Radiography xxx (2017) 1–6

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Development of a valid and reliable test to assess trauma radiograph interpretation performance

M.J. Neep [a, b, c, d, *], T. Steffens [e], V. Riley [f], P. Eastgate [c, d, g], S.M. McPhail [b, c, d]

[a] Department of Medical Imaging, Logan Hospital, Corner of Armstrong Road and Loganlea Road, Meadowbrook, Queensland, Australia
[b] Centre for Functioning and Health Research, Metro South Health, Buranda Plaza, Corner Ipswich Road and Cornwall Street Buranda, Brisbane, Australia
[c] School of Public Health & Social Work, Queensland University of Technology, Kelvin Grove, Brisbane, Australia
[d] Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Brisbane, Australia
[e] Department of Medical Imaging, Princess Alexandra Hospital, Ipswich Road, Brisbane, Australia
[f] Worcestershire Royal Hospital, Charles Hastings Way, Worcester, United Kingdom
[g] Department of Medical Imaging, Royal Brisbane and Women's Hospital, Bowen Bridge Road, Brisbane, Australia

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form
23 January 2017
Accepted 29 January 2017
Available online xxx

Keywords:
Image interpretation
Trauma
Performance
Validity
Reliability

ABSTRACT

Objectives: The purpose of this investigation was to develop and examine the preliminary validity and reliability among radiographers of a test to assess trauma radiograph interpretation performance suitable for use among health professionals.
Methods: Stage 1 examined 14,159 consecutive appendicular and axial examinations from a hospital emergency department over a 12 month period to quantify a typical anatomical region case-mix of trauma radiographs. A sample of radiographic cases representative of affected anatomical regions was then developed into the Image Interpretation Test (IIT). Stage 2 involved prospective investigations of the IIT's reliability (inter-rater, intra-rater, internal consistency) and validity (concurrent) among 41 radiographers.
Results: The IIT included 60 cases. The median (interquartile range) clinical experience of participants was 5 (2–10) years. Case scores were internally consistent (Cronbach's alpha = 0.90). Favourable inter-rater reliability (kappa > 0.70 for 58/60 cases, Intra-class correlation coefficient (ICC) > 0.99 for total score) and intra-rater reliability (kappa > 0.90 for 60/60 cases, ICC > 0.99 for total score) was observed. There was a positive association between radiographers' confidence in image interpretation and IIT score (coefficient = 1.52, r-squared = 0.60, $p < 0.001$).
Conclusions: The IIT developed during this investigation included a selection of radiographic cases consistent with anatomical regions represented in an adult trauma case-mix. This study has also provided foundational preliminary evidence to support the reliability and validity of the IIT among radiographers. The findings suggest that it is possible to assess image interpretation performance of adult trauma radiographs with this test.

Introduction

Accurate and timely interpretation of radiographic images is of paramount importance to the delivery of a high quality service in hospital emergency care settings. In these settings, health professionals from a range of clinical backgrounds are be required to interpret radiographic images as part of their clinical practice.[1] This may include medical professionals,[2] nurses,[3] physiotherapists[4] and preliminary clinical evaluations from radiographers[5] prior to the availability of a radiologist report. Education programs to enhance image interpretation ability among health professionals exist in both short-course format and formal university coursework.[6–8]

A valid and reliable Image Interpretation Test that clinicians could complete prior to and following image interpretation training would assist in the evaluation of the effectiveness of image interpretation education programs for enhancing image interpretation performance.[9] In the context of this study, validity refers to the ability of the interpretation test to appropriately quantify image

* Corresponding author. Department of Medical Imaging, Logan Hospital, Corner of Armstrong Road and Loganlea Road, Meadowbrook, Queensland, 4131, Australia.
E-mail address: Michael.neep@health.qld.gov.au (M.J. Neep).

http://dx.doi.org/10.1016/j.radi.2017.01.004
1078-8174/

interpretation performance. Reliability refers to the ability for test scores to be consistently reproduced under the same conditions either by the same test score rater (intra-rater reliability) or by another test score rater (inter-rater reliability), and internal consistency refers to whether the scores of individual cases within an interpretation test are correlated. Several Image Interpretation Tests have been reported in the literature.[2,10–16] However, there is a paucity of data describing the validity and reliability of these image tests which have, to the authors' knowledge, been developed for use in two studies.[3,17] Further, it is unclear whether the content of prior image tests represent the image case-mix that a clinician may expect to see in the context of their clinical practice.

There remains contention in the literature as to whether manufactured image tests (i.e., hand picked cases) are an accurate indicator of interpretive performance in comparison to image test banks that represent clinical practice. A typical manufactured test bank has an abnormality prevalence of 70%[17,18] in comparison to a lower reported abnormality prevalence of 20–30% in image banks that represent clinical practice.[19,20] A salient study by Hardy et al.[21] examined the influence of abnormality prevalence bias on the accuracy of interpretation of radiographs by investigating radiographers' performance on manufactured high abnormality prevalence image banks versus clinical practice image banks. The results indicated that the manufactured high abnormality prevalence test banks may overestimate abnormality detection ability.

The overarching aim of this investigation was to develop a valid and reliable Image Interpretation Test that could be used as an outcome measure to quantify radiographers' performance in interpreting adult trauma radiographs. This included three specific objectives. The first objective was to describe the anatomical region case-mix of axial and appendicular radiographic examinations in a hospital trauma setting over a 12 month period. The second objective was to generate an image test consisting of a sample of images from anatomical regions representative of the typical case-mix. The third objective was to develop an Image Interpretation Test (IIT) to quantify the performance of radiographers attempting to identify and describe potential abnormalities present in the selection of cases, and prospectively examine validity (concurrent) and reliability (intra-rater, inter-rater and internal consistency) of the IIT scoring among radiographers.

Methods

This investigation was approved by the Metro South Health and Queensland University of Technology Human Research Ethics Committees. Participants provided written informed consent.

Study design

A two-stage study design was implemented. Stage 1 involved the development of the IIT content, a reference standard for each case and scoring criteria. Stage 2 involved prospective examination of preliminary validity (concurrent) and reliability (intra-rater, inter-rater, and internal consistency) of the IIT scoring.

Stage 1 – IIT development

Selection of image test content for the IIT was consistent with guidelines suggested by Streiner and Norman.[22] Appendicular and axial skeleton radiographic images for suspected trauma were captured from digital radiographic examinations undertaken in a large metropolitan adult hospital emergency department. An audit of consecutive adult examinations over a 12 month period was undertaken to quantify the typical anatomical region case-mix referred for investigation in this setting. Quantifying the proportion of trauma cases from each anatomical region facilitated the selection of examinations representative of the anatomical region case-mix in typical clinical practice.

The total number of cases to be included in the test instrument was 60. The decision to include 60 radiographic examinations was considered carefully during study planning. The investigators were primarily influenced by consideration of the total duration of the assessment, as they wanted the utility of the IIT to be as adaptable to the clinical environment and feasible. This process began with the investigators examining a variety of university based assessments and observing that a 90 min assessment duration was common and thus would likely be a reasonable duration for the IIT. The authors then considered what would constitute a reasonable amount of time to interpret each examination. To the authors' knowledge, the only published guideline available that recommends an average time to interpret a radiographic examination is from the Royal College of Radiologists, United Kingdom. They recommend approximately 90 s per examination.[23] Considering these two factors (total assessment time divided by total time per examination), the total number of examinations to include in the IIT was elected to be 60.

The selection of the 60 cases followed a three-step process. First, the proportion of cases required for each body region was calculated (last row, Fig. 1). Second, a date was randomly selected as a starting point for identifying eligible consecutive cases. Third, consecutive cases from that date were examined against eligibility criteria (trauma-related referral, appendicular or axial skeleton, radiographic examination was complete) until the quota for each body region was filled.

A reference standard for each case and scoring criteria were then established. A reference standard is the correct interpretation for each case in the IIT and indicates whether a traumatic abnormality was present, and if present, the key elements (anatomical location and pathology) that should be described in an accurate interpretation. It was created by a panel of experienced experts (2 consultant radiologists and 1 consultant reporting radiographer) and involved two members of the panel, independent of one another, reporting each case in the IIT. The third independent expert was available to mediate any disagreements between the two primary experts. This consensus became the reference standard. When completing the IIT, participants were required to view each case and first decide whether an abnormality was present, and provide a brief description of the perceived abnormality (if present). Scoring an interpretation of a radiograph is not necessarily a binary decision between normal and abnormal. The authors of this study acknowledge that it is possible to have an incomplete or a partly correct or incorrect interpretation. After reviewing literature in the field, the authors of this study could not find an appropriate scoring criteria that incorporated the scoring of an interpretation that was potentially incomplete or only partly correct/incorrect. Therefore the investigators developed a new and novel scoring criteria to be trialled for cases with and without an abnormality present (described in Table 1).

Stage 2 – prospective validity and reliability testing

To examine the validity and reliability of the IIT and scoring system, a volunteer sample of radiographers who worked in the emergency medical imaging departments from three metropolitan hospitals in Brisbane, Australia were invited to participate ($n = 50$). Radiographers were considered eligible for inclusion if they had at least 12 months clinical experience and currently worked in an emergency setting. Forty-one radiographers met the inclusion criteria and provided informed consent.

Radiographers were first provided with guidelines for categorisation of radiographic examinations based on prior research by Robinson and colleagues.[24] These guidelines indicated that a

Please cite this article in press as: Neep MJ, et al., Development of a valid and reliable test to assess trauma radiograph interpretation performance, Radiography (2017), http://dx.doi.org/10.1016/j.radi.2017.01.004