Published in final edited form as:

J Am Coll Radiol. 2012 May ; 9(5): 344–351. doi:10.1016/j.jacr.2011.12.040.

Satisfaction-of-search for Subtle Skeletal Fractures May Not Be Induced by More Serious Skeletal Injury

Kevin S. Berbaum, PhD, Kevin M. Schartz, PhD, Robert T. Caldwell, MFA, George Y. El-Khoury, MD, Kenjiro Ohashi, MD, Mark Madsen, PhD, and Edmund A. Franken Jr., MD
Department of Radiology, The University of Iowa Carver College of Medicine 200 Hawkins Drive,
Iowa City, IA 52242-1007, USA

Abstract

Purpose—This experiment tested whether radiographs of *major* injuries, those having serious consequences for life and limb, produce a satisfaction of search (SOS) effect on the detection of subtle, non-displaced *test* fractures.

Materials and Methods—Institutional review board approval and informed consent from twenty-four participants were obtained. Seventy simulated patients with multiple trauma injuries were constructed from radiographs of three different anatomic areas demonstrated only skeletal injuries. Readers evaluated each patient under two conditions: first, in the non-SOS condition, no injuries present the first anatomic images; and second, in the SOS condition, the first anatomic images included a *major* injury requiring immediate medical intervention. The SOS effect was measured on detection accuracy using receiver operating characteristic (ROC) analysis for subtle *test* fractures presented on examinations of the second or third anatomic areas.

Results—SOS reduction in ROC area for detecting subtle *test* fractures with the addition of a *major* injury was not observed.

Conclusions—SOS was absent when *major* injuries were presented on radiographs. This finding rejects the hypothesis that SOS arises primarily from injuries requiring major intervention. Similar results have been found previously when *major* injuries were presented on CT, but *test* fractures were presented on radiographs. Our new finding rejects the possibility that SOS is absent because *added* and *test* fractures appear on different imaging modalities.

Keywords

Diagnostic radiology; observer performance; images; interpretation; quality assurance

Introduction

The satisfaction of search (SOS) effect, where one abnormality is missed in the presence of another, contributes to false-negative errors in radiology. An experimental paradigm for studying SOS in multiple organ trauma patients has been used in two receiver operating characteristic (ROC) experiments. Trauma patients often require a series of radiographs to evaluate all potentially damaged organs. In an earlier study (1994), detection of a subtle *test*

© 2012 American College of Radiology. Published by Elsevier Inc. All rights reserved.

Corresponding author: Kevin S. Berbaum, PhD, Telephone: 319-335-8122, Fax: 319-356-1503, kevin-berbaum@uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

fractures was substantially reduced when other injuries were included in the case series (1). In a second study (2007), the SOS effect was again demonstrated in multiple trauma patients using modern digital acquisition and display methods (2). The reduction in performance was about the same in both studies: a reduction in ROC curve area for detecting *test* fractures from 0.86 to 0.81 ($P < .01$). In both studies, the *test* fractures and other injuries were not serious and did not require immediate medical intervention.

It has long been suspected that a cause of diagnostic oversight in the multiply injured patient is the immediate need to treat life threatening injuries (3, 4). Therefore, the hypothesis that injuries having immediate implications for patient care should have stronger SOS effects than less serious injuries was tested in a third ROC experiment (2). This experiment did not find a significant SOS reduction in detecting *test* fractures on radiographs of other body parts, perhaps because a majority of the *major* injuries were shown in most cases with computed tomography (CT), while the *test* fractures were shown on radiographs. The absence of an SOS effect possibly could be explained as an effect that does not extend across imaging modalities.

Using a new group of radiology readers, this experiment tested whether *major* injuries, with serious consequences for life and limb but presented only on radiographs, produce a satisfaction of search (SOS) effect on the detection of subtle, non-displaced *test* fractures.

MATERIALS AND METHODS

In the SOS paradigm that has been used in laboratory experiments, a known abnormality is defined as the *test* abnormality because detection of that abnormality is measured. The *test* abnormality is always presented twice to observers: once alone and once with another abnormality within that same exam. Table 1 presents the SOS paradigm and terminology. SOS occurs when the *test* abnormality is missed in the presence of a distracting abnormality, but not in its absence. Although detection of any abnormality may affect detection of any other abnormality, SOS can only be measured rigorously on the *test* abnormality because it is the only abnormality that is presented by itself as a critical control condition.

In the current experiment, as in previous experiments testing SOS effects in skeletal trauma radiology, radiographic examinations of 3 anatomic areas were presented for each simulated patient. There were two experimental conditions, the first examination displayed was normal for the non-SOS control condition, but displayed a *major* injury in the SOS condition. Addition of injuries into the first examination was an experimental manipulation; we measured detection of the *test* fractures appearing in the second or third examinations, and gathered false-positive responses when both the second or third examinations were normal.

Imaging Material

Digital radiographs were collected from existing records and identified only with an alphanumeric code. Use of these images complied with federal guidelines protecting individual identities and approved by our institutional review board as compliant with Exemption 4 under National Institutes of Health rules. They were converted from Digital Imaging and Communications in Medicine format to Tagged Image File format and optimally resized to fill the display screen.

Simulated Multi-trauma Patients

Seventy cases were constructed to simulate multiple injury trauma patients. Each case depicted a series of examinations using only skeletal radiographs from three different body parts. A case was presented in a specific order so that the *major* injury could only appear as the first examination and the *test* fracture could appear as either the second or third

examination of the series. This display procedure ensured that images with *major* injuries would appear before those with *test* fractures. The radiographs from 304 unidentified patients presenting over 800 normal or abnormal examinations were used in the simulations. Although the examinations simulating each patient came from different sources, they were matched so that they would appear to belong to the same individual. To the extent possible, we used examinations from the same patient. Where this was not possible, we matched examinations by gender and age. The truth status of the cases was rigorously confirmed by two senior skeletal radiologists with 15–37 years of experience. They also confirmed that the examinations in each case appeared to have come from the same individual.

Figures 1, 2 and 3 illustrate a single simulated patient from our experiment (Case 34). Thumbnail images of the examination were presented as a group for each case, but for diagnostic purposes the images were enlarged and projected on a second display. Figure 2 shows the thumbnails in the initial display for Case 34 as it appeared in our two experimental conditions. Readers interpreted each series under two experimental conditions: when the first examination in the series included a *major* injury (Fig 2a), and when it did not (Fig 2b). Figure 3 shows how Case 34 appeared on the two-monitor display as a reader working through the three examinations of the series.

Detection accuracy was measured by scoring responses on the second and third examinations (e.g., radiographs of the pelvis and foot in Fig 2a–b). The second and third body parts presented for each simulated patient were radiographs of extremities, chest, or pelvis and usually included multiple views. The same examinations of the second and third body parts were presented for each patient in both control and experimental conditions and constituted 140 examinations. (These same second and third examinations were used and scored in previous experiments (2).) The type and number of radiographs used in this experiment are characterized in Table 2. There were 27 simulated patients in which both the second and third examinations contained only normal radiographs and 43 with a subtle *test* fracture. The appearance of the *test* fracture was randomized between the second or third examinations, but the same order was used for both experimental conditions. Thus there were 70 scored patients with 27 normal patients and 43 abnormal patients for analysis.

Display

Cases were presented on a workstation consisting of a Dell Precision 360 Mini-tower and two 3-megapixel LCD monitors (National Display Systems, San Jose, CA). Monitors were calibrated to the DICOM standard using manufacturer's specifications. Periodically throughout the 4-month course of the experiment, the calibration was checked and adjusted when necessary.

We used *WorkstationJ* to perform the reading study. It is available for download from <http://perception.radiology.uiowa.edu>. This software was developed from *ImageJ*, a public domain image processing and analysis package written in the Java programming language (available at <http://rsb.info.nih.gov/ij>). The *WorkstationJ* software also collected observer responses including location of abnormality and confidence rating for the abnormality (5).

Each patient's imaging examinations were first presented on the left monitor of a dual-monitor display station (Fig 3a). The reader saw the patient's age and gender, the case number, and thumbnail images of the studies comprising the case. (Fig 3a presents the thumbnails of the examinations shown in Fig 2a). Readers were instructed that these images were not meant to be diagnostic, but to give an impression as to the number and type of examinations available for the patient. The actual diagnostic reading was done on the right monitor (Fig 3b–d). A small menu at the bottom of the right monitor permitted the reader to display the images, one examination at a time, on the right monitor using “next image” and

“previous image” buttons. With these buttons, the reader could maximize the first examination on the right monitor as shown in Fig 3b, then maximize the second examination on the right monitor as shown in Fig 3c, and then maximize the third examination on the right monitor as shown in Fig 3d. They could use the “previous image” button to backup through the series. The small menu also contained a button that would allow readers to proceed to the next patient. The readers were fully aware that once they moved on to the next patient, they could not go back.

Readers

Twenty-four volunteer observers from our Department of Radiology fellows and residents were recruited as observers. All observers were given and signed an informed consent document that had been approved by our institutional review board for human subjects. None of these observers had participated in the prior experiments (1, 2). The experience of the current readers was matched to those of the previous experiments in terms of years of experience and certification (2). The experiment included nine second-year residents, four third-year residents, seven fourth-year residents, and four fellows totaling 24 readers.

Procedure

Before starting the experiment, each reader was read the instructions and, with a demonstration case, was shown how to display the images, make responses, and advance through the cases. They were told that the purpose of the study was to better understand how radiographic studies are read so that error in interpretation can be reduced or eliminated. Readers were instructed to search for all acute fractures and dislocations and to identify each abnormality by placing the mouse cursor over the abnormality and clicking with the right mouse button. This produced a menu box for rating their confidence that the finding was truly abnormal. The readers were directed to indicate their confidence that a fracture or dislocation was present by using discrete terms such as “definitely a fracture or dislocation,” “probably a fracture or dislocation,” “possibly a fracture or dislocation,” and “probably not a fracture or dislocation, but some suspicion.” These discrete terms were transformed into an ordinal scale where 1 represented no report, 2 represented suspicion, 3 represented possible abnormality, 4 represented probable abnormality, and 5 represented definite abnormality.

Data for each experiment were collected in two sessions separated in time by 2–3 months. Half of the cases presented in each session were from the SOS condition and half from the non-SOS condition. Thus in the course of the two sessions, each case appeared twice, once in the experimental condition and once in the control condition. Within each session, multi-trauma patients appeared in a pseudorandom order so that the occurrence of fractures was unexpected and balanced. Before each trial of an experiment, the reader was always informed of the patient’s age and sex. During the reading sessions, room lights were dimmed to about 5 foot candles of ambient illumination.

RESULTS

The multireader multicase (MRMC) ROC methodology developed by Dorfman, Berbaum, and Metz (DBM) (6, 7) has recently been extended in software (DBM MRMC 2.3, available from <http://perception.radiology.uiowa.edu>). Our primary method of analysis of the detection of *test* fractures used ROC analysis fitting the discrete rating data with the proper contaminated binormal model (CBM) (8), treating area under the ROC curve as the measure of detection accuracy. (CBM is a proper ROC model giving well-behaved ROC curves. It does not fail when false-positive rates are very low, as is often the case in skeletal radiology.) Because SOS affects readers rather than patients, generalization to the population of readers is fundamental. Therefore, patients were treated as a fixed factor and

readers as a random factor. These analytic choices mirrored those of the previous study (2). The DBM procedure did not demonstrate a reduction in ROC area for detecting subtle *test* fractures when a major injury was added to the first radiograph of the series (ROC area = 0.82 without *major* injury versus 0.81 with *major* injury, difference = 0.01, $F(1,23) = 2.58$, $P(\text{two-tailed}) = .12$).

We studied the possibility that decision thresholds might shift in the presence of the major added injury. Even with no difference in area under the ROC curve, sensitivity and specificity can vary inversely indicating shifts in decision thresholds. We analyzed the sensitivities at each ROC point using an analysis of variance with within-subject factors for threshold (established by grouping the ratings as 1/2345, 12/345, 123/45, and 1234/5) and experimental treatment (added injury absent vs. present). There was no significant difference for treatment and no significant treatment-by-threshold interaction. A similar analysis performed on specificities yielded similar results. This rules out the possibility of decision threshold shifts.

An additional analysis was performed to compare the detectability of the *major* injuries appearing in radiographs in the current experiment with the detectability of *minor* injuries appearing in radiography in the previous experiment (2). Thus, discrete ratings of the first anatomical region examined in each patient's series were treated as normal when an injury was absent and as abnormal when an injury was present. Therefore, for each reader in each experiment, there were 70 cases without injuries and 70 with injuries within the first examination in the patient's series. Accuracy parameters were estimated by fitting the contaminated binormal model to the rating data of individual readers in each treatment condition (experimental and control). The ROC areas for detecting injuries in the first examinations were compared using nonparametric Mann-Whitney rank-sum tests (9, 10). This test demonstrated a statistically significant difference in detecting *added* injuries presented on radiographs based on whether the fractures were associated with the need for *minor* or *major* intervention (ROC area = 0.97 vs. 0.88, $P < .0001$). Thus, the less serious injuries were more detectable than the more serious injuries on radiographs.

Another Mann-Whitney rank sum test showed that *major* injuries presented on radiography in the current experiment were reported less frequently than *minor* injuries presented on radiography in the previous experiment (58.1 versus 69.6, $P < .001$). These findings confirm that our *major* injuries were harder to detect than *minor* injuries.

Because major injuries were not as detectable as the less serious injuries in the previous experiment (2), we tested whether an SOS effect could be found when ROC analysis was limited to just those reader-case combinations in which the major added fractures were detected. Because different patients drop out of each reader's rating data, MRMC analysis could not be used. We tested the difference between the pairs of areas under CBM ROC curves with a Mann-Whitney rank-sum test (9, 10). We again found minimal SOS (ROC area = 0.81 vs. 0.80, $P(\text{one-tailed}) = 0.08$).

DISCUSSION

It has been supposed that the urgency of detecting and reporting life altering injuries leads to the neglect of additional injuries (3, 4, 11). We hypothesized that orthopedic injuries of greater clinical importance might induce greater SOS for subsequent detection of fractures of less severity. We tested this hypothesis with a carefully controlled set of simulated multi-trauma patients in which both the *major* injury and *test* fractures were presented on radiographs. Two independent studies presenting *added* and *test* fractures on radiographs have documented SOS reduction in *test* fracture detection where few of the distracting

injuries required major intervention (1, 2). In the current experiment, SOS was not induced by *major* injuries on radiographs so the hypothesis must be rejected. The result of this experiment is consistent with that of a previous experiment in which *major* injuries did not produce an SOS effect for other fractures (2). In that study, the major orthopedic injuries were presented primarily on CT; no SOS effect was noted. The possible explanation of that experiment – that there was an absence of SOS because SOS effects may not extend from one imaging modality to another – cannot apply in the current experiment because all *major* injuries were presented on digital radiographs.

In two experiments (1, 2) we have been able to demonstrate a statistically significant 5–6% ROC area difference with 10–12 readers per study. In the current experiment, there were more readers than both of the previous two studies combined. We did observe a very small difference of 1% AUC marginally significant in a one-tailed test. Although there are various approaches to statistical comparison of detection accuracy, when we note that the observed difference in this study is only 1% of total ROC area, it becomes clear that the magnitude of the effect is much less than observed with minor injuries (5% total area). While adding more readers would likely increase statistical significance, it would not likely increase the magnitude of the effect.

A possible reason contributing to the absence of SOS in the current experiment could be that the *major* injuries were not as detectible as the less serious injuries in the previous experiment. Reduced detection of the major (*added*) injuries might explain reduced SOS. Experience ranged from second-year in residency to fellowship training in both experiments (2). Perhaps readers with less experience than skeletal radiologists had seen fewer examples of major injuries than the minor ones and fractures of the cervical spine are difficult to see on radiographs. On the other hand, on average readers did detect 58 of the 70 major fractures, providing a substantial portion of examinations on which SOS could occur. Moreover, we tested this explanation directly by limiting ROC analysis to only those reader-case combinations in which the major fractures were detected. Minimal SOS was found; thus reduced detection of added fractures ultimately fails to explain the absence of SOS.

Scientific evidence does not generally support truncated search as a cause of SOS error in skeletal radiology (1, 2, 12). However, in all of the previous experiments and in the current one, each simulated patient examination series has been limited to a few body parts. In the real world, some trauma patients present with multiple serious injuries and many more anatomical areas may be studied with extensive radiography and advanced imaging. An SOS effect may occur in such a situation – but we are unaware of any published report of positive or negative results about this. For this reason, the SOS laboratory results may not apply to patients with very extensive radiographic examination.

For working skeletal radiologists, the current result may seem inconsistent with their experience. The experimental protocol necessarily differs from the conditions they face in the reading room. Our readers knew they were in an experiment; there were no interruptions and no time pressure. But a problem with comparing SOS effect in the laboratory with individual SOS errors in the clinic is that ascribing a cause for the latter is mostly guesswork. In general, attribution of radiologist errors to SOS cannot be verified: detection of the missed abnormality in the absence of the other abnormality is never studied. It is only certain that one abnormality was missed and another was found. Furthermore, for any particular miss in the reading room, there will be several uncontrolled factors. In the scientific laboratory, we control all factors; in the clinic, this is not possible. Moreover, no appeal to additional factors to explain the absence of SOS from *major* fractures can explain how other experiments demonstrated SOS effects with *minor* distracting fractures (1, 2).

Some readers may wonder whether contemporary residents and fellows might be better trained to avoid satisfaction of search, accounting for the lack of SOS effect in this experiment. Contemporary residents and fellows may be better trained to avoid SOS, but as recently as 2007, they were susceptible to SOS to about the same extent that radiologists were in 1994 (1, 2). In studies of misses revealed through quality assurance over-reading at our institution, many involve the presence of another abnormality that was reported (EA Franken, personal communication). Given that the study involved residents from one institution, it is theoretically possible though unlikely that training methodology affected the results.

A problem for scientific study of radiology errors is that each of the additional factors that could contribute to SOS in the clinic deserves its own laboratory experiment. How might fatigue contribute to missing further abnormalities once the first has been found? What about the interruption of inspection that occurs in reading rooms? There are more possible contributing factors than can easily be handled experimentally.

Another problem for generalization is that in our experiment it was quite easy to report abnormalities: just a mouse click on the fracture location and another on a radio button to report confidence. Part of the effect of severe injuries is the necessity to contact the patient's caregiver. Perhaps further experiments can construct some method to "rush" the observer to better simulate the clinic where (1) limited additional time is available for interpretation once a major injury is found and (2) where finding all abnormalities is secondary to treatment concerns. Perhaps requiring observers to report all major injuries immediately to caregivers would better represent the impact in the clinic.

A benefit of preparing our experiment for display in WorkstationJ software is that once the first experiment is created, it is essentially "packaged" for transmission to other sites where a new sample of radiology readers could be recruited. Questions about the effects of fatigue, interruption, requirement for phoned reports, and level of experience could be addressed without the time, effort and cost of developing the current experiment. We invite those interested in such collaborations to contact us.

Clinical Information

We provided limited clinical history: just the age and sex of the patients. We instructed readers that the experiment involved trauma radiology and asked them to look for fractures and dislocations. Location and type of pain is an effective cue when the detection task is already specific as it is in trauma radiology (13, 14). Moreover, accurate and specific clinical information may prevent SOS (15). Our minimal clinical information may actually approximate what radiologists cope with in the reading room: absent, partial, or incorrect clinical information.

A fundamental question for radiology is whether clinical history improves perception because requests for accurate and complete history rest on whether the information increases accuracy in image interpretation. In a previous paper, we provided a brief commentary and summary of papers by Swensson *et al.* and by Berbaum *et al.* on this question (16). The conclusion was that searching radiographs to explain clinical findings does improve performance, but searching based on the image interpretations of other observers does not. Given the rapid development of electronic medical records, we might wonder whether communication between the clinician and radiologist sufficiently improved by these records to improve radiologist performance. Although others emphasize the potential of electronic records in preventing diagnostic error, they generally admit that "evidence to support the existence of such a benefit is currently lacking" (17). They state the problem succinctly:

“The problem of having too much information is now surpassing that of having too little, and it will become increasingly difficult to review all the patient information that is electronically available.” – Schiff & Bates, (17), Page 1067

Implications

There are two primary ways to reduce diagnostic error in radiology. One method is to improve imaging technology so that abnormalities are more visible. The vast majority of technical research and clinical endeavor is concentrated in this activity (18, 19). The other approach is to improve the interpretive process itself, the most crucial, yet least understood part of the diagnostic process. For radiologists to avoid error, the source of errors must be known (20).

We expected, as do many radiologists, that the most serious injuries are responsible for the greatest portion of satisfaction of search errors. Our results challenge that conviction. The understanding of SOS that we have achieved is that the clinical significance of a detected abnormality does not determine whether it will suppress the detection of another abnormality. It may be that the relative significance of the abnormalities mediates SOS. While further research is needed, eliminating a conviction that is incorrect makes way for better explanations.

Legal implications of these findings are probably limited. Finding that life-threatening injuries have no special power to suppress the detection of other abnormalities should have no practical legal implication.

CONCLUSIONS

Satisfaction of search reduction in ROC area for detecting subtle *test* fractures with the addition of a major injury was not observed even though previous experiments have observed this reduction with the addition of less serious injuries. This result rejects the hypothesis that more severe detected injuries induce a greater satisfaction of search effect for subsequently viewed fractures.

Acknowledgments

Supported by USPHS Grants R01 EB/CA00145 and R01 EB/CA00863 from the National Cancer Institute, Bethesda, MD.

References

1. Berbaum KS, El-Khoury GY, Franken EA Jr, et al. Missed fractures resulting from satisfaction of search effect. *Emerg Radiol.* 1994; 1:242–249.
2. Berbaum KS, El-Khoury GY, Ohashi K, Schartz KM, et al. Satisfaction of search in multi-trauma patients: severity of detected fractures. *Academic Radiology.* 2007; 14:711–722. [PubMed: 17502261]
3. Rogers, LF. *Radiology of skeletal trauma.* New York: Churchill-Livingstone; 1982.
4. Rogers LF. Common oversights in the evaluation of the patient with multiple injuries. *Skel Radiol.* 1984; 12:103–111.
5. Schartz, KM.; Berbaum, KS.; Caldwell, RT.; Madsen, MT. WorkstationJ as ImageJ plugin for medical image studies. Presentation at the R&D Symposium at the Annual Meeting of the Society for Imaging Informatics in Medicine (SIIM); Charlotte, North Carolina. June 4–7, 2009;
6. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol.* 1992; 27:723–731. [PubMed: 1399456]

7. Hillis SL. Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. *Acad Radiol.* 2005; 12:1534–1541. [PubMed: 16321742]
8. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data. II. A formal model. *Acad Radiol.* 2000; 7:427–437. [PubMed: 10845402]
9. Dixon, WJ. BMDP statistical software manual. Vol. 1. Berkeley, CA: University of California Press; 1992. p. 155-174.p. 201-227.p. 521-564.
10. Copyright 1993 by BMDP Statistical Software, Inc. Statistical Solutions Ltd; Cork, Ireland: BMDP3D, BMDP7D, and BMDP2V, release: 8.0. (<http://www.statsol.ie>)
11. Rogers LF, Hendrix RW. Evaluating the multiply injured patient radiologically. *Orthopedic Clinics of North America.* 1990; 21:437–447. [PubMed: 2367099]
12. Berbaum, KS.; Franken, EA., Jr; Caldwell, RT.; Schartz, KM. Satisfaction of search in traditional radiographic imaging. In: Samei, E.; Krupinski, E.; Jacobson, F., editors. *The Handbook of Medical Image Perception and Techniques.* Cambridge, UK: Cambridge University Press; 2010. p. 107-138.
13. Berbaum KS, El-Khoury G, Franken EA, et al. The impact of clinical history on the detection of fractures. *Radiology.* 1988; 168:507–511. [PubMed: 3393672]
14. Berbaum KS, Franken EA, El-Khoury GY. Impact of clinical history on radiographic detection of fractures: A comparison of radiologists and orthopedists. *AJR.* 1989; 153:1221–1224. [PubMed: 2816635]
15. Berbaum KS, Franken EA Jr, Anderson KL, et al. The influence of clinical history on visual search with single and multiple abnormalities. *Invest Radiol.* 1993; 28:191–201. [PubMed: 8486483]
16. Berbaum KS, Franken EA Jr. Commentary: Does clinical history affect perception? *Acad Radiol.* 2006; 13:402–403. [PubMed: 16488852]
17. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? *NEJM.* 2010; 362:1066–1069. [PubMed: 20335582]
18. Samei E. Why medical image perception? *J Am Coll Radiol.* 2006; 3:400–401.
19. Beam CA, Krupinski EA, Kundel HL, Sickles EA, Wagner RF. The place of medical image perception in 21st-century health care. *J Am Coll Radiol.* 2006; 3:409–412. [PubMed: 17412095]
20. Alpert HR, Hillman BJ. Quality and variability in diagnostic radiology. *J Am Coll Radiol.* 2004; 1:127–32. [PubMed: 17411540]

Fig 1.

A 57-year-old woman with fractures of the spine and foot. There is a fracture (arrow) at the junction of the vertebral body with the pedicle (**a**). This was presented in Case 34 as an example of a *major* injury. There is a non-displaced intra-articular fracture (arrow) at the base of the proximal phalanx of the 4th toe (**b**). This represents a *test* fracture and was presented twice as Case 34, once with and once without a *major* injury.

Fig 2.

A 57-year-old woman with fractures of the spine and foot. Thumbnail images of Case 34 displayed in two conditions, the *test* fracture with the *major* injury (**a**) and the *test* fracture without the *major* injury (**b**).

Fig 3.

A dual-monitor display of case 34 as it appeared in the experiment. Each case was initially presented with patient information, case number, and thumbnail images comprising the complete study on the left monitor (**a**). Ordered search occurs when the reader presses a next image button. Images were displayed at maximum size, one examination at a time, on the right monitor for diagnostic interpretation (**b–d**). An image with the *major* injury (**as in b**), or a complementary image without an injury is presented first. The *test* fracture study always appears as either the second or third examination. Successively pressing the next image button replaces the previous image and maximizes the next image.

Table 1

The two treatment conditions of the SOS paradigm.

SOS Terminology	Non-SOS Condition		SOS Condition	
Paradigm design	Control		Experimental	
70 patients viewed twice – Once in each condition	43 Patients	27 Patients	43 Patients	27 Patients
1st examination in the patient series – Does it contain a major <i>added</i> injury?	No		Yes	
2 nd and 3 rd examinations of the patient – Does one of the two examinations contain a <i>test</i> fracture?	Yes	No	Yes	No
Parameter measured at each ROC point – Rating thresholds 5/4321, 54/321, 543/21, and 5432/1	TPF	FPF	TPF	FPF
Accuracy parameter – ROC AUCs for each of 24 readers in each condition	Non-SOS AUC		SOS AUC	
SOS effect = Non-SOS AUC – SOS AUC – Assessed with DBM MRMC	Generalization to population of readers (Readers treated as a random effect)			

Note: ROC = Receiver operating characteristic; FPF = False positive fraction = 1 – Specificity; TPF = True positive fraction = Sensitivity; AUC = Area under the ROC curve fitted by the proper contaminated binormal ROC model to the (FPF, TPF) coordinates for each reader-treatment combination; DBM = Dorfman, Berbaum, Metz methodology; MRMC = Multi-Reader, Multi-Case Methodology. The terminology associated with the SOS experimental paradigm relates chiefly the distinction between test fractures and added fractures. We measure the former and manipulate the latter. The manipulation creates the experimental conditions: no added abnormality (non-SOS condition), and added abnormality (SOS condition).

Table 2Type and number of *major* injuries, *test* fractures, and normal examinations

<i>Major injuries</i> Type (number)	<i>Test fractures</i> Anatomic area (number)	<i>Normal examinations</i> Anatomic area (number)
Cervical spine injuries (50)	Ankle (4)	Ankle (13)
Burst fractures (3)	Chest (0)	Chest (14)
Dens fractures (13)	Clavicle (1)	Clavicle (1)
Dislocations (3)	Elbow (1)	Elbow (5)
Hangman's fractures (5)	Foot (15)	Foot (7)
Posterior vertebral arch fractures (10)	Forearm (3)	Forearm (0)
Spinous process fractures (7)	Hand/finger (8)	Hand/finger (7)
Tear drop fractures (3)	Humerus (0)	Humerus (3)
Vertebral body fractures (6)	Knee (3)	Knee (20)
Non-spinal injuries (20)		
Humeral shaft fracture(1)	Pelvis (0)	Pelvis (19)
Metacarpal fracture(1)	Shoulder (3)	Shoulder (2)
Pelvic fractures (13)	Tibia/fibula (2)	Tibia/fibula (3)
Tibia/fibula fractures(5)	Wrist (3)	Wrist (2)