The Radiographer – vol. 51: 23-28

Measurement of Anatomy Contouring in EPI Review: A Practical Method for use in Radiation Therapy Departments

Michala Rybovic,[1,2] Georgia K.B. Halkett,[1] Marie T. Williams[1]

ABSTRACT

Radiation therapy treatment verification can be performed using hard copy portal films or digital Electronic Portal Images (EPI) of the treatment field, acquired at the time of treatment. This paper describes a practical method of assessing the accuracy of reference anatomy outlining, for treatment sites involving the pelvis, breast and lumbar spine. Seven original bone anatomy outlines contoured onto verification images of five patients, were printed on transparency sheets and reference points were marked at equal distances along the anatomy curves. Two sample anatomy contour sets were created by two independent radiation therapists who outlined visible bone anatomy on the same seven digitally reconstructed radiographs (DRR) and hard copy outlines were obtained. Three independent observers with differing levels of experience, assessed the discrepancies between the original anatomy contours and the sample sets on two occasions one week apart, by measuring the distances between the original and sample set contours (absolute values in mm). The degree of agreement between the same assessor on two occasions (intra-rater reliability) and between assessors (inter-rater reliability) was analysed using parametric analysis for levels of relationship and significant differences. This simple method of reference anatomy outline measurement was shown to be highly reliable within assessors and between assessors ($r > 0.87$ and $r^2 > 0.75$ for both intra- and inter-rater comparisons). This measurement process may be a suitable method, for undertaking quality assurance activities in image verification within radiation therapy departments.

INTRODUCTION

Successful radiation therapy treatment, relies on the correct delivery of radiation dose to the prescribed treatment area. It is paramount, that patients undergoing daily radiation therapy are set-up in the same position at every session or fraction. Inaccurate positioning at any treatment fraction may have serious implications for patients, especially in terms of achieving tumour control and reducing damage to surrounding healthy tissue.[1] The importance of reducing localisation errors in patient set-up has been widely reported[1-3] and highlights the rationale behind verifying patient positioning.

Verification of radiation therapy treatment involves comparing the treated radiation field against a reference image acquired during treatment planning. The reference image may either be in the form of a simulation film or a Digitally Reconstructed Radiograph (DRR). At present, there are two systems in widespread clinical use that enable monitoring and verification of patients' actual treatment field placement; port films and Electronic Portal Imaging Devices (EPID). Port films, encased inside a film cassette are hard copy x-ray images taken before or after treatment and capture exit radiation passing through the patient, to create a replicant image of the treatment field.[4] An EPID is a digital alternative to film cassettes and is able to convert exit radiation into an electrical signal, that can be displayed on a computer monitor in the form of a digital image.[5]

Port film verification, is commonly performed through visual comparison of anatomical landmarks observable in the reference image and the corresponding treatment image.[6] Discrepancies in the placement of the treatment field can be assessed and quantified through manual measurements and demagnification calculations between the two media. The difference in image quality between the reference image and treatment image may hinder the accuracy of port film verification as these techniques rely on interpretative skills of the reviewers and hence are prone to a high degree of variation and subjectivity.[7-9] Bissett et al. (1995)[6] highlighted the role of subjectivity in the interpretation of verification images for radiation treatment in a correlational study. Their findings demonstrated minimal agreement (kappa value of $\kappa = 0.096$) in image interpretation between four radiation oncologists, eight radiation therapists and four medical physicists all of varying levels of experience in image evaluation. The effect of previous experience in image interpretation of subjects within this study was not analysed.

Electronic Portal Image (EPI) analysis may also be performed through visual comparison, however, more objective EPI verification techniques use automatic computer-aided anatomy matching algorithms.[7,10,11] Commonly, EPI review consists of interactive contouring of anatomy landmarks on a reference DRR and anatomy matching of the contoured anatomy curves with a treatment EPI. Although the field placement accuracy is computer generated and hence objective, it should be acknowledged that contouring and anatomy matching processes rely on individual reviewers and their experience in image interpretation.

Reliability by definition, is the extent to which a test or measurement result is reproducible.[12] To date, there have been a number of studies[6,8,9] that have reported the accuracy of observers interpreting verification images, however, no specific studies focused on quantitative assessment of intra- and/or inter-rater reliability of the anatomy matching component of EPI review. In addition there has been a paucity of research exploring measurement procedures for investigating how differences in reference anatomy contours between observers could be measured. The difficulties posed by this area of research are likely to arise from

Authors:
1. Division of Health Sciences, University of South Australia
2. Department of Radiation Oncology, Royal Adelaide Hospital, South Australia

Author for Correspondence: M. Rybovic
Address: Department of Radiation Oncology, Royal Adelaide Hospital
North Terrace, Adelaide SA Australia 5000
Telephone: +61 8 8222 4390
Fax Number: +61 8 8222 2016
E-mail: mrybovic@mail.rah.sa.gov.au

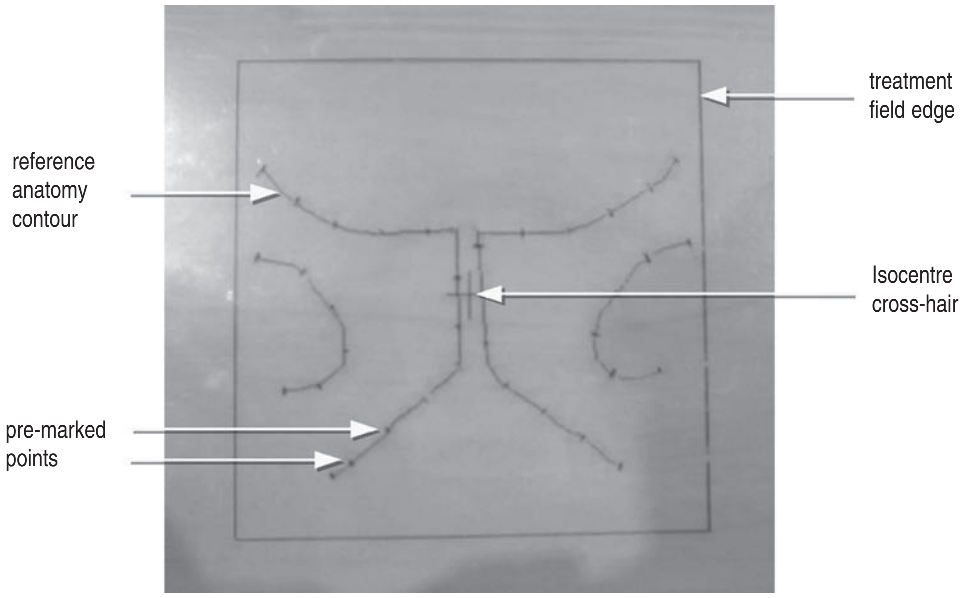


Figure 1: Prepared original reference anatomy image.

the lack of reliable and easily administered methods of measuring the accuracy of contoured anatomy curves. Therefore, the aim of this paper is to present a practical method of measuring contour accuracy for reference DRR images and report the level of reliability achieved between three independent assessors.

METHODS

Study Design

The study was a same subjects, repeated measures, observational study. Three assessors performed measurements on two separate occasions one week apart. Ethical approval was obtained from the University of South Australia Research Ethics Committee and the Royal Adelaide Hospital Ethics Committee.

Subjects

A convenient sample of three assessors (two qualified radiation therapists and one final year undergraduate radiation therapy student) was selected to participate in this study. Both qualified radiation therapists were currently working in a radiation therapy department, but had limited direct experience in anatomy contouring of DRR images.

Procedure

The study procedure can be delineated into three discrete steps;

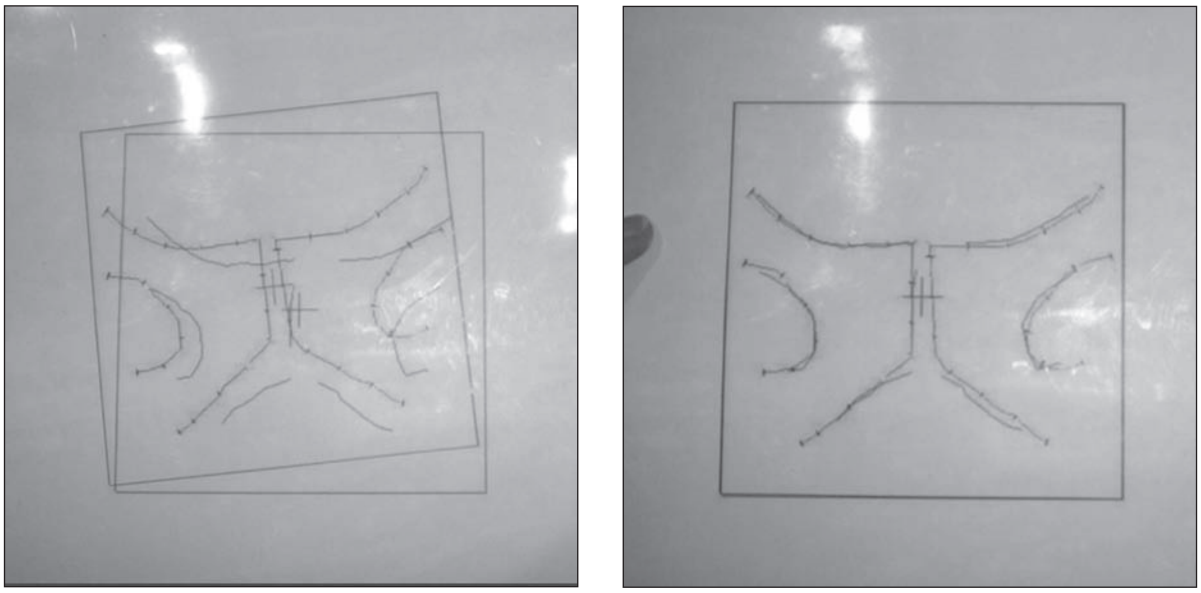


Figure 2: Superimposition of images to facilitate measurement of discrepencies.
a) Images prior to superimposition **b) Superimposed images**

retrieval of images, creation of sample sets of anatomy contours and measurement of discrepancies from the sample sets. The first step involved accessing the patient verification database at the Department of Radiation Oncology, Royal Adelaide Hospital. Seven reference DRR images were randomly selected from this archive by the principal author, using a computer generated random selection tool in Microsoft Excel. These seven images represented five previous patients who had completed a course of radiation therapy with treatment review using EPI. The seven reference images were:

1. AP (anterior-posterior) pelvis
2. AP pelvis
3. Left lateral pelvis
4. AP chest
5. Medial tangent using symmetric collimators
6. Medial tangent using asymmetric collimators
7. PA (posterior-anterior) lumbar spine

A trained EPI radiation therapist had previously contoured all seven DRRs. The reference anatomy outlines, treatment field edges and isocentre crosshairs for each image produced by the EPI radiation therapist, were retrieved and printed onto transparency sheets. For each of the seven images points were marked on the anatomy curves, using a permanent marker at equidistant 10 mm points. Figure 1 is an example of a prepared transparency of an original reference anatomy image.

The second step, involved the creation of two sample sets of anatomy contours. Two qualified radiation therapists (independent of the assessors), were asked to outline the visible bone anatomy on each of the seven original DRRs to create a total of fourteen sample images. Each radiation therapist completed this task independently and was blind to the original reference anatomy outlines. On completion of anatomy contouring for each image, the outlines and corresponding field edges were saved and hardcopies were printed at the same scale as the previously contoured images. In essence, sample sets of anatomy contouring for two independent radiation therapists (sample set A and sample set B), were created using the seven DRR images for which previous anatomy contours were available.

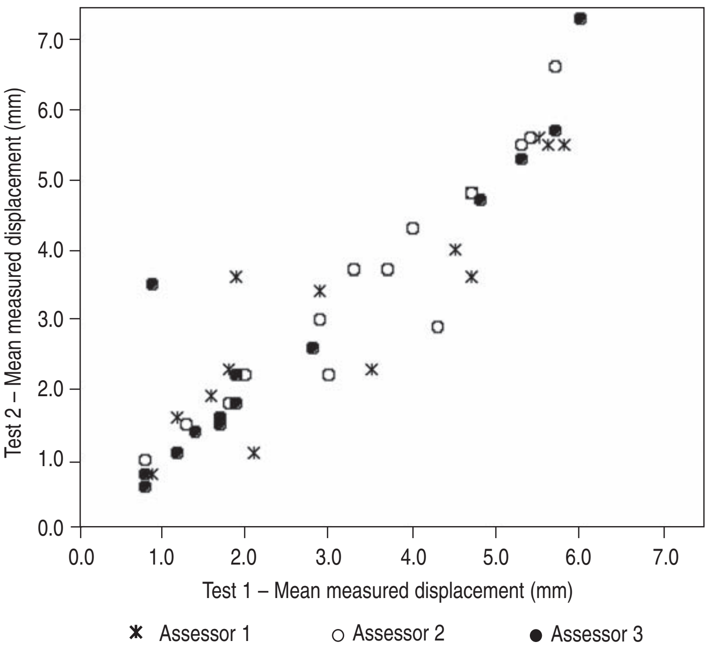
The third step, involved measuring contour displacements between the original contoured images and the paired sample sets, A and B. Three assessors (two qualified radiation therapists and one final year undergraduate radiation therapy student), were invited to measure discrepancies on two separate occasions, seven days apart. Each assessor was provided with a set of clear and unambiguous instructions or rules for measuring displacements. For each hardcopy image the appropriate transparency containing the original anatomy contouring was provided. The transparency was placed over the hard copy image and adjusted, so that the field edges and isocentre crosshairs were superimposed. Figure 2 demonstrates how the two images were overlayed to enable contour displacement measurements. All three assessors used the same small metal ruler, to manually measure distances from the original anatomy curves to the corresponding sample anatomy curves at each pre-marked point. All contour displacement measurements were taken perpendicularly from the original contour lines and recorded if they were no more than 15 mm away. If two sample contour lines were within the vicinity of a pre-marked point, the closest distance was recorded. Measurements were documented as absolute values to the nearest 0.5 mm. The assessor was required to enter each discrepancy measure into a Microsoft Excel spreadsheet. This information was not formally checked for transfer errors, as the assumption was made that the assessors entered data correctly. The identical measurement process

Table 1: Comparison of intra-rater reliability for each assessor

Image	Assess. 1 Test 1	Assess. 1 Test 2	Assess. 2 Test 1	Assess. 2 Test 2	Assess. 3 Test 1	Assess. 3 Test 2
Sample set A						
1. AP	2.1	1.1	2.0	2.2	0.8	0.8
2. AP	1.8	2.3	2.9	3.0	1.7	1.5
3. LLAT	1.6	1.9	1.8	1.8	1.4	1.4
4. AP	4.5	4.0	4.3	2.9	2.8	2.6
5. MT	5.8	5.5	5.7	6.6	5.7	5.7
6. MT	5.5	5.6	5.4	5.6	6.0	7.3
7. PA	0.9	0.8	0.8	1.0	0.8	0.6
Sample set B						
1. AP	1.9	3.6	3.3	3.7	1.9	1.8
2. AP	1.2	1.6	1.3	1.5	1.2	1.1
3. LLAT	3.5	2.3	3.0	2.2	1.7	1.6
4. AP	4.7	3.6	4.0	4.3	0.9	3.5
5. MT	4.7	4.8	4.7	4.8	4.8	4.7
6. MT	5.6	5.5	5.3	5.5	5.3	5.3
7. PA	2.9	3.4	3.7	3.7	1.9	2.2
Paired t-test p =	*0.81*		*0.77*		*0.29*	
r (df_{12}, r crit =0.532)	*0.90*		*0.95*		*0.93*	
r^2	*0.81*		*0.90*		*0.86*	
ANOVA (single factor, $df_{1,26}$ F crit = 4.23	*p = 0.94, F = 0.006*		*p = 0.95 F = 0.005*		*p = 0.77 F = 0.09*	
ICC	*-0.99*		*-0.99*		*-0.84*	

KEY: Assess. - assessor; AP - anterior-posterior; LLAT - left lateral; MT - medial tangent; PA - posterior-anterior; *df* - degrees of freedom; *crit* - critical; *ANOVA* - analysis of variance; *ICC* - intra-class correlation coefficients

Figure 3: Reliability of contour displacement measurements



was then repeated seven days later and data entered into a new spreadsheet.

Data Analysis
The mean of the discrepancy for all points in each image was calculated. All data collected was interval and was analysed using Microsoft Excel 2002 SP-2. Linearity of relationships between occasions of testing for each assessor (intra-rater reliability) was calculated using Pearson's correlation coefficient (r) (with significant relationships calculated to greater than r = 0.53, df 12) using a table of critical values for r and the coefficient of determination (r^2).[12] Differences between occasions of testing for each assessor were calculated using paired t-tests, single factor analysis of variance (ANOVA) and intra-class correlation coefficients (ICC). The relationships between assessors (inter-rater reliability) were calculated using Pearson's correlation coefficient, repeated measures ANOVA and unpaired t-tests. The level of significance was set at $p < 0.05$ and correlations greater than r = 0.5 were considered to be indicative of strong relationships.[13]

RESULTS
Complete data sets were obtained for all assessors for each image and on each occasion of testing. Table 1 demonstrates the results of mean displacements and comparisons between occasions of testing for each assessor (intra-rater reliability). For all three assessors, very high and significant linear relationships existed as illustrated in Figure 3 and no significant differences were calculated between the two occasions of testing.

As there were very strong relationships between occasions of testing and no significant differences between scores on either occasion of testing, data for tests 1 and 2 were pooled for each assessor (n = 28 data points for each assessor).

When assessors were compared with each other using r, r^2 and unpaired t-tests, very strong relationships and no significant differences were calculated. However, when a repeated measures ANOVA was calculated a significant difference between assessors was evident (p = 0.0001). Paired t-tests determined that there was a difference between Assessor 2 vs. Assessor 3 and Assessor 1 vs. Assessor 3. Hence Assessor 3 produced results which were different to the other two assessors. Table 2 demonstrates comparison of the three assessors using r, r^2, unpaired t-tests, repeated measures ANOVA and paired t-test.

DISCUSSION
The results of this study indicate that this system for measuring and reporting discrepancies between the original anatomy contours produced by the EPI radiation therapist and independent radiation therapists generates surprisingly stable and reliable results. Previous studies[6,8,9] have relied upon categorical scoring systems devised to indicate the degree of agreement between observers. This current system provides a method of recording specific displacements between contoured lines produced from two sources, in this case, a radiation therapist with expertise in reviewing EPI images and two independent radiation therapists (sample sets A and B) with limited EPI experience. The use of continuous data (mm) within this measurement process, allows for meaningful mathematical relationships to be analysed, both between assessors and between occasions of testing by the same assessor.

The instructions or rules governing the measurement process, though specific, were brief and simple to follow. All three assessors demonstrated very high degrees of relationships between the two occasions of testing (intra-rater reliability) and no significant differences in the measurements were found on either occasion of testing.

Differences between assessors were highlighted when inter-rater reliability analysis was undertaken. While a high degree of linearity was present between assessors, a significant difference

Table 2: Comparison of three assessors using r, r2, unpaired t-tests, repeated measures ANOVA and paired t-test

	Pearson's correlation coefficient r, df $_{26}$, *r crit = 0.381*	r^2	Unpaired t-test *p =*	Repeated measures *ANOVA* *F crit = 3.17*	Paired t-test *p =*
Assessor 1 vs Assessor 2	0.94	0.88	0.73		0.17
Assessor 2 vs Assessor 3	0.88	0.79	0.14	*p* = 0.0001* F = 11.0	0.0004*
Assessor 1 vs Assessor 3	0.87	0.75	0.25		0.006*

KEY: df - degrees of freedom; *crit* - critical; *ANOVA* - analysis of variance; *denotes significant difference, $p < 0.05$

emerged between the results of Assessor 3 and the other two assessors. In essence, Assessor 3 had a significantly lower mean for their measures of discrepancies compared to Assessors 1 and 2. It should be noted that smaller means in discrepancy measures in this case, do not actually reflect a greater accuracy of measurement. There could be a number of reasons for this finding. First, Assessor 3 was responsible for developing and fine tuning the measurement process used in this study, hence it is possible that prior experience of the measurement system and practice had an effect. Second, the reference anatomy contours were approximately 0.5 mm in width when printed on transparency sheets and paper hard-copies. This could have an impact on the final discrepancy measure, if assessors recorded displacements either from the outside or middle of the anatomy contour line.

This simple and time efficient measurement system provides a method for quantifying the ability of radiation therapists, to determine differences between reference anatomy contours. It is essential that contouring of reference anatomy on a DRR is accurate, because the outlines are directly transferred and used in the matching of observable anatomy in acquired treatment EPIs. If a discrepancy is evident between the pair of images, there may be several possible reasons for this observation, including an incorrect patient position, mobility or poor anatomy contouring on the DRR. Poor outlining of reference anatomy structures may result in anatomy matching errors that are within acceptable limits, when in actual fact there could be gross field placement errors or treatment deviations. This measurement system could be used to determine the consistency and accuracy of radiation therapists contouring anatomy. Such a system could be readily implemented into departmental quality assurance programs to determine the level of accuracy exhibited by radiation therapists when performing treatment verification tasks.

CONCLUSION

This study has presented a simple measurement process for determining the discrepancies between anatomy contours using equipment readily available. Determining reliability both within and between assessors, proved to be straightforward and used uncomplicated statistical analysis which could be completed by other radiation therapists. Such a measurement system could be adapted for investigating other aspects of accuracy in treatment verification and may provide a useful tool for quality assurance programs within radiation therapy departments.

ACKNOWLEDGEMENTS

The authors wish to thank the Department of Radiation Oncology at the Royal Adelaide Hospital for supporting this investigation and both Courtney Oake and Alison Martin for their participation.

REFERENCES

1. Dutreix, A.1984. When and how can we improve precision in radiotherapy? *Radiotherapy and Oncology;* 2: 275-292.
2. Herman, M., Abrams, R., Mayer, R. 1994. Clinical use of online portal imaging for daily patient treatment verification. *International Journal of Radiation Oncology Biology Physics;* 28(4): 1017-1023.
3. Marks, J., Haus, A., Sutton, H.1976. The value of frequent treatment verification films in reducing localization error in the irradiation of complex fields. *Cancer*; 37: 2755-2761.
4. Munro P.1995. Portal Imaging Technology: Past, Present, and Future. *Seminars in Radiation Oncology*; 5(2): 115-133.
5. Antonuk, L. 2002. Electronic portal imaging devices: a review and historical perspective of contemporary technologies and research. *Physics in Medicine and Biology;* 47: 31-65.
6. Bissett, R., Boyko, S., Leszczynski, K., Cosby, S., Dunscombe, P., Lightfoot, N. 1995. Radiotherapy portal verification: an observer study. *British Journal of Radiology;* 68(806): 165-174.
7. Bissett, R., Leszczynski, K., Loose, S., Boyko, S., Dunscombe, P. 1996. Quantitative vs. subjective portal verification using digital portal images. *International Journal of*

Radiation Oncology Biology Physics; 34(2): 489-495.

8. Lavertu, S., Girouard, L., Pouliot, J. 2000. Observation study of electronic portal images for off-line verification. *Radiotherapy and Oncology;* 54: 47-55.
9. Perera, T., Moseley, J., Munro, P. 1999. Subjectivity in Interpretation of Portal Films. *International Journal of Radiation Oncology Biology Physics;* 45(2): 529-534.
10. Gilhuijs, K., van Herk, M. 1993. Automatic on-line inspection of patient setup in radiation therapy using digital portal images. *Medical Physics;* 20(3): 667-677.
11. Gilhuijs, K., Touw, A., van Herk, M., Vijlbrief, R. 1995. Optimization of automatic portal image analysis. *Medical Physics;* 22(7): 1089-1099.
12. Polgar, S., Thomas, S. 2000. *Introduction to Research in the Health Sciences.* 4th ed. Churchill Livingstone; London, England.
13. Cozby, P. 1997. *Methods in behavioural research.* 6th ed. Mayfield publishing company; California, United States of America.

Peer Reviewed
Submitted: November 2003
Accepted: February 2004