PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Satisfaction of search (SOS) error and new lesions identification on imaging in central review for clinical trials

Sharma, Manish, Kota, Sree Sudha, Madasu, Madhuri, Bajpai, Surabhi, Shao, Yibin, et al.

Manish Sharma M.D., Sree Sudha Kota, Madhuri Madasu, Surabhi Bajpai, Yibin Shao, Srinivas Pasupuleti, Michael O'Connor, "Satisfaction of search (SOS) error and new lesions identification on imaging in central review for clinical trials," Proc. SPIE 12035, Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment, 1203509 (4 April 2022); doi: 10.1117/12.2612995

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

"Satisfaction of search" (SOS) error and new lesions identification on imaging in central review for clinical trials

Manish Sharma[1], Sree Sudha Kota[1], Madhuri Madasu[1], Surabhi Bajpai[2], Yibin Shao[3], Srinivas Pasupuleti[1], Michael O'Connor[4]
[1]Calyx.ai, Hyderabad, India, [2]Calyx.ai, Billerica, Boston, MA, US, [3]Calyx.ai, Shanghai, China, [4]University of Massachusetts, Lowell, MA, US

ABSTRACT

Purpose: Medical errors account for a third leading cause of death in the United States. Despite recent interventions, high error rates persist. Satisfaction of search (SOS) is a relatively less harmful type of bias that indicates an individual's decreased vigilance and/or awareness of additional abnormalities after the first abnormality has been identified. We studied SOS data in clinical practice and tried to correlate its best fit in clinical trial setting reads where diagnosis is typically already known.
Methods: SOS data from four different clinical trials including 8036 timepoints with assessments across 1655 subjects were reviewed by board-certified radiologist reviewers using Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria and analyzed for new lesion identification.
Results: We analyzed specific subset of subjects with progressive disease which usually is the critical clinical trial endpoint in oncology. We noticed that once progressive disease was detected by the radiologist reviewer, additional new lesions tend to be not marked or missed out on a statistically significant proportion. This might be not due to the incompetence of the reviewer but due to SOS error where satisfaction was reached on finding progressive disease, the trial endpoint analogous with the first abnormality in clinical practice.
Conclusions: With SOS, once an abnormality is detected and recognized, it requires additional attention to look for other possible abnormalities within an image. Additional abnormalities may be missed by the radiologist once the first abnormality is found. Several strategies can be used to mitigate SOS which includes the use of a systematic approach to ensure all relevant findings are identified, through a secondary search once the first finding is reported.

1. INTRODUCTION

Medical errors are one of the significant causes of morbidity and mortality. It is estimated that medical errors range from 44,000 to 400,000 annually. Such higher estimates rank medical errors as the third leading cause of death in the United States. Medical errors can be categorized and further subdivided into diagnostic errors, treatment errors, preventive errors, and other errors (e.g., communication failures, equipment failures, and other system failures). In radiology, diagnostic errors or other errors are the most common errors. In 1959, radiologists became aware of the high rates of diagnostic errors through an article by Garland who reported a 32% retrospective error rate in the interpretation of abnormal chest radiographs and an average daily error rate of 3%–4% when negative studies were included. Despite several interventions taken place, the high error rates persist. Thirty percent of these retrospective errors represented misses of positive findings, and 2% represented misinterpretations, specifically false-positive diagnoses of diseases that had not occurred.

Decision making involves type 1 and type 2 thinking. Heuristics also referred as type 1, or fast thinking, depicts the ability of the brain to think and act on intuition. Type 1 thinking requires continuous work as human beings at the conscious level experience less than 1% of the information the senses can process. Experienced radiologists can reach a diagnosis without significant conscious deliberation using a variety of heuristic techniques; however, heuristics may fail due to inherent errors called biases. Compared to type 1 thinking, type 2 thinking, is an analytical, slow, thoughtful, and effortful approach to decision making. In difficult situations, type 2 thinking dominates, and it helps the radiologist to reach a correct diagnosis in comparison to type 1 thinking. Complex decision making requires a combination of type 1 and type 2 thinking. A radiologist must recognize his or her susceptibility to cognitive biases, which can result from type 1 thinking [1].

Satisfaction of search (SOS) is a type of underreading error also referred as false-negative response which occurs when the lesion is undetected after detection of first lesion [2]. SOS describes a situation in which detection of one radiographic abnormality interferes with others. Ashman et al. studied whether SOS occurred in the interpretation of conventional

Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment, edited by Claudia R. Mello-Thoms, Sian Taylor-Phillips, Proc. of SPIE Vol. 12035, 1203509
· © 2022 SPIE · 1605-7422 · doi: 10.1117/12.2612995

radiographs of the musculoskeletal system. There was no statistically significant difference in the detection rate of one finding when comparing cases with one finding with multiple finding. Detection rate of second and third abnormalities in the multiple-finding cases was significantly lower [3]. It is a type of bias that indicates an individual's decreased vigilance and/or awareness of additional abnormalities after the first abnormality has been identified. Of all types of radiologic diagnostic errors in a study, 22% were related to SOS, which was second only to errors classified as underdiagnoses or misses, making this the most common cognitive bias in diagnostic radiology.

Radiologists may be subjected to answer a clinical question asked, but their actual role is to make a unifying diagnosis. An example related to it can be the call for a peripherally inserted central catheter line check, which appears to be a straightforward check, but it primarily asks whether the line is appropriately positioned. If only the clinical question is answered, it conceals the search for possible additional findings, such as pulmonary nodules. The introduction of an artificial second finding on a chest radiograph can reduce the sensitivity for detection of the initial abnormality. Even in the absence of introduced bias, lung cancer on chest radiographs is commonly missed [1].
In SOS, once an abnormality is detected and recognized, it requires additional attention to look for other possible abnormalities within an image. If this extra effort is not taken, subsequent lesions in the same image or case can be missed. Estimates of SOS errors vary, ranging from one-fifth to one-third of misses in radiology and possibly as high as 91% in emergency medicine. An in-depth study has found that premature termination of search is generally not the basic cause of SOS; rather, it is faulty pattern recognition and/or faulty decision making as more likely culprits [4].

Samuel et al. concluded from their study that SOS is an important source of error in the detection of subtle abnormalities but not of obvious abnormalities. Obvious abnormalities capture visual attention and decrease vigilance for subtler abnormalities [5].

Krupinski et al. ran a series of studies bringing faculty and residents into the lab before and after a long day of clinical reading (about 8 hours on average) and measured their diagnostic performance (using Receiver Operating Characteristic (ROC) analysis) interpreting bone images with subtle fractures, plain film chest with nodules and Computerized Tomography chest with nodules. There was a statistically significant drop in diagnostic accuracy of about 4 percent in each case. A series of SOS studies were replicated to find whether the combination of fatigue and the well-known SOS source of errors results in additional changes in performance. Early results suggested that being fatigued seems to reduce radiologists' willingness to report multiple lesions on an exam and may have a significant impact on time to report [6].

Berbaum et al. demonstrated that SOS associated with the presentation of a simulated pulmonary nodule on film-based chest radiographs resulted in a statistically significant decrease in the area under the receiver operating characteristic (ROC) curve. Prior studies using film radiographs interpreted on a light box postulated that satisfaction of search was due to decreased accuracy, specifically defined as a change in the height of the ROC curve [2].

Krupinski et al. viewed on a workstation using digital images to suggest that SOS is secondary to a shift in the decision threshold, reflected by a change in both true-positive fractions and false-positive fractions. SOS produced a threshold shift in the ROC curves for both faculty and resident readers, with a decrease in median true and false-positive fractions. The authors interpreted it as a decreased willingness to identify an abnormality. Interpretation times were not significantly different between the SOS and non-SOS conditions; a native abnormality was observed to require more interpretation time. A multitiered experimental design with independent assessments of radiologist performance in the fatigue condition as well as the satisfaction of search condition, and a subsequent combined assessment can provide more insight. The ROC analysis will be able to more effectively analyze the combined effects of fatigue and satisfaction of search. Fatigue has previously been hypothesized to be a decrease in accuracy, whereas SOS appears to manifest as a reluctance to report additional abnormalities. The interaction between these two mechanisms remains incompletely understood, although the authors suggest an additive effect [7].

Several strategies can be used against SOS which includes using a systematic approach to ensure all relevant findings are identified, particularly common, and commonly missed diagnoses. Once the primary search is completed and the first finding is reported or response to the clinical question is complete, there should be an initiation for a secondary search. The radiologist should adhere to primary and secondary search patterns [1].

The Response Evaluation Criteria in Solid Tumors (RECIST) guideline proposed a method for measurement of antitumor activity of cytotoxic drugs based on tumor shrinkage. The RECIST guideline for treatment response assessment relies on total tumor size, including both enhancing (viable) and nonenhancing (necrotic) components of the tumor and based on the longest unidimensional measurement of the lesion. In 2009, RECIST 1.1 guidelines came as a revision of RECIST 1.0 guidelines for the radiologic imaging-based evaluation of tumor response. In RECIST 1.1 guideline, Disease progression is clarified in several aspects, for measurable disease progression is considered for 20% increase in sum with 5 mm absolute increase. Furthermore, 'unequivocal progression' of non-measurable/non-target disease is made clear which was a source of confusion in the original RECIST guideline. Detection of new lesions, including the interpretation of FDG-PET scan assessment is also considered PD.

For new lesions: The appearance of new malignant lesions denotes disease progression. There are no specific criteria for the identification of new radiographic lesions; however, the finding of a new lesion should be unequivocal: i.e. not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumor [8]. As long as new lesion is identified for confirmation of PD, the number of new lesions is not relevant from SOS error point of view.

2. METHODS

Data Acquisition: A retrospective analysis of four oncology trials including 1655 subjects with 8036 post-baseline timepoints with assessments was reviewed by board-certified radiologist reviewers using Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria and analyzed to evaluate the frequency and significance of SOS error. Data was blinded with respect to study names, study sponsors, treatment arm, subjects and reviewers. Similar study designs were selected to focus on analyzing RECIST 1.1 criteria assessments discordance rates and SOS error was analyzed.

Data Analysis Methods: After obtaining an assessment dataset, the data were aggregated and prepared for analysis using Excel - a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS and Python - a high-level, cross-platform, and open-sourced programming language released under a GPL-compatible license. Python Software Foundation (PSF), a non-profit organization, holds the copyright of Python. Url - https://www.python.org Version 3.9.0. All the plots are worked on Power BI - a Microsoft business analytics service version 2.91.383.0

3. RESULTS

Table 1 includes data from four oncology trials including 1655 subjects with 8036 follow-up timepoints with assessments. The total progressive disease (PD) timepoints were 2384, total non-PD assessments were 5652, total PD timepoints by new lesion were 2021. New lesion PD by both readers (827) was less than new lesion PD by only one reader (1194). As this is retrospective analysis, reader with more new lesions identified was marked as "Gold Standard" readers for benchmark setting for comparison for "missed" new lesions.

Table 1: Comparison of new lesion PD by both readers versus only one reader in 4 studies

STUDY	TOTAL SUBJECTS	TIMEPOINTS WITH ASSESMENTS	TOTAL PD TIMEPOINTS	TOTAL NON PD ASSESMENTS	**TOTAL PD TIMEPOINTS BY NEW LESION**	NEW LESION PD BY BOTH READERS	**PD WITH NEW LESION BY ONLY ONE READER**
I	705	3752	994	2758	**952**	440	**512**
II	213	1227	452	775	**422**	194	**228**
III	388	1728	425	1303	**284**	89	**195**
IV	349	1329	513	816	**363**	104	**259**
TOTAL	**1655**	**8036**	**2384**	**5652**	**2021**	**827**	**1194**

Figure 1: % of tumors identified based on count of new lesions between gold standard and comparator

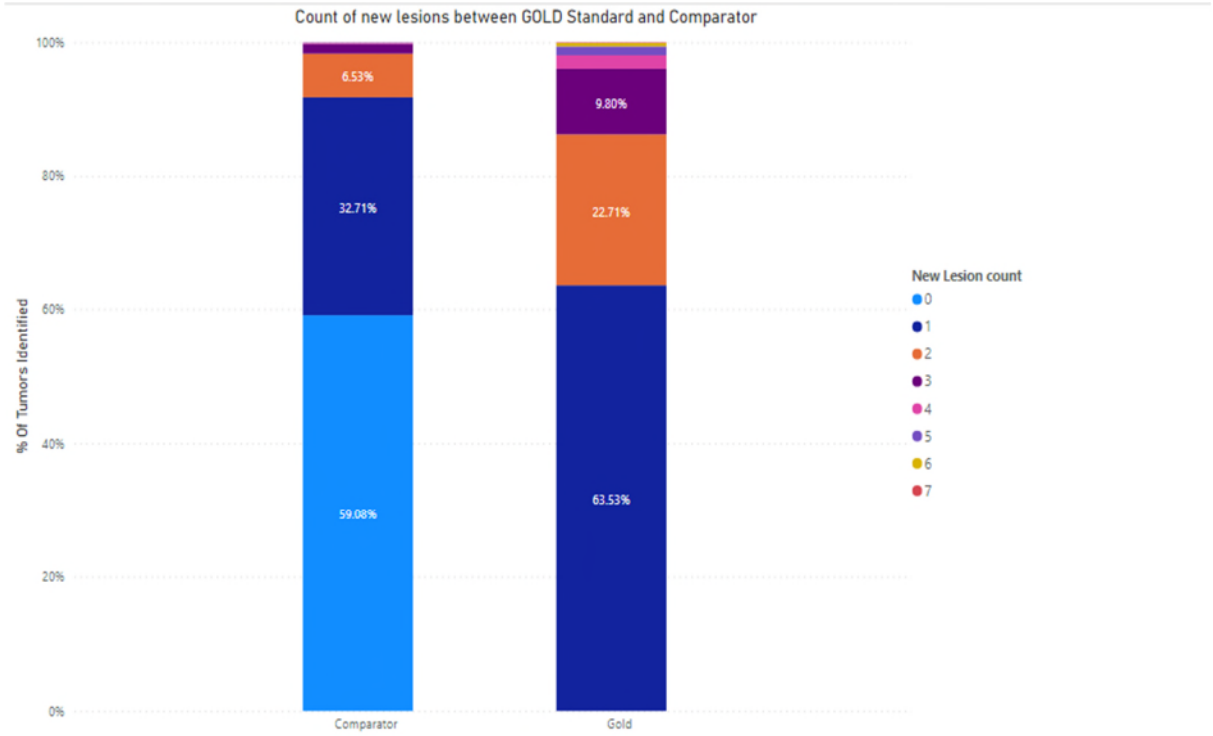

Figure 1 shows % of new lesion counts for gold standard and comparator reader. Gold standard was the reader who detected at more new lesions per visit. The reader being compared against gold standard was referred as comparator. The stacked bar shows that comparator could not find new lesion in 59.08% of tumors identified. Gold standard identified one new lesion in 63.53% of tumors identified as compared to 32.71% for comparator. The Comparator reader missed or ignored to focus on the new lesions once PD was identified either by target, non-target or other new lesions.

Table 2: Comparison of evaluation by comparator and gold standard reader

	PD by New Lesion	PD without New Lesion	TOTAL
Gold Standard Reader	1194	1190	2384
Comparator Reader	827	1557	2384
TOTAL	2021	2747	4768

The total counts by gold standard were much higher than comparator. The average count by gold standard was 1.57 > 0.51 by comparator. Minimum lesion count was 1 for gold standard reader and 0 for comparator. The maximum of lesion count was 7 for gold standard reader and 5 for comparator. There is statistically significant difference between the Gold Standard reader and comparator reader when it comes to identification of new lesion in a PD setting. The chi-square statistic is 115.676. The p-value is < 0.00001. Significant at $p < .05$.

Figure 2: Difference between gold and comparator

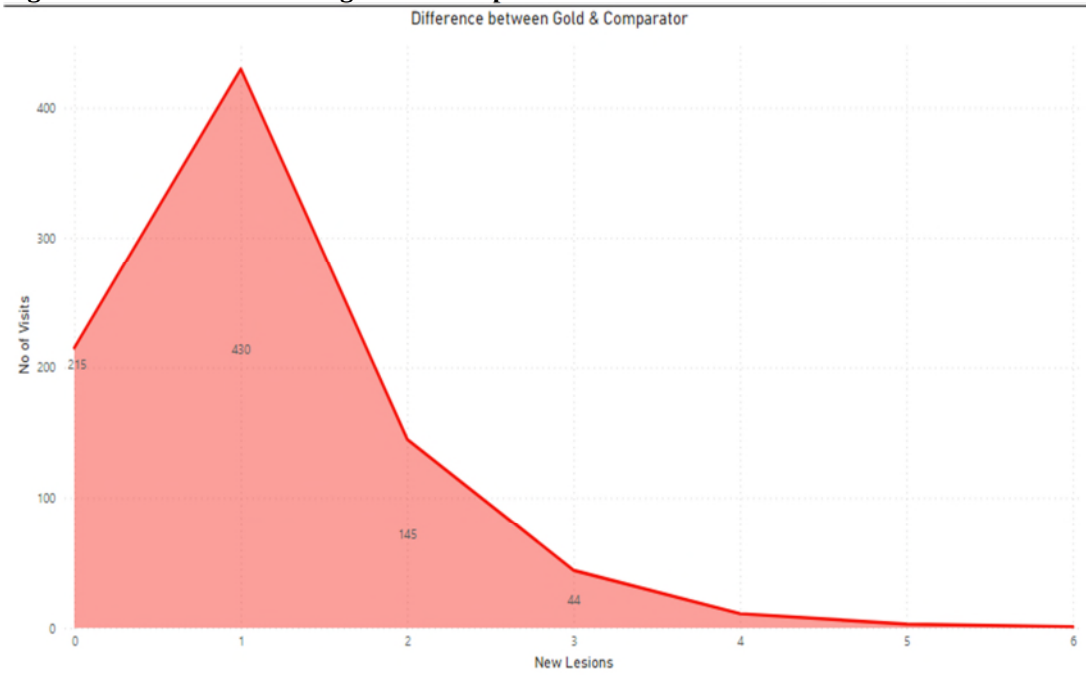

Figure 2 clearly depicts that for PD vs non-PD cases, the scenario of difference in new lesion identification is in fact more than agreement on no new lesion between the two readers. This suggests a bias towards ignoring new lesion in a PD setting as this would typically not have any impact on clinical outcome but is significant from total tumor burden capture perspective.

Figure 3:

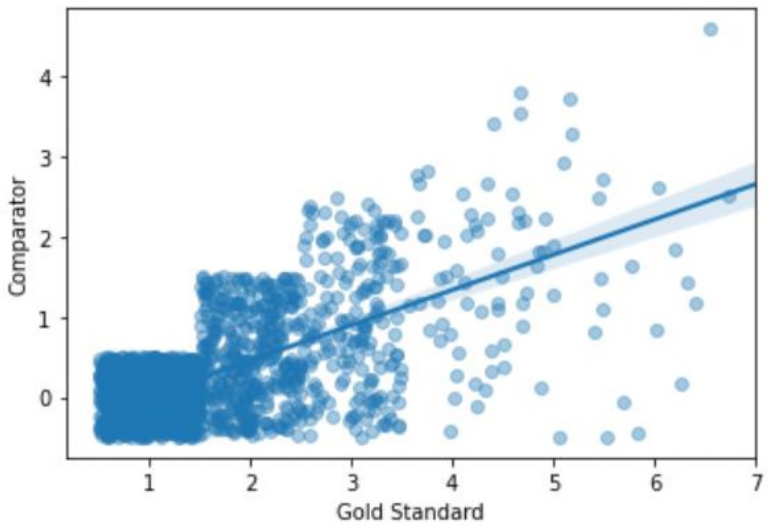

Figure 3 The scatter plot shows that most of the difference in identification of new lesions in PD setting is towards lower numbers like 1 and 2 for Gold Standard reader. As the patient is already PD, such difference tends happen when subtle missed new lesions do not have any impact on clinical outcome for the patients. We strongly believe this is related to SOS error rather than other types of cognitive error by reader in missing new lesions.

4. DISCUSSION & CONCLUSION

Satisfaction of search is a type of cognitive bias which refers to a radiologist's tendency to reduce attentiveness for recognition of additional abnormalities after detection of the first abnormality [1]. This diagnostic error which accounts for 22% of diagnostic errors by a radiologist is often related to reader fatigue and impacts the ability to detect subtle abnormalities [5,6]. While this diagnostic error has important clinical implications with potential adverse effects on patient management in routine practice, it also impacts precise evaluation of patterns of treatment response in oncological clinical trials. In this study, our objective was to determine the occurrence of this error in multi-institutional oncological trials based on independent radiologist data using the standardized response assessment criteria, Response Evaluation Criteria in Solid Tumors 1.1 (RECIST 1.1). We found that there was significant variability among independent readers for recognition of new lesions in patients categorized as progressive disease (PD). In comparison to the reference standard, independent readers missed new lesions in 59% of subjects after PD assessment based on target, non-target or other new lesions.

Several investigators have studied the impact of satisfaction of search as a diagnostic error in clinical practice [2,5,9]. It can have adverse impact on patients with trauma and pulmonary nodules. This diagnostic error could also influence detection of subtle incidental findings, recognition of which has substantial impact on long term patient outcome such as early detection of incidental pancreatic and renal abnormalities in patients undergoing abdominal CT. In oncological patients on various systemic and loco-regional therapies, this error could lead to lack of recognition of early metastases which has critical ramifications in triage of patients. Our study shows that independent readers often ignore or fail to report additional sites of new disease when an assessment of PD is determined. While this does not impact the primary end point in clinical trials, it limits comprehensive evaluation of patterns of metastatic disease and various organ involvement, thereby affecting more detailed understanding of the tumor biology and patterns of response to newer therapies. In the era of precision medicine, this is especially important with the emergence of newer targeted and immunotherapeutic agents which often effect specific tumor cells and identification of entire tumor burden is essential in patients on newer trial criteria such as iRECIST. Structured reporting templates and use of a systematic approach to search pattern have been shown to reduce the occurrence of this diagnostic error. Training and education of radiologist on the occurrence of this error and reinforcing the need for secondary search after addressing the primary clinical or trial question is important. Other strategies which have been used to address this error include reduction of workplace interruptions with emphasis on task-oriented work without multi-tasking. In a clinical trial setting, this error can offset by incorporating their discussion during training and providing information on the common imaging manifestations and combination of disease presentation. For example, in patients with pancreato-biliary malignancies, a discussion on the common sites of metastases such as liver, lymph nodes and peritoneum would allow the reader to evaluate these sites during secondary search.

This study has few limitations. Firstly, this was a retrospective study with multiple readers which can have a selection bias. Secondly, the patient population was heterogeneous with patients with different malignancies on various different therapies. Thirdly, our findings are based on an assumption that SOS could be key contributor to variability in identification of new lesions at PD diagnosis. It is possible that the variability in identification of new lesions could be related to variation in reader experience and competency.

To conclude, satisfaction of search is factor contributing to variability in the diagnosis of new lesions in patients on oncological trials in a progressive disease setting. Multiple strategies are essential to mitigate this error including radiologist training, use of structure templates and a systematic search pattern.

FUTURE WORK

We intend to further explore the interaction between SOS and fatigue and analyze to find if there is any correlation between SOS and fatigue and its relationship to reader performance in blinded independent central read setting.

ACKNOWLEDGMENTS

The authors wish to acknowledge Calyx (https://www.calyx.ai/) for its willingness to allow us to submit this abstract. This data has not been submitted elsewhere for publication or presentation.

REFERENCES

1. Busby L, Courtier FL, Glastonbury CM. Bias in Radiology: The How and Why of Misses and Misinterpretations. RadioGraphics. 2018; 38:236–247. https://doi.org/10.1148/rg.2018170107.
2. Berbaum KS, Franken EA Jr, Dorfman DD, Rooholamini SA, Kathol MH, Barloon TJ, Behlke FM, Sato Y, Lu CH, el-Khoury GY, et al. Satisfaction of search in diagnostic radiology. Invest Radiol. 1990;25(2):133-40. doi: 10.1097/00004424-199002000-00006. PMID: 2312249.
3. Ashman CJ, Yu JS, and Wolfman D. Satisfaction of Search in Osteoradiology. American Journal of Roentgenology 2000;175(2);541-544.
4. Krupinski EA. Current perspectives in medical image perception, Atten Percept Psychophys. 2010; 72(5). doi:10.3758/APP.72.5.1205.
5. Samuel S, Kundel HL, Nodine CF, Toto LC. Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. Radiology. 1995;194(3).https://doi.org/10.1148/radiology.194.3.7862998
6. Krupinski EA. Fatigue in Radiology: What is its Impact and What Can Be Done? https://www.radiologybusiness.com/topics/care-optimization/fatigue-radiology-what-its-impact-and-what-can-be-done?nopaging=1
7. Cook TS. Now You See It, But Would You Later? Examining the Mechanisms of Satisfaction of Search in the Fatigued Radiologist. Acad Radiol 2017; 24:1055–1057.
8. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, et al. New response evaluation in solid tumors: Revised RECIST guideline (version 1.1). Eur J Cancer 2009; 45: 228-247.
9. Schartz KM, Madsen MT, Kim J, Ohashi R, Ohashi K, El-Khoury GY, Caldwell RT, Franken EA Jr, Berbaum KS. Trauma in CT: The Role of Severe Injury on Satisfaction of Search Revised. J Am Coll Radiol. 2016 Aug;13(8):973-978.e4. doi: 10.1016/j.jacr.2016.04.014.