BMJ Open

Now you see me: a pragmatic comparison between first and final radiologic diagnosis in the emergency department

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020230
Article Type:	Research
Date Submitted by the Author:	23-Oct-2017
Complete List of Authors:	Mattsson, Björn; Hopital neuchatelois, Departement des urgences Ertman, David; Inselspital, University of Bern, Department of Emergency Medicine Exadaktylos, Aristomenis; Inselspital, University of Bern, Department of Emergency Medicine Martinolli, Luca; Klinik Linde, Dept. of Emergency Medicine Hautz, Wolf; Inselspital, University of Bern, Department of Emergency Medicine
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Radiology and imaging
Keywords:	diagnostic error, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, ACCIDENT & EMERGENCY MEDICINE, RADIOLOGY & IMAGING

Now you see me: a pragmatic comparison between first and final radiologic diagnosis in the emergency department

Björn Mattsson[1*], David Ertman[2], Aristomenis Exadaktylos[2], Luca Martinolli[3], Wolf E Hautz[2]

1 Departement des urgences, Hôpital neuchâtelois, Neuchâtel, Switzerland
2 Department of Emergency Medicine, Inselspital, University Hospital Berne, Berne, Switzerland
3 Notfallstation, Privatklinik Linde, Biel, Switzerland

Correspondence to:
Wolf Hautz, M.D., M.M.E.
Dept. of Emergency Medicine
Freiburgstrasse 16c
3010 Bern, Switzerland
Email: wolf.hautz@insel.ch
Phone: +41 31 632 45 87

Words : 1928

Abstract

Objective

To 1) compare timely but preliminary and definitive but delayed radiologic reports in a large urban level one trauma center, to 2) assess the clinical significance of their differences and to 3) identify clinical predictors of such differences.

Design, Setting and Participants

We performed a retrospective record review for all 2'914 patients who presented to our university affiliated emergency department during a 6-week period. In those that underwent radiologic imaging, we compared the patients discharge letter from the emergency department to the definitive radiologic report. All identified discrepancies were assessed regarding their clinical significance by trained raters, independent and in duplicate. A binary logistic regression was performed to calculate the likelihood of discrepancies based on readily available clinical data.

Results

1'522 patients had radiographic examinations performed. Rater agreement on the clinical significance of identified discrepancies was substantial (kappa = 0.86). We found an overall discrepancy rate of 20.35% of which about one third (7.48% overall) are clinically relevant. A logistic regression identified patients age, the imaging modality and the anatomic region under investigation to be predictive of future discrepancies.

Conclusions

Discrepancies between radiologic diagnoses in the emergency department are frequent and readily available clinical factors predict their likelihood. Emergency physicians should reconsider their discharge diagnosis especially in older patients undergoing CT scans of more than one anatomic region.

Strengths and limitations of this study

- Retrospective record review of a real-world patient sample
- Clinically valid comparison between immediate first and delayed final radiologic diagnosis
- Situated in a large urban emergency room, where many diagnoses are first made
- Designed to identify readily available predictors of misdiagnosis such as age, imaging modality and anatomic region

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Introduction

Annually, between 100,000 and 250,000 patients in the United States alone die from medical errors [1,2]. Diagnostic errors are a frequent and the most consequential medical error [2–6], and misdiagnosis thus is one of the greatest concerns for patients in the emergency department (ED) [7]. It furthermore has important economic and legal consequences [8]. Errors in the assessment of radiographs are a potential source of such diagnostic errors. Especially in the ED, diagnostic errors might lead to iatrogenic harm to the patient [9].

In most EDs, plain film radiographs (Xr) are initially interpreted by the treating emergency physician (EP), and a definitive diagnosis by a radiologist is provided hours to days later. More complex exams, including computed tomography scans (CT) or magnetic resonance imaging (MRI), are often interpreted immediately by a (junior) radiologist on duty and findings communicated to the EP [10], while a seniors definitive approval of such reports might follow much later. Often, EPs additionally informally consult radiologists on duty on findings the EPs are uncertain about. Thus, two interpretations of radiographs typically exist in most EDs: an immediately available reading by EPs and potentially junior radiologists and a delayed but more reliable reading by senior radiologists. Treatment and discharge decisions in the ED are typically based on the former due to the time constraints in most EDs.

Previous studies have shown overall discrepancy rates in the interpretation of radiographic images between radiologists and emergency physicians to range between 1.1% [11] and 9.2% [12], although much higher discrepancy rates have been reported for specific types of exams [13]. However, missed radiological findings that would have resulted in an immediate change in the management of a patient have been reported to be exceedingly rare [14].

Whereas differences in the interpretation of radiographic images between radiologists and EPs have been extensively researched, the discrepancies between preliminary results reported in the EDs discharge letter and the definitive radiology report are less well examined. We thus aimed to compare the preliminary findings reported by the ED to the definitive radiologic reports and determine the clinical significance of any differences in order to estimate the resulting degree of consequential diagnostic errors. We further aimed to model the binary outcome discrepancy / no discrepancy based on clinical data readily available to the EP before discharge to provide the EP with an a priori estimate of the probability of error.

Patients and Methods

This study is a retrospective review of all radiologic studies ordered between December 2012 and January 2013 in a large urban academic ED and level one trauma center that saw approximately 38'000 patients in 2013. The ED is staffed by physicians certified in internal medicine, surgery, traumatology and emergency medicine [15]. We retrieved records of all adult patients presenting with traumatic or non-traumatic injury, medical or neurological chief complaints during the study period. Of these patients, we included all those for whom radiologic studies had been ordered. Patients consulting directly with specialist clinics (orthopedics, neurosurgery, hand surgery, plastic surgery, nephrology and urology) for non-urgent reasons were excluded since the procedures of how and when radiological findings are reported to the requesting physicians differ strongly depending on requesting departments. All relevant data, including age, gender, time of day, diagnosis and clinical management as noted in the discharge documents were retrieved from the ED patient management system (ECare ED 2.1.3.0, E.care bvba, Turnhout, Belgium) and entered into a database (Microsoft Excel 14.0, Redmond Washington, USA). Definitive radiologic reports were retrieved from our digital radiologic database (Spectra Workstation IDS 7, Sectra AB, Linköping, Sweden) and imaging modality categorized as either Xr, CT, MRI, Ultrasound (US) or Scintigraphy (SCI). We further coded the body part examined as either head (including face and neck), chest, abdomen, skeletal system or other. The total number of imaging studies in each category was recorded. All patient data was anonymized and the study design was approved by the IRB at Inselspital (Kantonale Ethikkommission Bern) and received ethical approval (KEK BE 394/15).

We subsequently analyzed the preliminary radiologic report as given in the ED-discharge documents and compared them to the definitive radiologic report, which was defined as the gold standard. Two independent reviewers analyzed the data set and noted discrepancies between preliminary and definitive findings. Discrepancies were subsequently categorized by two independent EPs in duplicate as either “clinically significant”, i.e. changing clinical management or “clinically insignificant”. Rater agreement was calculated as Cohens kappa and disagreements resolved by discussion.

Statistical analysis was conducted in SPSS 22 (IBM Corporation, Armonk, USA) and include descriptive statistics (frequency, mean and standard deviation) and a logistic regression model. With the regression we aimed to predict the binary outcome discrepancy versus no discrepancy between the radiologic studies based on the patients age, gender, the imaging modality, the anatomic location of the radiologic study and the time of day. Metric predictor variables (age and time of day) were z-standardized prior to the analysis to ensure comparability within the model. We refined the model stepwise by removing all non-significant predictor variables and report Nagelkerkes R^2 as measure of the models fit together with p-values from Wald statistics and the respective regression coefficients. P-values <0.05 were considered significant.

Results

In the six week study period from December 1st 2012 to January 15th 2013 a total of 2'914 patient visits were recorded in the ED. Of these, a total of 1'522 patients, which corresponds to just over half (52.0%) of all patients, had at least one radiologic study taken and were thus included in the study. Upon presentation, 608 of these patients had been triaged as surgical, 544 patients as medical and 360 patients as neurological emergencies. A majority of patients were male (n=868, 57.0%), the median age was 53.74 years (min 16, max 98, SD 20.9). The majority of studies were ordered during daytime between 07:00h and 20:00h (n=1'086) (Table 1).

Table 1: Total number of patients, overall and clinically significant discrepancies.

Variable	Total number of patients [*N*]	Overall discrepancies [*N* (%)]	Clinically significant discrepancies [*N* (%)]	*p*-value (significant discrepancies)
specialty				
surgery	608	146 (24.0)	39 (6.4)	
medicine	504	130 (23.9)	36 (6.6)	0.031
neurology	360	57 (16.1)	10 (2.8)	
gender				
women	654	135 (20.6)	36 (5.5)	0.911
men	868	230 (23.4)	50 (5.8)	
age				
65 and older	543	176 (32.4)	45 (8.3)	0.002
under 65	979	162 (16.6)	41 (4.2)	
time of presentation				
day time	1'086	238 (21.9)	51 (5.6)	0.903
night time	436	100 (22.9)	25 (5.7)	

A total of 1875 radiological studies were performed, including 776 Xr, 680 CT, 367 MRI, 49 US and 3 SCI. Due to their small number, SCI were excluded from further analysis. The most common radiologic studies ordered were CT of the head and neck (n=343), Xr of the chest (n=319) and MRI of the head (n=329).

Rater agreement on whether or not discrepancies between discharge report and final radiological report were clinically significant was substantial (kappa = 0.86). Overall, 381 discrepancies (20.35%) were found, of which 149 (7.48%) were judged to be clinically significant (Table 2).

Table 2: Number of radiologic studies, overall and clinically significant discrepancies classified according to type of radiologic study.

Radiologic study	Overall	Overall discrepancies [*N* (%)]	Clinically significant discrepancies [*N* (%)]
Xr hand/wrist	87	12 (13.79)	2 (2.3)
Xr thorax	319	74 (23.2)	22 (6.9)
Xr spine	48	11 (22.92)	4 (8.34)
Xr pelvis	51	14 (27.45)	4 (7.84)
Xr knee	56	4 (7.14)	1 (1.79)
Xr ankle/foot	62	10 (16.13)	5 (8.06)
Xr other[a]	153	14 (9.15)	6 (3.92)
Sum XR	**776**	**139 (17.91)**	**44 (5.67)**
CT head/neck	343	63 (18.37)	29 (8.46)
CT thorax	115	39 (33.91)	17 (14.78)
CT abdomen	114	31 (27.19)	9 (7.9)
CT whole body	57	27 (47.39)	14 (24.56)
CT other[b]	51	12 (23.53)	5 (9.8)
Sum CT	**680**	**172 (25.29)**	**74 (10.88)**
MRI head	329	55 (16.72)	12 (3.65)
MRI spine	32	7 (21.88)	6 (18.75)
MRI other[c]	6	4 (66.67)	3 (50)
Sum MRI	**367**	**66 (17.98)**	**21 (5.77)**
Sum US	**49**	**4 (8.16)**	**1 (2.04)**
Sum Total	**1872**	**381 (20.35)**	**140 (7.48)**

[a] XR skull/OPT and body parts not otherwise specified

[b] CTs of soft tissue and bone and from body parts not otherwise specified

[c] body parts not otherwise specified

Whether or not a discrepancy between an initial radiological assessment and the definitive report by the department of radiology did or did not occur was predicted by several clinical variables. Logistic regression identified patients age, modality of imaging and anatomic region of the radiological study to be significant predictors (all *p*-values <0.05), while time of day and patient gender had no significant predictive value. The model fits the data fairly well (R^2=0.112) and would correctly predict outcome on 77.8% of the cases. Details of the model are given in Table 3.

Table 3: Results of the refined logistic regression model to predict a discrepancy between ED discharge report and definitive radiological report based on clinical characteristics.

Predictor	Regression coefficient	*p*-value
Age	0.472	<0.001
Imaging modality	-0.649	0.006
Anatomic region	-1.085	<0.001
Constant	-0.584	<0.001

Figs 1 and 2 show the change in probability of a discrepancy predicted by the regression model based on age, imaging modality and anatomic region of the study.

Fig 1. Probability of a discrepancy between first and final radiological diagnosis depending on body part over patients age.

Fig 2. Probability of a discrepancy between first and final radiological diagnosis depending on image modality over patients age.

Discussion

Radiologic images are an important part of medical diagnosis. In many EDs, patients radiographs are initially assessed by ED physicians as well as junior radiologists and treatment is determined based on their joint interpretation. A more definitive interpretation by senior radiologists is typically only available with a considerable delay.

Comparing the interpretation of radiographs in the discharge letter of ED patients to the final report from radiology, we found an overall discrepancy rate of 20.35%. Slightly more than one third of these (7.48% overall) were deemed clinically relevant by two independent expert raters. The estimates of error from our study are well within the range of previous publications [11,12,14,16–18], which however mainly compared EPs reading of radiographs to senior radiologists. Such a comparison is only directly applicable to very small EDs as lager center such as the one under investigation in our study typically have at least a junior radiologist on duty around the clock. Thus, the study reported here extends previous findings to the clinical reality in tertiary centers.

Using a liner regression to model the likelihood of a discrepancy between first and final radiological diagnosis, we found several readily available clinical factors to be predictive of an error. These factors namely are patient age, imaging modality and region of the body under investigation. The factors are both, plausible from a clinical perspective as well as in line with the sparse previous findings on the issue. Age has been previously found to be associated with diagnostic error [19] and adverse events in the ED [20], likely because radiographs become harder to interpret in the presence of age-related or chronic findings.

We further found imaging modality and region of the body under investigation to be predictive of a discrepancy. From a clinical as well as a mathematical perspective, it is plausible that both, more than one modality as well as more than one body region under investigation increase the likelihood of a discrepancy. Furthermore, two well-known cognitive sources of error are premature closure, i.e. the failure to consider alternative diagnoses [21] as well as satisfaction of search, i.e. the termination of a diagnostic search after successful identification of one pathological finding [22]. Both phenomena are less likely to occur with increasing expertise on a subject [23].

One counterintuitive finding at first sight is the rather low discrepancy rates in MRIs of the head as well as in patients triaged as neurological emergencies. We assume these findings to be related because most MRIs of the head are ordered in patients with neurological chief complaints. One reason why the discrepancy rate in these patients is rather low may be the fact that neurologists are highly trained in interpreting cerebral MRI [24]. Furthermore, the variety of possible interpretations is lower in cerebral MRIs than in a patient population with highly diverse body regions under investigation commonly triaged as medical or surgical chief complaints.

Our study has several limitations. First, this is a retrospective study susceptible to both, documentation bias and hindsight bias [25]. Prospective studies of diagnostic error are imperative and currently ongoing [26]. Second, our study design does not allow us to discern whether the discrepancies identified between the final radiologic report and the findings documented in the ED discharge documentation are due to misinterpretations by the junior radiologist, the discharging EPs or failed communication between the radiologist and the EP. However, regardless of where the error originates, it is the differences pragmatically assessed in this study that arguably matter most to the patient. Future studies focusing on collaboration in healthcare are needed [27] because failed teamwork has been repeatedly identified as an important source of diagnostic error [6]. Last, one obvious question is why the estimates of error with and without consequence very by an order of magnitude from author to author. We would offer two potential explanations. First, the definition of what constitutes a diagnostic error in general [28] and a clinically significant differences in radiologic diagnosis specifically is highly variable between publications, potentially resulting in different estimates. Second, due to time constrains, EPs may tend to only report findings they deem significant, which may explain the comparatively large number of insignificant differences found in our study.

In conclusion, we found a comparatively large number of discrepancies between radiologic findings in patients discharge documentation compared to the final radiological report and identified age, imaging modality and body parts under investigation to be predictive of such discrepancies. All three predictors are readily available in clinical practice and should prompt EPs to reconsider their discharge diagnosis especially in older patients undergoing CT scans of more than one anatomic region.

Competing interests

WEH received financial compensation for educational consultancies from the AO foundation, Zurich and speakers honorarium from Mundipharma Medical, Basel. All other authors declare no competing interests.

Acknowledgements and Contributions

The authors would like to thank Sabina Utiger for her support in data management.

Authors contributions were as follows:

Study Design: BM, DE, AE, LM, WEH

Data Collection: BM, DE, LM

Expert Raters: LM, AE

Statistical Analysis: WEH

Interpretation of Results: BM, DE, AE, LM, WEH

Drafting the manuscript: BM, DE, WEH

Substantial contribution to the manuscript for important intellectual content: BM, DE, AE, LM, WEH

Original data are available from the corresponding author upon request, provided the requesting person or party fulfills the data sharing requirements laid out by Kantonale Ethikkomission Bern.

References

1. Andel C, Davidow SL, Hollander M, Moreno DA. The economics of health care quality and medical errors. *J Health Care Finance*. 2012;39(1):39-50.
2. Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human*. Washington, D.C.: National Academies Press; 2000.
3. Schwartz A, Elstein AS. Clinical reasoning in medicine. In: *Clinical reasoning in the health professions*. Edinburgh: Elsevier; 2008:223-34.
4. Leape LL, Brennan TA, Laird N, et al. The Nature of Adverse Events in

Hospitalized Patients. *N Engl J Med*. 1991;324(6):377-384.

5. Lu T-C, Tsai C-L, Lee C-C, et al. Preventable deaths in patients admitted from emergency department. *Emerg Med J*. 2006;23(6):452-5.
6. Balogh EP, Miller BT, Ball JR. *Improving Diagnosis in Health Care*. (Balogh EP, Miller BT, Ball JR, eds.). Washington, D.C.: National Academies Press; 2015.
7. Burroughs TE, Waterman AD, Gallagher TH, et al. Patient concerns about medical errors in emergency departments. *Acad Emerg Med*. 2005;12(1):57-64.
8. Brown TW, McCarthy ML, Kelen GD, Levy F. An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. *Acad Emerg Med*. 2010;17(5):553-60.
9. Lawson CM, Daley BJ, Ormsby CB, Enderson B. Missed injuries in the era of the trauma scan. *J Trauma*. 2011;70(2):452-456-458.
10. Hunter TB, Krupinski EA, Hunt KR, Erly WK. Emergency department coverage by academic departments of radiology. *Acad Radiol*. 2000;7(3):165-70.
11. Warren JS, Lara K, Connor PD, Cantrell J, Hahn RG. Correlation of emergency department radiographs: results of a quality assurance review in an urban community hospital setting. *J Am Board Fam Pract*. 6(3):255-9.
12. Fleisher G, Ludwig S, McSorley M. Interpretation of pediatric x-ray films by emergency department pediatricians. *Ann Emerg Med*. 1983;12(3):153-8.
13. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(5 Suppl):S2-23.
14. Petinaux B, Bhat R, Boniface K, Aristizabal J. Accuracy of radiographic readings in the emergency department. *Am J Emerg Med*. 2011;29(1):18-25.
15. Exadaktylos A, Hautz WE. Emergency Medicine in Switzerland. *ICU Manag Pract*. 2015;15(4):160-2.
16. Gratton MC, Salomone J a, Watson W a. Clinically significant radiograph misinterpretations at an emergency medicine residency program. *Ann Emerg Med*. 1990;19(5):497-502.
17. Brunswick JE, Ilkhanipour K, Seaberg DC, McGill L. Radiographic interpretation in the emergency department. *Am J Emerg Med*. 1996;14(4):346-8.
18. Nitowski L a, 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.

19. Peng A, Rohacek M, Ackermann S, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. *Swiss Med Wkly*. 2015;145(March):1-9.
20. Aminzadeh F, Dalziel WB. Older adults in the emergency department: A systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann Emerg Med*. 2002;39(3):238-247.
21. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106.
22. Berbaum KS, Franken EA, Dorfman DD, Caldwell RT, Lu CH. Can order of report prevent satisfaction of search in abdominal contrast studies? *Acad Radiol*. 2005;12(1):74-84.
23. Ericsson KA, Charness N, Feltovich PJ, Hoffman RR. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press; 2006.
24. Flanders AE, Flanders SJ, Friedman DP, Tartaglino LM, Russell KM. Performance of neuroradiologic examinations by nonradiologists. *Radiology*. 1996;198(3):825-30.
25. Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, Norman G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf*. 2017;26(2):104-110.
26. Hautz SC, Schuler L, Kämmer JE, et al. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. *BMJ Open*. 2016;6(5):e011585.
27. Hautz WE, Kämmer JE, Exadaktylos A, Hautz SC. How thinking about groups is different from groupthink. *Med Educ*. 2017;51(2):229-229.
28. Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. *BMJ Qual Saf*. 2013;22(Suppl 2):ii52-ii57.

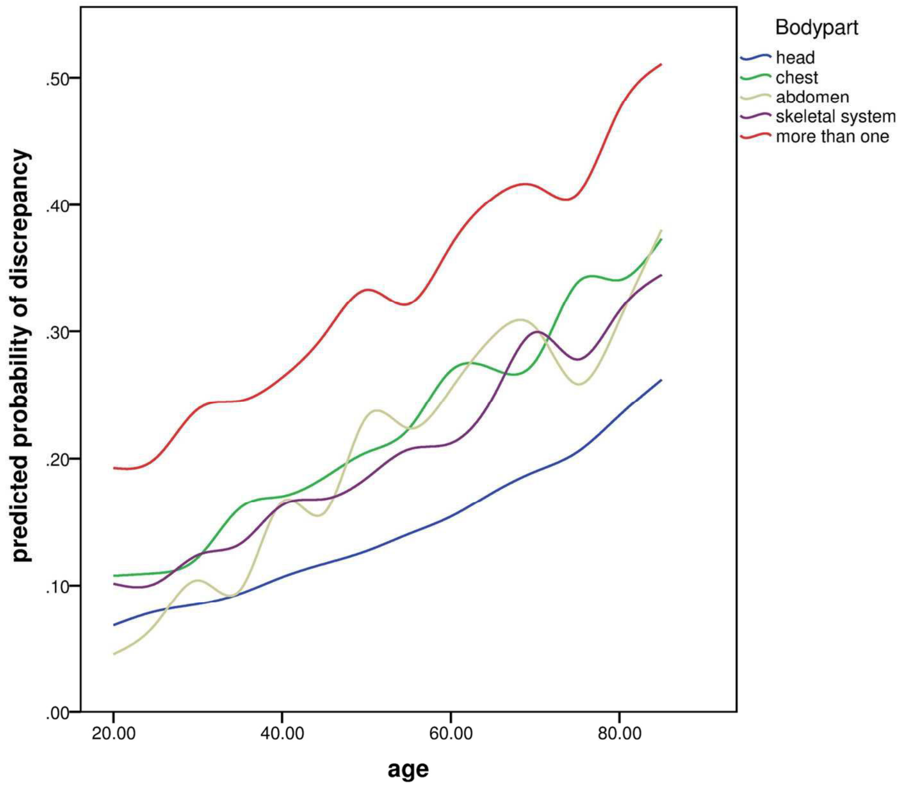

Fig 1. Probability of a discrepancy between first and final radiological diagnosis depending on body part over patients age.

131x105mm (300 x 300 DPI)

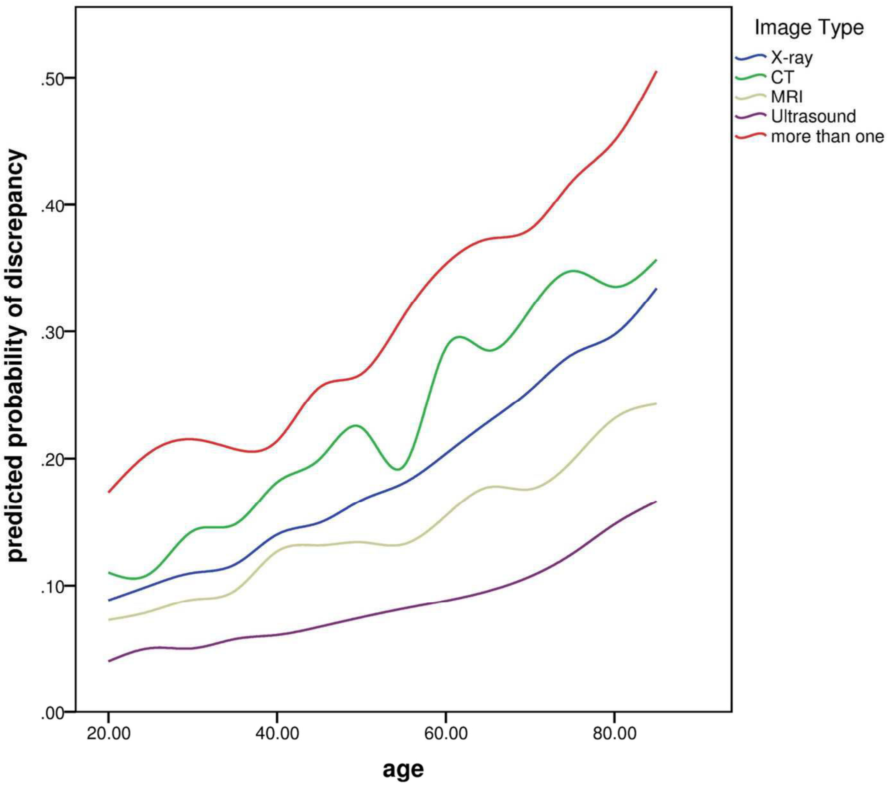

Fig 2. Probability of a discrepancy between first and final radiological diagnosis depending on image modality over patients age.

131x105mm (300 x 300 DPI)

BMJ Open

Now you see me: a pragmatic cohort study comparing first and final radiologic diagnoses in the emergency department

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020230.R1
Article Type:	Research
Date Submitted by the Author:	29-Nov-2017
Complete List of Authors:	Mattsson, Björn; Hopital neuchatelois, Departement des urgences Ertman, David; Inselspital, University of Bern, Department of Emergency Medicine Exadaktylos, Aristomenis; Inselspital, University of Bern, Department of Emergency Medicine Martinolli, Luca; Klinik Linde, Dept. of Emergency Medicine Hautz, Wolf; Inselspital, University of Bern, Department of Emergency Medicine
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Radiology and imaging
Keywords:	diagnostic error, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, ACCIDENT & EMERGENCY MEDICINE, RADIOLOGY & IMAGING

Now you see me: a pragmatic cohort study comparing first and final radiologic diagnoses in the emergency department

Björn Mattsson1*, David Ertman2, Aristomenis Exadaktylos2, Luca Martinolli3, Wolf E Hautz2

1 Departement des urgences, Hôpital neuchâtelois, Neuchâtel, Switzerland
2 Department of Emergency Medicine, Inselspital, University Hospital Berne, Berne, Switzerland
3 Notfallstation, Privatklinik Linde, Biel, Switzerland

Correspondence to:
Wolf Hautz, M.D., M.M.E.
Dept. of Emergency Medicine
Freiburgstrasse 16c
3010 Bern, Switzerland
Email: wolf.hautz@insel.ch
Phone: +41 31 632 45 87

Words : 1928

Abstract

Objective

To 1) compare timely but preliminary and definitive but delayed radiologic reports in a large urban level one trauma center, to 2) assess the clinical significance of their differences and to 3) identify clinical predictors of such differences.

Design, Setting and Participants

We performed a retrospective record review for all 2'914 patients who presented to our university affiliated emergency department during a 6-week period. In those that underwent radiologic imaging, we compared the patients discharge letter from the emergency department to the definitive radiologic report. All identified discrepancies were assessed regarding their clinical significance by trained raters, independent and in duplicate. A binary logistic regression was performed to calculate the likelihood of discrepancies based on readily available clinical data.

Results

1'522 patients had radiographic examinations performed. Rater agreement on the clinical significance of identified discrepancies was substantial (kappa = 0.86). We found an overall discrepancy rate of 20.35% of which about one third (7.48% overall) are clinically relevant. A logistic regression identified patients age, the imaging modality and the anatomic region under investigation to be predictive of future discrepancies.

Conclusions

Discrepancies between radiologic diagnoses in the emergency department are frequent and readily available clinical factors predict their likelihood. Emergency physicians should reconsider their discharge diagnosis especially in older patients undergoing CT scans of more than one anatomic region.

Strengths and limitations of this study

- Retrospective record review of a real-world patient sample
- Clinically valid comparison between immediate first and delayed final radiologic diagnosis
- Single center study, situated in a large urban emergency room, where many diagnoses are first made
- Designed to identify readily available predictors of misdiagnosis such as age, imaging modality and anatomic region
- Unable to determine long term consequences due to retrospective design

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Introduction

Annually, between 100,000 and 250,000 patients in the United States alone die from medical errors [1,2]. Diagnostic errors are a frequent and the most consequential medical error [2–6], and misdiagnosis thus is one of the greatest concerns for patients in the emergency department (ED) [7]. It furthermore has important economic and legal consequences [8]. Errors in the assessment of radiographs are a potential source of such diagnostic errors. Especially in the ED, diagnostic errors might lead to iatrogenic harm to the patient [9].

In most EDs, plain film radiographs (Xr) are initially interpreted by the treating emergency physician (EP), and a definitive diagnosis by a radiologist is provided hours to days later. More complex exams, including computed tomography scans (CT) or magnetic resonance imaging (MRI), are often interpreted immediately by a (junior) radiologist on duty and findings communicated to the EP [10], while a seniors definitive approval of such reports might follow much later. Often, EPs additionally informally consult radiologists on duty on findings the EPs are uncertain about. Thus, two interpretations of radiographs typically exist in most EDs: an immediately available reading by EPs and potentially junior radiologists and a delayed but more reliable reading by senior radiologists. Treatment and discharge decisions in the ED are typically based on the former due to the time constraints in most EDs.

Previous studies have shown overall discrepancy rates in the interpretation of radiographic images between radiologists and emergency physicians to range between 1.1% [11] and 9.2% [12], although much higher discrepancy rates have been reported for specific types of exams [13]. However, missed radiological findings that would have resulted in an immediate change in the management of a patient have been reported to be exceedingly rare [14].

Whereas differences in the interpretation of radiographic images between radiologists and EPs have been extensively researched, the discrepancies between preliminary results reported in the EDs discharge letter and the definitive radiology report are less well examined. We thus aimed to compare the preliminary findings reported by the ED to the definitive radiologic reports and determine the clinical significance of any differences in order to estimate the resulting degree of consequential diagnostic errors. We further aimed to model the binary outcome discrepancy / no discrepancy based on clinical data readily available to the EP before discharge to provide the EP with an a priori estimate of the probability of error.

Patients and Methods

This study is a retrospective review of all radiologic studies ordered between December 2012 and January 2013 in a large urban academic ED and level one trauma center that saw approximately 38'000 patients in 2013. The ED is staffed by physicians certified in internal medicine, surgery, traumatology and emergency medicine [15]. We retrieved records of all adult patients presenting with traumatic or non-traumatic injury, medical or neurological chief complaints during the study period. Of these patients, we included all those for whom radiologic studies had been ordered. Patients consulting directly with specialist clinics (orthopedics, neurosurgery, hand surgery, plastic surgery, nephrology and urology) for non-urgent reasons were excluded since the procedures of how and when radiological findings are reported to the requesting physicians differ strongly depending on requesting departments. All relevant data, including age, gender, time of day, diagnosis and clinical management as noted in the discharge documents were retrieved from the ED patient management system (ECare ED 2.1.3.0, E.care bvba, Turnhout, Belgium) and entered into a database (Microsoft Excel 14.0, Redmond Washington, USA). Definitive radiologic reports were retrieved from our digital radiologic database (Spectra Workstation IDS 7, Sectra AB, Linköping, Sweden) and imaging modality categorized as either Xr, CT, MRI, Ultrasound (US) or Scintigraphy (SCI). We further coded the body part examined as either head (including face and neck), chest, abdomen, skeletal system or other. The total number of imaging studies in each category was recorded. All patient data was anonymized and the study design was approved by the IRB at Inselspital (Kantonale Ethikkommission Bern) and received ethical approval (KEK BE 394/15).

We subsequently analyzed the preliminary radiologic report as given in the ED-discharge documents and compared them to the definitive radiologic report, which was defined as the gold standard. Two independent reviewers analyzed the data set and noted discrepancies between preliminary and definitive findings. Discrepancies were subsequently categorized by two independent EPs in duplicate as either "clinically significant", i.e. changing clinical management or "clinically insignificant". Rater agreement was calculated as Cohens kappa and disagreements resolved by discussion. For some patients, e.g. those with known comorbidities or after major trauma, more than one abnormal finding may be present in any radiographic imaging. Whenever rater encountered a discrepancy between first and final radiologic report, we counted this as a discrepancy. However, each image with a discrepancy was counted only once, regardless of the total number of discrepancies present on that image, leading to a conservative estimate of the total number of discrepancies. For each discrepancy identified per image, the clinical relevance was assessed separately but again counted only once if present.

Statistical analysis was conducted in SPSS 22 (IBM Corporation, Armonk, USA) and include descriptive statistics (frequency, mean and standard deviation) and a logistic regression model. With the regression we aimed to predict the binary outcome discrepancy versus no discrepancy between the radiologic studies based on the patients age, gender, the imaging modality, the anatomic location of the radiologic study and the time of day. Metric predictor variables (age and time of day) were z-standardized prior to the analysis to ensure comparability within the model. We refined the model stepwise by removing all non-significant predictor variables and report Nagelkerkes R2 as measure of the models fit together with p-values from Wald statistics and the respective regression coefficients. P-values <0.05 were considered significant. We planned to assume data to be missing at random and thus impute missing data by means of a maximum likelihood estimation. We did however not encounter any missing data in the variables assessed in this study, which may result from the fact that we only retrieved very basic patient data such as gender or age.

Results

In the six week study period from December 1st 2012 to January 15th 2013 a total of 2'914 patient visits were recorded in the ED. Of these, a total of 1'522 patients, which corresponds to just over half (52.0%) of all patients, had at least one radiologic study taken and were thus included in the study. Upon presentation, 608 of these patients had been triaged as surgical, 544 patients as medical and 360 patients as neurological emergencies. A majority of patients were male (n=868, 57.0%), the median age was 53.74 years (min 16, max 98, SD 20.9). The majority of studies were ordered during daytime between 07:00h and 20:00h (n=1'086) (Table 1).

Table 1: Total number of patients, overall and clinically significant discrepancies.

Variable	Total number of patients [*N*]	Overall discrepancies [*N* (%)]	Clinically significant discrepancies [*N* (%)]	*p*-value (significant discrepancies)
specialty				
surgery	608	146 (24.0)	39 (6.4)	
medicine	504	130 (23.9)	36 (6.6)	0.031
neurology	360	57 (16.1)	10 (2.8)	
gender				
women	654	135 (20.6)	36 (5.5)	0.911
men	868	230 (23.4)	50 (5.8)	
age				
65 and older	543	176 (32.4)	45 (8.3)	0.002
under 65	979	162 (16.6)	41 (4.2)	
time of presentation				
day time	1'086	238 (21.9)	51 (5.6)	0.903
night time	436	100 (22.9)	25 (5.7)	

A total of 1875 radiological studies were performed, including 776 Xr, 680 CT, 367 MRI, 49 US and 3 SCI. Due to their small number, SCI were excluded from further analysis. The most common radiologic studies ordered were CT of the head and neck (n=343), Xr of the chest (n=319) and MRI of the head (n=329).

Rater agreement on whether or not discrepancies between discharge report and final radiological report were clinically significant was substantial (kappa = 0.86). Overall, 381 discrepancies (20.35%) were found, of which 149 (7.48%) were judged to be clinically significant (Table 2).

Table 2: Number of radiologic studies, overall and clinically significant discrepancies classified according to type of radiologic study.

Radiologic study	Overall	Overall discrepancies [*N* (%)]	Clinically significant discrepancies [*N* (%)]
Xr hand/wrist	87	12 (13.79)	2 (2.3)
Xr thorax	319	74 (23.2)	22 (6.9)
Xr spine	48	11 (22.92)	4 (8.34)
Xr pelvis	51	14 (27.45)	4 (7.84)
Xr knee	56	4 (7.14)	1 (1.79)
Xr ankle/foot	62	10 (16.13)	5 (8.06)
Xr othera	153	14 (9.15)	6 (3.92)
Sum XR	**776**	**139 (17.91)**	**44 (5.67)**
CT head/neck	343	63 (18.37)	29 (8.46)
CT thorax	115	39 (33.91)	17 (14.78)
CT abdomen	114	31 (27.19)	9 (7.9)
CT whole body	57	27 (47.39)	14 (24.56)
CT otherb	51	12 (23.53)	5 (9.8)
Sum CT	**680**	**172 (25.29)**	**74 (10.88)**
MRI head	329	55 (16.72)	12 (3.65)
MRI spine	32	7 (21.88)	6 (18.75)
MRI otherc	6	4 (66.67)	3 (50)
Sum MRI	**367**	**66 (17.98)**	**21 (5.77)**
Sum US	**49**	**4 (8.16)**	**1 (2.04)**
Sum Total	**1872**	**381 (20.35)**	**140 (7.48)**

a XR skull/OPT and body parts not otherwise specified

b CTs of soft tissue and bone and from body parts not otherwise specified

c body parts not otherwise specified

An example for a discrepancy judged as not clinically relevant is the CT scan of the head of patient #27, who was found unconscious. The radiographic report of the ED documents “no pathologies in contrast enhanced CT scan of the head”, while the final report points to “no explanation for acute unconsciousness identifiable, signs of chronic sinusitis”. The patient was found to be intoxicated with mixed substances. A relevant discrepancy for example was identified in patient #51, who presented with an acute abdomen due to a perforated sigma-diverticulitis. While the ERs report mentions this diagnosis of the CT of the abdomen, it fails to mention the infiltrate in the lower sections of the left lung, that the final report identified.

Whether or not a discrepancy between an initial radiological assessment and the definitive report by the department of radiology did or did not occur was predicted by several clinical variables. Logistic regression identified patients age, modality of imaging and anatomic region of the radiological study to be significant predictors (all *p*-values <0.05), while time of day and patient gender had no significant predictive value. The model fits the data fairly well (R2=0.112) and would correctly predict outcome on 77.8% of the cases. Details of the model are given in Table 3.

Table 3: Results of the refined logistic regression model to predict a discrepancy between ED discharge report and definitive radiological report based on clinical characteristics.

Predictor	Regression coefficient	*p*-value
Age	0.472	<0.001
Imaging modality	-0.649	0.006
Anatomic region	-1.085	<0.001
Constant	-0.584	<0.001

Figs 1 and 2 show the change in probability of a discrepancy predicted by the regression model based on age, imaging modality and anatomic region of the study.

Fig 1. Probability of a discrepancy between first and final radiological diagnosis depending on body part over patients age.

Fig 2. Probability of a discrepancy between first and final radiological diagnosis depending on image modality over patients age.

Discussion

Radiologic images are an important part of medical diagnosis. In many EDs, patients radiographs are initially assessed by ED physicians as well as junior radiologists and treatment is determined based on their joint interpretation. A more definitive interpretation by senior radiologists is typically only available with a considerable delay.

Comparing the interpretation of radiographs in the discharge letter of ED patients to the final report from radiology, we found an overall discrepancy rate of 20.35%. Slightly more than one third of these (7.48% overall) were deemed clinically relevant by two independent expert raters. The estimates of error from our study are well within the range of previous publications [11,12,14,16–18], which however mainly compared EPs reading of radiographs to senior radiologists. Such a comparison is only directly applicable to very small EDs as larger center such as the one under investigation in our study typically have at least a junior radiologist on duty around the clock. Thus, the study reported here extends previous findings to the clinical reality in tertiary centers. A previous review of diagnostic error in medicine in general found the rate of critical discrepancies between a first and a second reading of images in visual specialties such as radiology, dermatology or pathology to range between 2-5% [19], just below the rate of discrepancies the raters deemed clinically relevant in our study.

Using a liner regression to model the likelihood of a discrepancy between first and final radiological diagnosis, we found several readily available clinical factors to be predictive of an error. These factors namely are patient age, imaging modality and region of the body under investigation. The factors are both, plausible from a clinical perspective as well as in line with the sparse previous findings on the issue. Age has been previously found to be associated with diagnostic error [20] and adverse events in the ED [21], likely because radiographs become harder to interpret in the presence of age-related or chronic findings.

We further found imaging modality and region of the body under investigation to be predictive of a discrepancy. From a clinical as well as a mathematical perspective, it is plausible that both, more than one modality as well as more than one body region under investigation increase the likelihood of a discrepancy. Furthermore, two well-known cognitive sources of error are premature closure, i.e. the failure to consider alternative diagnoses [22] as well as satisfaction of search, i.e. the termination of a diagnostic search after successful identification of one pathological finding [23]. Both phenomena are less likely to occur with increasing expertise on a subject [24]. Consequently, some authors have argued that the interpretation of any medical image should be exclusively left to experienced radiologists [25], while others argue that non radiologists should simply be better trained [26], especially given the increasing availability of radiographic imaging.

One counterintuitive finding at first sight is the rather low discrepancy rates in MRIs of the head as well as in patients triaged as neurological emergencies. We assume these findings to be related because most MRIs of the head are ordered in patients with neurological chief complaints. One reason why the discrepancy rate in these patients is rather low may be the fact that neurologists are highly trained in interpreting cerebral MRI [27]. Furthermore, the variety of possible interpretations is lower in cerebral MRIs than in a patient population with highly diverse body regions under investigation commonly triaged as medical or surgical chief complaints. Also, the likelihood of a coincidental finding in an MRI of the head, that is not related to the ER presentation and thus not actively searched for, is likely smaller than in e.g. a CT-scan of the abdomen, where there simply is more to see an therefor a higher probability of an abnormality.

Our study has several limitations. First, this is a retrospective study susceptible to both, documentation bias and hindsight bias [28]. Prospective studies of diagnostic error are imperative and currently ongoing [29]. Second, our study design does not allow us to discern whether the discrepancies identified between the final radiologic report and the findings documented in the ED discharge documentation are due to misinterpretations by the junior radiologist, the discharging EPs or failed communication between the radiologist and the EP. However, regardless of where the error originates, it is the differences pragmatically assessed in this study that arguably matter most to the patient. Future studies focusing on collaboration in healthcare are needed [30] because failed teamwork has been repeatedly identified as an important source of diagnostic error [6]. Third, due to the retrospective nature of this study, we are unable to determine if and how the identified discrepancies were acted upon. Future prospective investigations should include a follow up on diagnostic discrepancies. Fourth, the study is a single center cohort study. Results may vary between centers and levels of care.

Last, one obvious question is why the estimates of error with and without consequence very by an order of magnitude from author to author. We would offer two potential explanations. First, the definition of what constitutes a diagnostic error in general [31] and a clinically significant differences in radiologic diagnosis specifically is highly variable between publications, potentially resulting in different estimates. Second, due to time constraints, EPs may tend to only report findings they deem significant, which may explain the comparatively large number of insignificant differences found in our study.

In conclusion, we found a comparatively large number of discrepancies between radiologic findings in patients discharge documentation compared to the final radiological report and identified age, imaging modality and body parts under investigation to be predictive of such discrepancies. All three predictors are readily available in clinical practice and should prompt EPs to reconsider their discharge diagnosis especially in older patients undergoing CT scans of more than one anatomic region.

Competing interests

WEH received financial compensation for educational consultancies from the AO foundation, Zurich and speakers honorarium from Mundipharma Medical, Basel. All other authors declare no competing interests.

Acknowledgements and Contributions

The authors would like to thank Sabina Utiger for her support in data management.

Authors contributions were as follows:

Study Design: BM, DE, AE, LM, WEH

Data Collection: BM, DE, LM

Expert Raters: LM, AE

Statistical Analysis: WEH

Interpretation of Results: BM, DE, AE, LM, WEH

Drafting the manuscript: BM, DE, WEH

Substantial contribution to the manuscript for important intellectual content: BM, DE, AE, LM, WEH

Original data are available from the corresponding author upon request, provided the requesting person or party fulfills the data sharing requirements laid out by Kantonale Ethikkomission Bern.

References

1. Andel C, Davidow SL, Hollander M, Moreno DA. The economics of health care quality and medical errors. *J Health Care Finance*. 2012;39(1):39-50.

2. Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human*. Washington, D.C.: National Academies Press; 2000. doi:10.17226/9728.

3. Schwartz A, Elstein AS. Clinical reasoning in medicine. In: *Clinical Reasoning in the Health Professions*. Edinburgh: Elsevier; 2008:223-234.

4. Leape LL, Brennan TA, Laird N, et al. The Nature of Adverse Events in Hospitalized Patients. *N Engl J Med*. 1991;324(6):377-384. doi:10.1056/NEJM199102073240605.

5. Lu T-C, Tsai C-L, Lee C-C, et al. Preventable deaths in patients admitted from emergency department. *Emerg Med J EMJ*. 2006;23(6):452-455. doi:10.1136/emj.2004.022319.

6. Balogh EP, Miller BT, Ball JR. *Improving Diagnosis in Health Care*. (Balogh EP, Miller BT, Ball JR, eds.). Washington, D.C.: National Academies Press; 2015. doi:10.17226/21794.

7. Burroughs TE, Waterman AD, Gallagher TH, et al. Patient concerns about medical errors in emergency departments. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2005;12(1):57-64. doi:10.1197/j.aem.2004.08.052.

8. Brown TW, McCarthy ML, Kelen GD, Levy F. An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2010;17(5):553-560. doi:10.1111/j.1553-2712.2010.00729.x.

9. Lawson CM, Daley BJ, Ormsby CB, Enderson B. Missed injuries in the era of the trauma scan. *J Trauma*. 2011;70(2):452-456; discussion 456-458. doi:10.1097/TA.0b013e3182028d71.

10. Hunter TB, Krupinski EA, Hunt KR, Erly WK. Emergency department coverage by academic departments of radiology. *Acad Radiol*. 2000;7(3):165-170.

11. Warren JS, Lara K, Connor PD, Cantrell J, Hahn RG. Correlation of emergency department radiographs: results of a quality assurance review in an urban community hospital setting. *J Am Board Fam Pract*. 6(3):255-259.

12. Fleisher G, Ludwig S, McSorley M. Interpretation of pediatric x-ray films by emergency department pediatricians. *Ann Emerg Med*. 1983;12(3):153-158.

13. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(5 Suppl):S2-23. doi:10.1016/j.amjmed.2008.01.001.

14. Petinaux B, Bhat R, Boniface K, Aristizabal J. Accuracy of radiographic readings in the emergency department. *Am J Emerg Med*. 2011;29(1):18-25. doi:10.1016/j.ajem.2009.07.011.

15. Exadaktylos A, Hautz WE. Emergency Medicine in Switzerland. *ICU Manag Pract*. 2015;15(4):160-162.

16. Gratton MC, Salomone J a, Watson W a. Clinically significant radiograph misinterpretations at an emergency medicine residency program. *Ann Emerg Med*. 1990;19(5):497-502.

17. Brunswick JE, Ilkhanipour K, Seaberg DC, McGill L. Radiographic interpretation in the emergency department. *Am J Emerg Med*. 1996;14(4):346-348. doi:10.1016/S0735-6757(96)90045-5.

18. Nitowski L a, O'Connor RE, Reese CL. The rate of clinically significant plain radiograph misinterpretation by faculty in an emergency medicine residency program. *Acad Emerg Med Off J Soc Acad Emerg Med*. 1996;3(8):782-789.

19. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. 2013;22(Suppl 2):ii21-ii27. doi:10.1136/bmjqs-2012-001615.

20. Peng A, Rohacek M, Ackermann S, et al. The proportion of correct diagnoses is low in emergency patients with nonspecific complaints presenting to the emergency department. *Swiss Med Wkly*. 2015;145(March):1-9. doi:10.4414/smw.2015.14121.

21. Aminzadeh F, Dalziel WB. Older adults in the emergency department: A systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann Emerg Med*. 2002;39(3):238-247. doi:10.1067/mem.2002.121523.

22. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39(1):98-106. doi:10.1111/j.1365-2929.2004.01972.x.

23. Berbaum KS, Franken EA, Dorfman DD, Caldwell RT, Lu CH. Can order of report prevent satisfaction of search in abdominal contrast studies? *Acad Radiol*. 2005;12(1):74-84. doi:10.1016/j.acra.2004.11.007.

24. Ericsson KA, Charness N, Feltovich PJ, Hoffman RR. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press; 2006.

25. Naeger DM, Webb EM, Zimmerman L, Elicker BM. Strategies for Incorporating Radiology into Early Medical School Curricula. *J Am Coll Radiol*. 2014;11(1):74-79. doi:10.1016/j.jacr.2013.07.013.

26. Zwaan L, Kok EM, van der Gip A. Radiology education: a radiology curriculum for all medical students? *Diagn Berl Ger*. 2017;4(3):185-189.

27. Flanders AE, Flanders SJ, Friedman DP, Tartaglino LM, Russell KM. Performance of neuroradiologic examinations by nonradiologists. *Radiology*. 1996;198(3):825-830. doi:10.1148/radiology.198.3.8628878.

28. Zwaan L, Monteiro S, Sherbino J, Ilgen J, Howey B, Norman G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf*. 2017;26(2):104-110. doi:10.1136/bmjqs-2015-005014.

29. Hautz SC, Schuler L, Kämmer JE, et al. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. *BMJ Open*. 2016;6(5):e011585. doi:10.1136/bmjopen-2016-011585.

30. Hautz WE, Kämmer JE, Exadaktylos A, Hautz SC. How thinking about groups is different from groupthink. *Med Educ*. 2017;51(2):229-229. doi:10.1111/medu.13137.

31. Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. *BMJ Qual Saf*. 2013;22(Suppl 2):ii52-ii57. doi:10.1136/bmjqs-2012-001624.

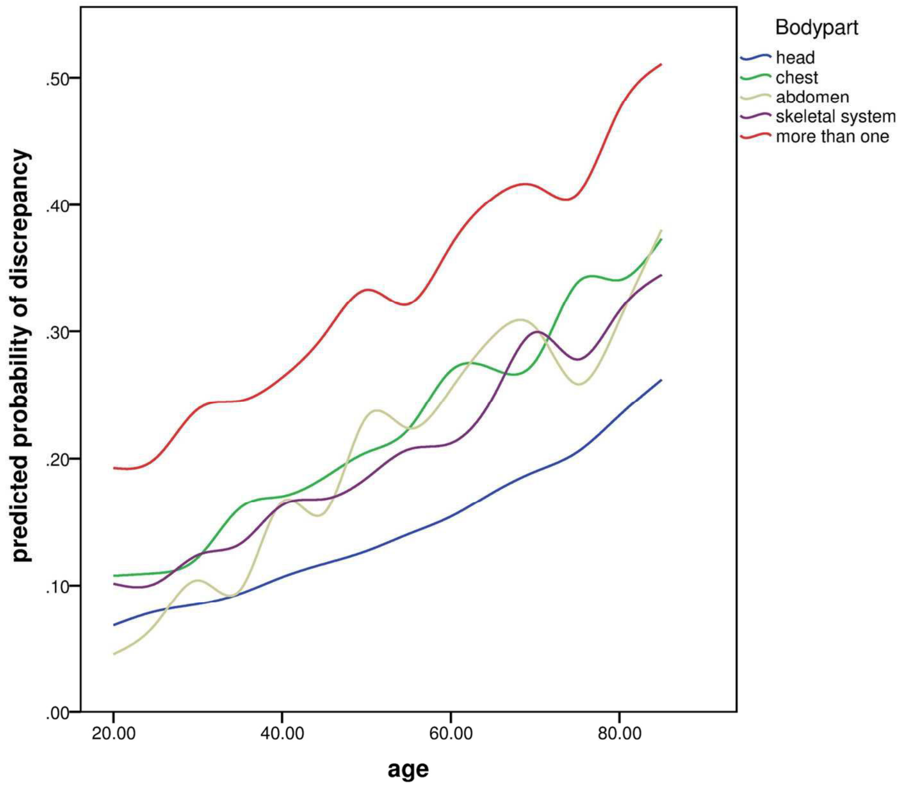

Fig 1. Probability of a discrepancy between first and final radiological diagnosis depending on body part over patients age.

131x105mm (300 x 300 DPI)

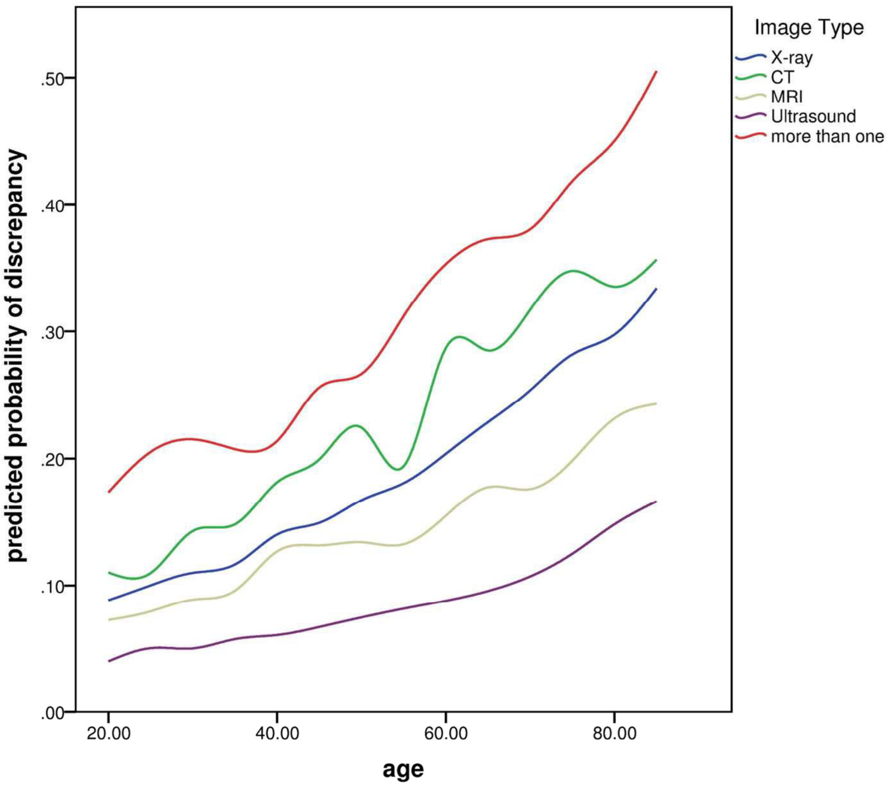

Fig 2. Probability of a discrepancy between first and final radiological diagnosis depending on image modality over patients age.

131x105mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Pages 1 & 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page 4 & 5
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 5 & 6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Pages 5 & 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Pages 5 & 6
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Page 7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Page 6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Page 6
		(*b*) Describe any methods used to examine subgroups and interactions	n/a
		(*c*) Explain how missing data were addressed	Page 6
		(*d*) If applicable, explain how loss to follow-up was addressed	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 7 & Table 1
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	done
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Tables 1 & 2
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	Report numbers of outcome events or summary measures over time	Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted	Tables 2

		estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	& 3
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Page 9, Fig 1 & 2
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Pages 11 & 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Pages 10 & 11
Generalisability	21	Discuss the generalisability (external validity) of the study results	Pages 11 & 12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 3

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.