915 other. Of the 293 discrepancies recorded, 40 proved to be not discrepant. In these cases, the charts were interpreted but not documented on the radiology view box; instead, they were interpreted and managed appropriately based on the chart notes. They were therefore considered non-discrepant. There were 252 (1.00%) true discrepancies, 123 female and 129 male. The average age was 7.4 years (range 3 days – 17.6 years, SD 5.4 years). Table 1 summarizes patient demographics based on body image. Of the discrepancies, there were 207 false-negative (82.1%) and 45 false-positive interpretations. Table 2 shows a breakdown of the type of X-ray and the category of discrepancy.

The clinically significant error rate for all radiographs completed during the study period was 0.46% (116/25,304). Clinically significant changes included returning patients to the ED (51 patients), filling prescriptions (25), and calling patient or family members to arrange follow-up appointments (38). One patient's follow-up was not documented. Of the follow-ups done, none resulted in permanent morbidity or any mortality. Table 3 shows the overall discrepancy rate by type of radiograph and clinical significance. CXR was the most frequent study ordered, comprising 31.3% of the total ordered. It also had the highest error rate (0.17%), followed by upper extremity X-rays. A discrepancy occurring in an X-ray of the lower limbs had the highest rate of clinical significance.

Seventy-seven charts were randomly selected and reviewed for accuracy and consistency. The calculated inter-rater reliability in the charts selected for review was found to be Kappa 0.89 ($p < 0.01$, 95% CI 0.79–0.99). For the abstraction of clinical significance, inter-rater reliability was found to be Kappa 0.76 ($p < 0.01$,

Table 2. Type of discrepancy by body imaging type

	FN	FP	Total
	N	N	
Abdomen	10	1	11
Axial skeleton	19	1	20
Chest	81	24	105
Upper extremity	54	12	66
Lower extremity	37	7	44
Other	3	0	3
Soft-tissue neck	3	0	3

FN = false negative; FP = false positive.

95% CI 0.63–0.89). As described by Altman's qualitative classification system, these Kappa values represent good to very good inter-rater agreement.¹⁵

DISCUSSION

This retrospective study shows that the rates of discrepancy between EPs' and radiologists' interpretations of radiographs in a pediatric ED are quite low. The number and proportion of clinically significant discrepancies are even lower, affecting patient management in only 0.46% of all radiographs ordered. Prior studies, summarized in Table 4, show that EPs do well in interpreting adult X-ray images. However, there has been considerable variability in the interpretation of pediatric X-rays, due perhaps to anatomic changes that vary with age as well as a lack of expertise in interpreting these images. As pediatric emergency medicine has evolved into a recognized sub-speciality, the ability of EPs to interpret X-ray images has improved. Higginson (2004) asks whether, given the low error rate, we still need radiologists to interpret X-rays of pediatric patients in the ED.²

Table 3. Clinically significant errors stratified by body imaging type

	Total discrepancies	Overall discrepancy rate (%)	Clinically significant	Overall clinical significance (%)
Abdomen	11	4.3%	5	0.02
Axial skeleton	20	7.94	10	0.04
Chest	105	41.67	43	0.17
Upper extremity	66	26.19	30	0.1
Lower extremity	44	17.46	26	0.10
Other	3	1.19	1	0.00
Soft-tissue	3	1.19	1	0.00

The overall discrepancy rate is the number of cases for this body part divided by the total number of discrepant cases. Overall clinical significance is the rate of clinically significant cases divided by the total number of radiographs completed during the study period.

Table 4. Summary of prior studies

Authors	Age	Discordant rate (%)	Clinically significant (%)	N
Gratton ¹	>15 years	3.4	2.8	12,395
Higgonson ²	<14 years	26.1	4.8	1,039
Klein ³	Pediatric	15.95	1.1	2,083
Nitowski ⁴	All ages	0.95	0.2	14,046
Petinaux ⁵	>18 years	3.73	0.056	151,693
Walsh-Kelly ⁶	<18 years	9.53	0.7	1,994
Shirm ⁷	Pediatric	3.50	0.6	8,174
Simon ⁸	Pediatric	9.8	6.8	707
Soudack ^{9*}	<36 years	28	6.3	1,196
Fleisher ¹⁰	≤19 years	8.9	1.2	564
Benger ¹¹	All ages	0.28	0	1,489
Kim ^{12†}	All ages	0.77	0.47	10,243
Minnes ^{13‡}	<18 years	16.2	8.7	205

*Included only CXR images.

†Looked at minor trauma patients only. The study is based on the number of patients, not radiographs.

‡This study looked at extremities only.

Several factors can explain the low error rates found in our study. In contrast to most earlier studies, all radiographs were interpreted and documented by the staff EP, not by trainees or house staff. Our results are therefore not surprising, given that most of the staff reporting these images are either fellowship trainees or have significant experience in pediatric emergency medicine. Another possible factor is the low rate of false positives as a proportion of total discrepancies (45, 15.4%). Radiologists may be less inclined to flag a discrepancy if they consider it a clinically unimportant “overcall” that will not affect patient management, such as an EP’s decision to treat a patient with antibiotics for presumed pneumonia. EPs are likely to prescribe antibiotics if there is clinical concern or suspicion, even in the absence of definitive findings of pneumonia on imaging. Earlier studies of inter-observer reliability among staff radiologists in

diagnosing pneumonia on the basis of chest radiographs found a wide range of agreement, with Kappa values between 0.54 and 0.92.^{16–20} This variation in interpretation suggests that with the added benefit of clinical findings, EPs may be more likely to make an accurate diagnosis.

This study was limited to a single-site tertiary care pediatric centre, and all of the cases were flagged and placed in the discrepancy folder by the interpreting EP. Although all of the images and reports may not have been highlighted, reminders were emailed to all staff, and an information sheet was placed in the radiology viewer system. Finally, no discrepancy was found in some cases initially flagged as discrepant. These errors occurred when EPs failed to place their readings in the view box, indicating there are systems issues rather than discrepancies. Further studies are needed to assess the validity of our results as well as generalizability by EPs

who work in the community or are not pediatric emergency fellowship trained.

CONCLUSION

In summary, the accuracy of ED staff physicians in interpreting radiographs is high, and the frequency of errors requiring a change in patient management is very low. The majority of errors occurred with radiographs of the chest and upper extremities. The low rate of clinically significant discrepancy allows safe management based on EP interpretation.

Acknowledgements: The abstract for this paper was presented at the European Congress on Emergency Medicine, European Society of Emergency Medicine, Torino, Italy, October 2015. It won the Best Young Scientist Award. JT and RV conceived the study, designed the trial, and obtained research ethics approval. JT, SS, and RV supervised the conduct of the trial and data collection. JT, SS, and RV undertook the recruitment of patients and managed the data, including quality control. JT, SS, and RV provided statistical advice on study design and analysed the data. JT drafted the manuscript, and all authors contributed substantially to its revision. RV takes responsibility for the paper as a whole.

Competing interests: None declared.

REFERENCES

1. Gratton MC, Salomone JA 3rd, Watson WA. Clinically significant radiograph misinterpretations at an emergency medicine residency program. *Ann Emerg Med* 1990; 19(5):497-502.
2. Higginson I, Vogel S, Thompson J, et al. Do radiographs requested from a paediatric emergency department in New Zealand need reporting? *Emerg Med Australas* 2004; 16(4):288-94.
3. Klein EJ, Koenig M, Diekema DS, et al. Discordant radiograph interpretation between emergency physicians and radiologists in a pediatric emergency department. *Pediatr Emerg Care* 1999;15(4):245-8.
4. Nitowski LA, O'Connor RE, Reese CL. The rate of clinically significant plain radiograph misinterpretation by faculty in an emergency medicine residency program. *Acad Emerg Med* 1996;3(8):782-9.
5. Petinaux B, Bhat R, Boniface K, et al. Accuracy of radiographic readings in the emergency department. *Am J Emerg Med* 2011;29(1):18-25.
6. Walsh-Kelly CM, Melzer-Lange MD, Hennes HM, et al. Clinical impact of radiograph misinterpretation in a pediatric ED and the effect of physician training level. *Am J Emerg Med* 1995;13(3):262-4.
7. Shirm SW, Graham CJ, Seibert JJ, et al. Clinical effect of a quality assurance system for radiographs in a pediatric emergency department. *Pediatr Emerg Care* 1995;11(6):351-4.
8. Simon HK, Khan NS, Nordenberg DF, et al. Pediatric emergency physician interpretation of plain radiographs: is routine review by a radiologist necessary and cost-effective? *Ann Emerg Med* 1996;27(3):295-8.
9. Soudack M, Raviv-Zilka L, Ben-Shlush A, et al. Who should be reading chest radiographs in the pediatric emergency department? *Pediatr Emerg Care* 2012;28(10):1052-4.
10. Fleisher G, Ludwig S, McSorley M. Interpretation of pediatric x-ray films by emergency department pediatricians. *Ann Emerg Med* 1983;12(3):153-8.
11. Benger JR, Lyburn ID. What is the effect of reporting all emergency department radiographs? *Emerg Med J* 2003; 20(1):40-3.
12. Kim SJ, Lee SW, Hong YS, et al. Radiological misinterpretations by emergency physicians in discharged minor trauma patients. *Emerg Med J* 2012;29(8):635-9.
13. Minnes BG, Sutcliffe T, Klassen TP. Agreement in interpretation of extremity radiographs if injured children and adolescents. *Acad Emerg Med* 1995;2(9):826-30.
14. Gilbert EH, Lowenstein SR, Koziol-McLain J, et al. Cart reviews in emergency medicine research: what are the methods? *Ann Emerg Med* 1996;27:305-8.
15. Altman D. *Practical statistics for medical research*. London (UK): Chapman and Hall; 1991.
16. Albaum MN, Hill LC, Murphy M, et al. Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT investigators. *Chest* 1996;110:343-50.
17. Loeb MB, Carusone SBC, Marrie TJ, et al. Interobserver reliability of radiologists' interpretations of mobile chest radiographs for nursing home - acquired pneumonia. *J Am Med Dir Assoc* 2006;7(7):416-9.
18. Melbye H, Dale K. Interobserver variability in the radiographic diagnosis of adult outpatient pneumonia. *Acta Radiol* 1992;33(1):79-81.
19. Young M, Marrie TJ. Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. *Arch Intern Med* 1994;154:2729-32.
20. Moncada DC, Rueda ZV, Macías A, et al. Reading and interpretation of chest X-ray in adults with community-acquired pneumonia. *Braz J Infect Dis* 2011;15(6):540-6.