ELSEVIER

Radiotherapy and Oncology 74 (2005) 275–281

RADIOTHERAPY & ONCOLOGY
JOURNAL OF THE EUROPEAN SOCIETY FOR THERAPEUTIC RADIOLOGY AND ONCOLOGY

www.elsevier.com/locate/radonline

Observer variability when evaluating patient movement from electronic portal images of pelvic radiotherapy fields

D. Geraint Lewis[a,*], Karen R. Ryan[b], Cyril W. Smith[a]

[a]*Department of Medical Physics, Velindre Hospital, Cardiff CF14 2TL Wales, UK*
[b]*Department of Radiotherapy, Velindre Hospital, Cardiff CF14 2TL Wales, UK*

Received 28 November 2003; received in revised form 12 November 2004; accepted 19 November 2004
Available online 23 December 2004

Abstract

Background and purpose: A study has been performed to evaluate inter-observer variability when assessing pelvic patient movement using an electronic portal imaging device (EPID).

Materials and methods: Four patient image sets were used with 3–6 portal images per set. The observer group consisted of nine radiographers with 3–18 months clinical EPID experience. The observers outlined bony landmarks on a digital simulator image and used matching software to evaluate field placement errors (FPEs) on each portal image relative to the reference simulator image. Data were evaluated statistically, using a two-component analysis of variance technique, to quantify both the inter-observer variability in evaluating FPEs and inter-fraction variability in patient position relative to the residuals of the analysis. Intra-observer variability was also estimated using four of the observers carrying out three sets of repeat readings.

Results: Eight sets of variance data were analysed, based on FPEs in two orthogonal directions for each of the four patient image sets studied. Initial analysis showed that both inter-observer variation and inter-fraction-patient position variation were statistically significant ($P<0.05$) in seven of the eight cases evaluated. The averaged root-mean-square (RMS) deviation of the observers from the group mean was 1.1 mm, with a maximum deviation of 5.0 mm recorded for an individual observer. After additional training and re-testing of two of the observers who recorded the largest deviations from the group mean, a subsequent analysis showed the inter-observer variability for the group to be significant in only three of the eight cases, with averaged RMS deviation reduced to 0.5 mm, with a maximum deviation of 2.7 mm. The intra-observer variability was 0.5 mm, averaged over the four observers tested.

Conclusions: We have developed a quantitative approach to evaluate inter-observer variability in terms of its statistical significance compared to inter-fraction patient movement. This will assist us in training and assessing observers required to perform this task on a routine basis.

Keywords: Electronic portal imaging; Observer variability; Pelvic radiotherapy

1. Introduction

Over the last 15 years electronic portal imaging devices (EPIDs) have become widely available as an alternative to conventional portal films for clinical radiotherapy verification. The development and deployment of the technology has been the subject of several comprehensive review articles during this period [2,7,11–13].

The main application of EPIDs to date has been for the evaluation of 'set-up' or 'field placement' errors (FPEs), i.e. displacements of patient position within the treatment field compared to that originally specified by the treatment plan and visualised by either simulator images or digitally reconstructed radiographs (DRRs). Hurkmans et al. [12] have reviewed the large number of studies performed in this area and provided recommendations on procedures to quantify, report and reduce set-up errors, based on the differing correction strategies and protocols developed in the various active centres.

Only a few investigations have been published on the degree of variability between reviewers reporting FPEs based on portal images. This can be important, as

* Corresponding author.

0167-8140/$ - see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.radonc.2004.11.009

inter-observer variations can introduce systematic errors in the determination of FPEs. One of these studies, performed by Bissett et al. [6] involved comparison of EPID images on a monitor with simulator films viewed on an adjacent light box. The observers (comprised of three groups: oncologists, radiation therapists/technologists and medical physicists) were required to assess the degree of conformity between the treated and prescribed fields on a five-point scale (ranging from 'definitely acceptable' to 'definitely unacceptable'). The analysis of agreement between individual observers and within the observer groups was carried out using 'kappa statistics' [1,10], kappa being the observed proportion of agreements between observers corrected for the expected proportion of random agreements.

Perera et al. [14] assessed observer variability in identifying geometric errors and clinical decision-making when using portal films. The observer cohort consisted of radiation oncologists and radiation therapists/technologists. The observers were required to view portal films side-by-side with simulator films on light boxes, estimate FPEs and to decide, using a yes/no response, whether the treatment needed adjustment or not. The study found that there was a high degree of variability both in assessing FPEs and in the number of recommended adjustments by each observer. One of the conclusions of the study was that centres involved in portal imaging should attempt to quantify the degree of variability in assessing FPEs and that as EPIDs with image registration capabilities became more widely available, this type of study should become more practical.

Subsequently, Suter et al. [15] and Dalen et al. [8] have reported on studies comparing the performance of oncologists and radiographers/technologists when evaluating portal images. The motivation in each case was to establish whether the task of portal image approval could be reliably devolved from the oncologists. Suter et al. [15] found that the radiographers demonstrated a 97% agreement with the oncologists (based on statistical predictivity tests [1]) when assessing displacements between portal and simulator films. The study involved assessment of several hundred films by 60 observers [16]. The study concluded that the radiographers demonstrated sufficient parity with the clinicians to justify a change in departmental protocols. In the work of Dalen et al. [8] the observers were required to assess displacements between digital portal and simulator images of the pelvis. Dalen et al. presented their results in terms of correlation coefficients describing the levels of agreement within and between two 'groups' consisting of three technologists and two oncologists. The coefficient of agreement between the groups (0.84) was higher than that within either group (0.72 for the technologists and 0.58 for the oncologists). The range of observations (i.e. the difference between the largest reading and the smallest) averaged across all five observers and 18 patient datasets was 2.3 mm for anterior images and 3.6 mm in the laterals.

Our centre is considering transferring responsibility for portal image assessment and approval from oncologists to the radiographer staff group. We have therefore investigated the degree of observer variability to be found within the latter group. The training task involved assessing FPEs over a series of fractions for given patients, in line with the decision-making process required in clinical practice. For example, if one were to adopt a correction protocol such as 'shrinking action level' (SAL) [4,5] or 'no action level' (NAL) [9], the task would involve assessing patient displacement from a series of portal images and implementing a correction strategy on the basis of the results obtained. In the NAL case, the protocol would require assessment of average FPE over a preset number of fractions and implementing a correction accordingly.

2. Materials and methods

2.1. Study details

The study involved the assessment of four test sets of pelvic radiotherapy verification images acquired using the Varian PortalVision Mk II EPID [17]. Treatment fields were delivered by a Varian Clinac 2100CD accelerator. Tests 1 and 3 involved images of actual treatment fields acquired at a beam energy of 10 MV and dose rate of 400 machine Monitor Units (MU)/min. Approximately 20 MU were delivered during the acquisition of each image. Test 2 was based on open field images, obtained prior to the treatment field in a 'double exposure' sequence (conventionally employed when imaging treatment fields $<(10\times10)\ \mathrm{cm}^2$). These images were acquired at a setting of 100 MU/min to reduce extraneous dose to the patient, images being acquired with <5 MU. A beam energy of 6 MV is used in these cases as image quality is somewhat better than that found at 10 MV. Test 4 involved a $(10\times10)\ \mathrm{cm}^2$ non-treatment 'set-up' field. The MV and MU parameters were the same as described for Test 2.

All images were obtained with the EPID operating in the 'fast acquisition/full resolution' mode. Image pixel resolution was $\sim$1 mm (for typical source-EPID distance of 130 cm). For evaluation purposes a standard processing filter from the PortalVision software suite was applied to all portal images in the tests, namely 'Enhance 2+ELW' which performs edge enhancement of bony structures and automatic adjustment of level (brightness) and window (contrast). All images were viewed under normal room lighting conditions. Representative images from the tests are shown in Fig. 1. The image sets were chosen to sample the range of matching tasks encountered clinically, thus differing degrees of difficulty were associated with each test. Test 1 involved a large anterior field with a large amount of pelvic bony structure visible and was deemed to be the easiest matching task. In contrast, Test 4, involving a lateral image with little more than the femoral head for outlining, was expected to be the most difficult. Tests 1, 2, 3