Is a Picture Worth a Thousand Words? The Effect of Viewing Patient Photographs on Radiologist Interpretation of CT Studies

John Ryan, MD, Gul-e Khanda, MBBS, Rebecca Hibbert, MD, Shawna Duigenan, MD, Adam Tunis, MD, Najla Fasih, MBBS, Blair MacDonald, MD, Mohammed El-Khoudary, MBBS, Ania Kielar, MD, Matthew McInnes, MD, Vivek Virmani, MD, Nitin Ramamurthy, MBBS, Nick Kolenko, MD, Adnan Sheikh, MD

DESCRIPTION OF THE PROBLEM

The past decade has seen sweeping changes to radiologic practice. Year-on-year increases in CT utilization [1,2] and the introduction of teleradiology [3] have inexorably shifted practice toward rapid reporting of cross-sectional imaging. An unintended effect of these changes has been a perceptible reduction in patient contact, with radiologists becoming increasingly remote from patient care [4,5] and feeling personally detached from patients [6].

To "personalize" the reporting process, the idea of presenting patient photographs to reporting radiologists was investigated in an unpublished study [7]. The authors suggested that radiologists produced longer reports containing more incidental findings, and subjectively reported more meticulously and empathetically, when presented with patients' photographs. The study attracted considerable media attention and was even included in a published nonfiction book [8]. However, its significance remains unclear, because the incidental abnormalities were not classified in terms of clinical importance, and the findings have not been reproduced elsewhere.

The purpose of this study was to assess whether presenting patients' photographs to reporting radiologists has a clinically important impact on reported findings.

WHAT WE DID

This prospective study was approved by The Ottawa Hospital Institutional Review Board and was compliant with the Personal Information Protection and Electronic Documents Act and HIPAA. Written informed consent was obtained from all subjects.

Design

CT reports issued for the same examination, by the same radiologist, at two temporally separate interpretation sessions, one with an attached patient photograph ("face study") and one without a photograph ("no-face study"), were compared. A crossover design assigned patients to two groups: (1) face study reported at the first reading session and no-face study reported at the second reading session and (2) no-face study reported first and face study reported second. This minimized systematic bias associated with read order.

Recruitment and Study Population

Over a 6-month period, patients undergoing CT Abdomen and Pelvis examinations on lists supervised by coauthors were invited to participate. Exclusion criteria included age < 18 years, refusal or inability to consent, vascular studies, limited or incomplete imaging, and pregnancy. A photograph of the patient's face was taken before the scan using an off-the-shelf digital camera. Forty-seven eligible patients were recruited, but because of one of the coauthors leaving the institution midway through the study, 5 cases could not be analyzed. The final study group consisted of 42 patients: 25 with the face study reported on the first read and 17 with the no-face study reported on the first read. Fifty-two percent of participants were men and 48% women, and the mean age was 60 years. Sources of referral were the emergency department (48%), oncology (31%), general outpatients (5%), and general inpatients (16%).

Reporting

All studies were reported on PACS workstations using institutional voice recognition technology (PowerScribe 360; Nuance, Burlington, Massachusetts). For face studies, the patient's nonanonymized photograph opened in a separate window when the case was read (not integrated with DICOM images). Eleven staff radiologists with 1

1546-1440/14/$36.00 ■ http://dx.doi.org/10.1016/j.jacr.2014.09.028

to 10 years' experience participated, from the abdominal/emergency and musculoskeletal imaging divisions. Participants were cognizant of the study protocol but blinded to prior reports. The interval between first and second reads was 4 to 10 months. First reads were performed in the clinical environment (emergency room, outpatient, and inpatient reporting); second reads were performed in a nonclinical environment (research-protected time). In 26 cases, the first read was coreviewed by a resident or fellow and reported by the staff radiologist, and the second read was performed by the staff radiologist alone. In 16 cases, both reads were performed exclusively by a staff radiologist. Standardized reporting templates were used in both reads in 27 cases, in neither in 9 cases, and in one read but not the other in 6 cases.

Classification of Findings

Findings were classified into 8 categories according to whether they were related or unrelated to the clinical question (expected or incidental) and their clinical importance (no significance, minimal or doubtful significance, important or potentially important, and urgent or emergent). Clinical importance was determined on the basis of patient information, clinical details on requisition, and severity of reported abnormality (eg, trace pleural effusion in an elderly outpatient was classified as "minimal or doubtful significance," whereas moderate pleural effusion in an unwell inpatient was classified as "important or potentially important").

Impression concordance (presence in the conclusion of all urgent or emergent and important or potentially important findings listed in main body of report), number of recommendations made, and typographic errors were also assessed. The latter were included only if they subjectively reduced reader fluency and were categorized as significant only if they could cause clinical misinterpretation (eg, missed words, wrong laterality). Numbers of negative, normal, and qualifying statements were also assessed, defined on a per finding basis; for example, "no renal calculi, hydronephrosis, or solid lesion" was counted as 3 negative findings, whereas "normal kidneys" was counted as 1 negative finding. Qualifying statements included comments such as "pelvic organ assessment limited by streak artefact from hip prosthesis." These findings were summated and assessed for differences between groups.

Statistical Analysis

Data were analyzed using PSPP version 0.8.1 (http://www.gnu.org/software/pspp/). Validity was evaluated using the sign test or the Wilcoxon signed rank test to assess for significant differences in the presence and number of findings between the first and second reads and the Fisher exact test or χ^2 test to assess for significant differences between potentially confounding variables for each of the 8 classification categories.

The effect of viewing the patient's photograph on report content was assessed using the sign test and the Wilcoxon signed rank test to evaluate for significant differences in the presence and number of findings (in each classification category) between the face and no-face groups. Similar analyses were performed for impression concordance, typographic errors, recommendations made, and numbers of negative and normal findings. Because the numbers of negative and normal findings depended on congruent reporting and template use, 6 cases with discordant template use and 1 case that included the thorax were excluded from this subanalysis. Finally, a χ^2 test was performed on the same data subset to assess whether fellow or resident coreporting of one of the studies could have influenced the results.

OUTCOMES

Validity and Bias: Differences Between First and Second Reads

Seven of the 8 classification categories demonstrated no significant differences between the first and second reads (only expected findings of minimal or doubtful clinical significance were significant; $P = .03$). Chi-square testing of potentially confounding variables associated with read order (face or no-face study read first; fellow or resident coreporting of first read; reporting templates used in both, neither, or only one study) also demonstrated no significant difference in the number of findings between the face and no-face groups for the majority of categories.

Any systematic read-order effect (eg, reader memory) or systematic bias due to the aforementioned confounding variables was therefore unlikely. The few significant differences observed on χ^2 testing are difficult to interpret because of diminutive group size but include more incidental findings of important or potentially important significance on face studies when reporting templates were used for one but not the other study ($P = .02$) and fewer significant typographic errors in no-face studies when reporting templates were used for both reads ($P = .04$).

There was a significant increase in the presence and number of typographic errors and significant typographic errors on the first read compared with the second read ($P = .00$, $P = .00$, and $P = .02$, respectively). This was a significant source of bias for these variables, which were excluded from further analysis.

Effect of Viewing Patient Photographs: Differences Between Face and No-Face Studies

No significant difference was demonstrated between the face and no-face groups for the presence and number of expected or incidental findings of