Published in final edited form as:

J Am Coll Radiol. 2016 August ; 13(8): 973–978.e4. doi:10.1016/j.jacr.2016.04.014.

Trauma in Computed Tomography: The Role of Severe Injury on Satisfaction of Search Revised

Kevin M. Schartz, PhD¹, Mark T. Madsen, PhD¹, John Kim, MD², Riko Ohashi³, Kenjiro Ohashi, MD¹, George Y. El-Khoury, MD¹, Robert T. Caldwell, MFA¹, Edmund A. Franken Jr., MD¹, and Kevin S. Berbaum, PhD¹

¹Department of Radiology, The University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City, IA 52242

²Section of Neuroradiology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115

³Northwestern University, Evanston, IL 60201

Abstract

Purpose—The satisfaction of search (SOS) effect occurs when an abnormality on an image is missed because another is found. This experiment tested whether severe distracting fractures control the magnitude of SOS on other fractures when both appear in a single computed tomograph (CT).

Materials and Methods—Our Institutional Review Board (IRB) approved the study. The experimental (SOS) condition included 35 cervical spine CT cases, all of which contained severe cervical spine injuries. For each of these cases, a similar case was found that had no injuries. We developed image modification software to add simulated fractures to each pair of cases, with and without a major injury. Sixteen different minor fractures were added to 16 of the 35 pairs of images. The thirty-five cases without native injuries constituted a control (non-SOS) condition mixed in a random order. Twenty radiologists read 35 mixed cases in each of two sessions. False-positive evaluations were only collected for cases without simulated fractures.

Results—An SOS effect on detection of simulated fractures was not observed. There was a non-significant ($P = 0.07$) finding of poorer detection in the presence of cases with severe injuries. However, the magnitude of the effect was no greater than has been observed for less severe distracting injuries.

Conclusions—The outcome agrees with the results of two previous experiments that failed to yield an SOS effect associated with detecting severe injuries, suggesting that the severity of a distracting injury does not determine whether a second injury is discovered.

One Sentence Summary

Corresponding author: Kevin M. Schartz, PhD, Telephone: 319-335-8489, Fax: 319-356-2220, kevin-schartz@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.

One of the standard explanations for satisfaction of search errors – that finding severe injuries leads to subsequent errors – was not supported.

Keywords

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

Introduction

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 [1, 2]. The satisfaction of search (SOS) effect, where one abnormality is overlooked in the presence of another, could be the mechanism whereby serious injuries interfere with finding other fractures. A 2012 article [3] in the *Journal* provided experimental evidence on whether the severe injuries are the root cause of SOS on other fractures. The experiment tested whether radiographs showing a variety of major injuries, those having serious consequences for life and limb, produce a SOS effect on the detection of subtle, non-displaced test fractures. Including a major injury among a patient's images did not reduce detection of subtle test fractures.

Cervical spine fractures are among the most clinically significant skeletal injuries. Computed tomography (CT) examination is often used in the case of trauma to the neck rather than radiography because of better sensitivity and specificity [4, 5], but with many more images to evaluate than a radiograph, the chance of information overload may be greater [6]. The current study tested for SOS effects of major fractures on detection of other fractures present within the same CT examinations. It controlled the presence of fractures within CT examinations using new special-purpose image processing techniques.

MATERIALS AND METHODS

The SOS paradigm in laboratory experiments uses abnormalities defined as the *test* abnormalities because detection of those abnormalities is measured. Test abnormalities are always presented twice to observers: once alone and once with another abnormality within that same imaging examination. SOS occurs when the test abnormality is missed in the presence of a distracting abnormality, but not in its absence. The experimenter controls the presence of the *added* distracting abnormality.

Imaging Material

CT examinations 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. We converted images in Digital Imaging and Communications in Medicine (DICOM) format to a file format that removed tags that identified patients and their physicians and retained only the gender, age, and diagnosis.

Cases consisted of CT studies of the cervical spine, half with serious injuries to the cervical spine and half with no apparent spinal injuries. Thirty-five CT examinations showing high morbidity fractures and 35 CT examinations showing no fractures were paired to create 35 dyads, matched in age and general physical characteristics. The 35 major fractures and/or dislocations included: burst (1), dens (2), facet (6), facet dislocation (2), pedicle (5), lamina (2), lateral mass (2), facet-lamina, rib (3), spinous process (3), bilateral arch (1), vertebral body-lamina (2), Jefferson (2), superior facet (2), and burst-lamina (1). Two senior skeletal radiologists with 15–37 years of experience rigorously confirmed the truth status of the cases. In the SOS condition, CT examinations presented major fractures of the cervical spine; in the non-SOS condition, CT examinations of the same anatomy presented no major fractures.

Sixteen dyads had test fractures inserted into both examinations of the pair. For each CT examination of a dyad, the same test fracture was simulated. The 16 simulated test fractures included: transverse, lamina, rib (4), transverse process, clavicle (2), spinous process (3), anterior arch (2), scapula, and transverse foramen. These examinations were used to collect true-positive (TP) reports for ROC analysis. Simulated test fractures were made to be subtle, yet recognizable and placed to best replicate images that may appear on a CT in a reading room. They did not overlap CT slices containing the native injuries.

The other nineteen dyads were used to gather false-positive (FP) reports for ROC analysis. These FP reports were only collected from regions within the matched pairs of examinations without an abnormality on either examination. Thus, even for the examination that lacked a serious native injury, the corresponding region on the other examination was ignored. The size of the region from which FP reports were collected was the same for the two examinations.

Fracture Simulation Methodology

Our experiment required simulation of subtle test fractures on spine CT, which represented a technical challenge. Placing the test fracture in the same location of two different patients required expert guidance and some artistic flair. A pair required matching patient age, general physical characteristics, and number of slices in the study. Additionally, both studies needed to share the same size, shape, intensity, and contrast.

A medical physicist (and a co-author) developed software to place simulated fractures [7] in nearly identical locations of two different images. The software allowed the operator to draw a fracture region on the bone image using the mouse (see online Appendix for a detailed description). Bony structures limited the location of simulated test fractures and were deemed realistic by our musculoskeletal radiologist authors.

Figure 1 illustrates a typical case pair. It demonstrates slices from a patient with a serious fracture as shown in Figure 1A and indicated by the white rectangle. A matched patient without a fracture (shown in Figure 1C) appears in a similar location as Figure 1A. A simulated rib fracture (the target) appears in nearly identical locations of the paired patients as shown by white rectangles in the images Figure 1B and Figure 1D.

Display

A Dell Precision 360 Mini-tower and two 3-megapixel LCD monitors (National Display Systems, San Jose, CA) displayed the images on monitors 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* display software to perform the reading study. This software is available at <http://perception.radiology.uiowa.edu>. It 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 [8]. We presented only the axial views of both the test and distracting abnormalities.

Readers

Twenty radiologists from our Department of Radiology acted as observers. All observers were given and signed an informed consent document that had been approved by our IRB for use with human subjects. None of these observers had participated in the prior experiments [2, 3]. The experience of the current readers was matched to those of the previous experiments in terms of experience and certification [3]. The experiment included six second-year residents, eight third-year residents, four fourth-year residents, and two fellows totaling 20 readers.

Procedure

Before starting the experiment, each reader was read the instructions and shown how to display the images, make responses, and advance through the cases using a demonstration case. 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 indicated their confidence that a fracture or dislocation was present by using percentage certainty: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100%.

Data 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 case of the 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.

Statistical Analysis

Detection accuracy for test fractures in the two conditions was measured and compared to determine whether high morbidity fractures on spine CTs result in other fractures being missed. Detection accuracy was measured by scoring responses on simulated test fractures.

The multireader multicase (MRMC) ROC methodology developed by Dorfman, Berbaum, and Metz (DBM) [9, 10] has recently been extended in software (OR-DBM MRMC 2.5, available from <http://perception.radiology.uiowa.edu>). Our primary method of analysis of the detection of test fractures used ROC analysis fitting the rating data with the contaminated binormal model (CBM) [11], treating area under the ROC curve as the measure of detection accuracy. (CBM is a proper ROC model producing non-degenerate ROC curves.) We also studied average true-positive fraction (TPF) measured at the false-positive fraction (FPF) of 0.1 (TPF_{FPF=0.1}, alternatively referred to as sensitivity at a fixed specificity of 0.9) because this index of 0.1 focuses on a part of the ROC curve well-supported with empirical ROC points. 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 [3].

RESULTS

OR-DBM MRMC did not demonstrate a reduction in ROC area for detecting subtle test fractures when a major injury was present in the CT examination (ROC area = 0.880 without *major* injury versus 0.855 with *major* injury, difference = 0.025, $F(1,19) = 1.95$, $P(\text{two-tailed}) = 0.179$). The same was true using TPF at FPF = 0.1 as the dependent measure (TPF_{FPF=0.1} = 0.764 without *major* injury versus 0.713 with *major* injury, difference = 0.051, $F(1,19) = 2.40$, $P(\text{two-tailed}) = 0.138$). Because only a reduction in performance in the SOS condition is under consideration to test the hypothesis of an SOS effect, it is appropriate to use the one-tailed P; in this case $P(\text{one-tailed}) = 0.07$. However, the one-tailed test is not statistically significant either.

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 studied decision thresholds using the FP fraction associated with each ROC point, perhaps the best and simplest index of decision threshold [12]. The FP fractions were established by grouping the ratings as 10%|20–100%, 10–20%|30–100%, 10–30%|40–100%, 10–40%|50–100%, 10–50%|60–100%, 10–60%|70–100%, 10–70%|80–100%, 10–80%|90–100%, and 10–90%|100%.) For each reader-treatment combination, the midpoint of the FP range of ROC points was computed as the average of most and least conservative ROC points. As such we get an array of midpoint values consisting of the two SOS conditions $\times$ 20 readers (40 values). Likewise, the width of the FP range of ROC points was computed as least conservative minus the most conservative ROC point. We tested for differences in FP fractions between non-SOS and SOS conditions using non-parametric Wilcoxon tests. There was no statistically significant difference in either the average mid-point of FP points (0.13 vs. 0.13, Wilcoxon $p = 0.98$) or for the range of FP points (0.12 vs. 0.10, Wilcoxon $P = 0.32$).

Median inspection times were computed for each reader for each experimental condition for examinations with and without test fractures. The four median times per reader were analyzed with an ANOVA with within-subject factors for SOS condition and examination type (with or without test abnormality). It took 18 seconds longer to read cases that had a test fracture than those that did not (119 seconds vs. 101 seconds, $F(1,19) = 34.24$, $P < 0.0001$). It took 42 seconds longer to read cases that had a major native injury than those that did not (131 seconds vs. 89 seconds, $F(1,19) = 61.28$, $P < 0.0001$). The interaction of factors was not statistically significant, so the impact of the two factors on time was additive. Therefore, reading time was directly related to how much injury was present.

DISCUSSION

Some thought the urgency of detecting and reporting life altering injuries lead to a neglect of additional injuries [13–16]. In other words, body injuries of greater clinical importance induce greater SOS for subsequent detection of fractures.

Two independent experiments presented less severe injuries and test fractures on radiographs. They documented SOS reduction in test fracture detection [2, 3]. These experiments demonstrated a small but statistically significant ROC area differences with modest reader sample sizes.

Two additional experiments (both described in [3]) studied whether more serious injury would yield more robust SOS effect. Major orthopedic injuries presented on CT failed to produce an SOS effect for other fractures presented on radiographs [3]. In the first experiment distracting and test fractures appeared in different modalities, CT and radiography, respectively. An absence of SOS might reflect a limit on how far SOS effects extend, for example from one imaging modality to another. Perhaps SOS from major injuries would still yield a stronger SOS effect, so long as they were presented within the same imaging modality.

The second experiment described in [3] tested this hypothesis. That experiment included more readers than the previous two studies and presented minor distractor fractures on radiographs combined. Yet with major injuries on radiographs only a small reduction of 1% in ROC area was observed, which could be considered marginally significant in a one-tailed test but non-significant in a two-tailed test. Although opinions may differ, one point of view is that the SOS reduction was not significant. Regardless of this point, it is clear that the magnitude of the effect is much less than that repeatedly observed with minor injuries. The small magnitude of the effect cannot support the notion that SOS magnitude depends on severity of the distracting injury in radiography.

The current experiment also tested whether major injuries would yield a stronger SOS effect, so long as they were presented within the same imaging modality. This is the first study in which distracting and test fractures were presented within a single CT examination. With 20 radiology readers, there was a 3% ROC area reduction in the SOS condition which was not statistically significant. This finding does not support the severity hypothesis. The 3% reduction is smaller than the 5% and 6% reductions observed with minor distracting injuries

[13, 15, 16]. This smaller reduction of ROC area with major injuries than with minor injuries is further evidence of the failure of the severity hypothesis.

Implications

We had expected, as do many radiologists, that the most serious injuries are responsible for the greatest portion of satisfaction of search errors. Our current results together with those of a prior study [3] fail to support this hypothesis. The clinical significance of a detected abnormality does not seem to determine whether it will suppress the detection of another abnormality. The current experiment extends that conclusion beyond radiography to computed tomography.

CONCLUSIONS

We observed no significant SOS reduction in detecting subtle test fractures in the presence of serious injury in computed tomography examination of the cervical spine. Our results do not support the hypothesis that more severe detected injuries induce a greater satisfaction of search effect for subsequently viewed fractures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by USPHS Grant R01 EB 00145 from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), Bethesda, Maryland.

REFERENCES

1. Rogers, LF. Radiology of skeletal trauma. New York: Churchill-Livingstone; 1982.
2. Rogers LF. Common oversights in the evaluation of the patient with multiple injuries. *Skel Radiol*. 1984; 12:103–111.
3. Berbaum KS, Schartz KM, Caldwell RT, et al. Satisfaction-of-search for subtle skeletal fractures may not be induced by more serious injury. *J Am Coll Radiol*. 2012; 9:344–351. PMID: PMC 3345121, NIHMSID: NIHMS347566. [PubMed: 22554633]
4. Cain G, Shepherdson J, Elliot V, Svensson J, Brennan P. Imaging suspected cervical spine injury: plain radiography or computed tomography? *Radiography*. 2010; 16:68–77.
5. Nuñez DB, Zuluaga A, Fuentes-Bernardo DA, Rivas LA, Becerra JL. Cervical spine trauma: how much more do we learn by routinely using helical CT? *Radiographics*. 1996; 16:1307–1318. [PubMed: 8946537]
6. Andriole KP, Wolfe JM, Khorasani R, Treves ST, Getty DJ, Jacobson FL, Steigner ML, Pan JJ, Sitek A, Seltzer SE. Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section CT scan can ruin your whole day. *Radiology*. 2011; 259:346–362. [PubMed: 21502391]
7. Madsen MT, Berbaum KS, Ellingson AN, et al. A new software tool for removing, storing, and adding abnormalities to medical images for perception research studies. *Acad Radiol*. 2006; 13:305–312. [PubMed: 16488842]
8. Schartz KM, Berbaum KS, Caldwell RT, Madsen MT. WorkstationJ: Workstation emulation software for medical image perception and technology evaluation research. *Proceedings of the International Society for Optical Engineering*. 2007; 6515:65151I–65151I1.

9. 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]
10. 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]
11. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data. II. A formal model. *Acad Radiol.* 2000; 7:427–437. [PubMed: 10845402]
12. Swets, JA.; Pickett, RM. *Evaluation of diagnostic systems: Methods from Signal Detection Theory.* New York: Academic Press; 1982.
13. Berbaum KS, El-Khoury GY, Ohashi K, et al. Satisfaction of search in multi-trauma patients: severity of detected fractures. *Acad Radiol.* 2007; 14:711–722. [PubMed: 17502261]
14. Rogers LF, Hendrix RW. Evaluating the multiply injured patient radiologically. *Orthopedic Clinics of North America.* 1990; 21:437–447. [PubMed: 2367099]
15. 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.
16. Berbaum KS, El-Khoury GY, Franken EA Jr, et al. Missed fractures resulting from satisfaction of search effect. *Emerg Radiol.* 1994; 1:242–249.

Figure 1.

A and B images are from a patient with a serious fracture as shown in the upper left slice and indicated by the white rectangle. The C and D images are from a matched patient without a fracture (as shown in C at a similar location as image A). The same simulated fracture was added to a rib of the paired patients shown in the images B and D indicated by white rectangles.