

Evaluation of atlas-based auto-segmentation software in prostate cancer patients

Stuart Greenham, DipAppSc(TherRad), MIR, BAppScComp,¹ Jenna Dean, BMedRadSc(RT),² Cheuk Kuen Kenneth Fu, BSc Rad (RT), MBT,³ Joanne Goman, BMedRadSc(RT),⁴ Jeremy Mulligan, BMedRadSc(RT),² Deanna Tune, BAppSc(MRT),¹ David Sampson, BAppSc(MRT),³ Justin Westhuyzen, MSc, PhD,¹ & Michael McKay, MBBS(Hons), PhD, MD³

¹Department of Radiation Oncology, North Coast Cancer Institute, Coffs Harbour Health Campus, Coffs Harbour, New South Wales, Australia

²North Coast Cancer Institute, Port Macquarie Health Campus, Port Macquarie, New South Wales, Australia

³North Coast Cancer Institute, Lismore Health Campus, Lismore, New South Wales, Australia

⁴Department of Radiation Oncology, Calvary Mater Newcastle, Newcastle, New South Wales, Australia

Keywords

Computer assisted, computer-generated 3D imaging, diagnostic imaging, pelvis, prostate, radiotherapy planning

Correspondence

Stuart Greenham, Department of Radiation Oncology, North Coast Cancer Institute, Coffs Harbour Health Campus, Coffs Harbour, New South Wales 2450, Australia.
Tel: +61 2 66565760; Fax: +61 2 5556 5128;
E-mail: stuart.greenham@ncahs.health.nsw.gov.au

Funding Information

No funding information provided.

Received: 17 February 2014; Revised: 17 July 2014; Accepted: 17 July 2014

Journal of Medical Radiation Sciences 61 (2014) 151–158

doi: 10.1002/jmrs.64

Abstract

Introduction: The performance and limitations of an atlas-based auto-segmentation software package (ABAS; Elekta Inc.) was evaluated using male pelvic anatomy as the area of interest. **Methods:** Contours from 10 prostate patients were selected to create atlases in ABAS. The contoured regions of interest were created manually to align with published guidelines and included the prostate, bladder, rectum, femoral heads and external patient contour. Twenty-four clinically treated prostate patients were auto-contoured using a randomised selection of two, four, six, eight or ten atlases. The concordance between the manually drawn and computer-generated contours were evaluated statistically using Pearson's product-moment correlation coefficient (r) and clinically in a validated qualitative evaluation. In the latter evaluation, six radiation therapists classified the degree of agreement for each structure using seven clinically appropriate categories. **Results:** The ABAS software generated clinically acceptable contours for the bladder, rectum, femoral heads and external patient contour. For these structures, ABAS-generated volumes were highly correlated with 'as treated' volumes, manually drawn; for four atlases, for example, bladder $r = 0.988$ ($P < 0.001$), rectum $r = 0.739$ ($P < 0.001$) and left femoral head $r = 0.560$ ($P < 0.001$). Poorest results were seen for the prostate ($r = 0.401$, $P < 0.05$) (four atlases); however this was attributed to the comparison prostate volume being contoured on magnetic resonance imaging (MRI) rather than computed tomography (CT) data. For all structures, increasing the number of atlases did not consistently improve accuracy. **Conclusions:** ABAS-generated contours are clinically useful for a range of structures in the male pelvis. Clinically appropriate volumes were created, but editing of some contours was inevitably required. The ideal number of atlases to improve generated automatic contours is yet to be determined.

Introduction

Manual contouring of targets and organs at risk (OAR) for radiotherapy planning is a tedious and labour intensive process. Consistent and accurate delineation of OAR and target structures for prostate patients is vital when using dose escalation and intensity modulated

radiation therapy (IMRT) planning techniques. All prostate patients at our institution are contoured using departmental guidelines that are based on eviQ recommendations¹ in which structures are contoured as a solid organ to ensure consistent delineation of OAR and target volumes. However, delineation of regions of interest (ROIs) still result in intra- and inter-observer

variations due to factors such as user preferences and experience.² Streamlining the contouring process using computer-assisted atlas-based methods could reduce these intra- and inter-observer variations as well as reduce the time required to complete contouring tasks.²⁻⁴

The atlas-based auto-segmentation software program (ABAS; Elekta Instrument AB Stockholm, Stockholm, Sweden) provides a computerised auto-contouring functionality. The software uses atlases with pre-defined ROI as templates to automatically delineate contours on a new patient's computed tomography (CT) data set. ABAS has the potential to not only reduce the burden of manual contouring but also to minimise intra- and inter-observer variations.

ABAS deforms atlases of anatomy from a user-defined reference atlas onto a new patient's CT data set, creating a structure set fitted to the patient's unique anatomy. Users have the option of running a single atlas or a fusion of multiple atlases using the simultaneous truth and performance level estimation (STAPLE) method to auto-generate contours.⁵ The STAPLE method involves running a patient's CT data set through ABAS with a number of appropriate atlases selected. ABAS generates a contour for each selected atlas before creating a fusion or 'stapled' structure set which is designed to be the best fit from all of the chosen atlases for the anatomy of the current patient. Due to variations in patient anatomy, especially the bladder and rectum in patients with pelvic malignancy, it could be preferable to run multiple atlases to attain a better result. Utilising ABAS for this group of patients could potentially provide substantial gains in efficiency with improved consistency, as has been reported from studies conducted in the head and neck region and the brain.⁶⁻⁸

The primary goal of the present study was to evaluate the accuracy and clinical applicability of the ABAS software when auto-contouring male pelvic anatomy, specifically for prostate patients. We sought to improve our understanding of the performance and limitations of the ABAS software to help facilitate its introduction into routine clinical use for all prostate patients at our institution.

Methods

The present evaluation of the ABAS software package (Elekta Inc.) employed data from patients treated between 2010 and 2012; auto-generated contours were not used prospectively in the treatment planning of patients. The work met the criteria for a Quality Improvement project according to the NSW Health Ethics Guideline document GL2007_020 and did not require formal ethical review.

Initial familiarisation and review of the software was completed on ABAS 1.0 using the data sets of 24 patients.

Each patient in the group was contoured with an arbitrarily selected atlas as well as the demonstration atlas provided with the program. The upgrade to ABAS 2.0 provided the opportunity to test multiple atlases simultaneously, thus allowing the software to access a larger range of solutions to refine contours to the individual patients. Making use of the earlier review/assessment methodology the analysis was repeated on all cases in order to review stability and in particular determine if improvements in the software, specifically the STAPLE algorithm translated into improvements in contour accuracy.

The contours of 10 randomly selected clinically treated prostate patients were used to create atlases in ABAS. All prostate patients at our cancer institute are planned and treated supine with bladder full and rectum empty and contoured according to departmental protocol, based on The Royal Australian and New Zealand College of Radiologists Faculty of Radiation Oncology Genito-Urinary Group (FROGG) guidelines.⁹ All structure contours undergo a stringent checking process as part of the overall planning quality assurance before a patient is treated. Contours are initially contoured by the planning radiation therapist, checked by the attending radiation oncologist and by a peer radiation therapist. It should be noted that the prostate volumes are contoured with the benefit of computed tomography-magnetic resonance imaging (CT-MRI) registration unless MRI is contraindicated for the patient. The addition of MRI data provides enhanced soft tissue definition which is valuable for the delineation of target volumes; however, at present, ABAS is unable to utilise the information from this imaging modality.

Using the 10 atlases, CT data sets of 24 randomly selected prostate patients clinically treated at our cancer institute with IMRT and image-guided radiation therapy (IGRT) were auto-contoured using ABAS. Each case was run through the system with a single 'demo' atlas (supplied by Elekta Inc.) as well as a random selection of two, four, six, eight or ten stapled atlases.

The manually drawn 'as treated' contours were regarded as the 'gold standard' volumes and were used as a reference for those generated in ABAS. These contours included the bladder, rectum, femoral heads and external patient contours. Although not required for clinical purposes, the pelvis structure was also contoured as this structure functionally assisted the ABAS software in the placement of adjacent structures. The 10 cases used to create atlases were chosen as they provided a range of patient shape and size for both internal and external anatomy. The atlases were named according to the bladder or rectal volumes. This allowed for small, regular and large atlases of either the bladder or the rectum to be

created. The volumes of the rectum or bladder were arbitrarily chosen and did not correlate with any published sizing system.

Cases were evaluated by six radiation therapists comparing ABAS auto-segmented structures against 'as treated' structures. Bladder, prostate, rectum, femoral heads and patient contours were assessed on two criteria – a qualitative evaluation of the structure's clinical appropriateness and a quantitative evaluation based on the volume of ABAS structures versus the manually drawn equivalent.

The qualitative evaluation classified the degree of agreement of each structure into a descriptive category as shown in Table 1. The categories ranged from 'good agreement', minor to major editing required, to 'gross error'; and all were specifically defined. This methodology provided a means of quantifying the frequency and 'goodness of fit' ABAS was able to achieve in a typical clinical situation; therefore this approach was considered a suitable surrogate measure of clinical performance.

The degree to which ABAS-generated volumes correlated with 'as treated' volumes, manually drawn, were assessed using the Pearson's product-moment correlation coefficient (r). Although some authors have used a more sophisticated dice similarity coefficient to assess the similarity of structures,^{2,6,10} our approach quantifies the 'goodness of fit' of the ABAS-generated volumes with respect to our reference method. Correlation coefficients were calculated for the patient, bladder, prostate, rectum and femoral head contours (left and right) and were tested among various atlas groups, that is, the Demo atlas (supplied by Elekta Inc.), five stapled atlas groups (2, 4, 6, 8 and 10 atlas staples) as well as a pooled group of all the atlas staples. A probability value $P < 0.05$ (two tailed) was considered statistically significant.

Results

Twenty-four patients were contoured in the study; only one gross error was recorded. In this case the external

patient contour consistently failed regardless of the number of atlases used; ABAS generated 'streaks' throughout the data set, and missed the majority of the superior and inferior slices completely. Missing the superior and inferior slices of the external patient contour was seen in 19 (80%) of the remaining 23 patients; but in terms of qualitative analysis still scored either 1 (acceptable) or 2 (very minor edits required) over 90% of the time as missing these slices did not affect the overall outcome in terms of planning or treatment for the patient.

Femoral heads, external patient contour and bladder contours were the most consistently acceptable in the qualitative evaluation (Table 2). Femoral heads and pelvis were clinically acceptable in >90% of the time. Bladder contours were clinically acceptable in ~80% of the time. It was noted that ABAS tended to contour the inner bladder wall rather than the whole bladder and this necessitated some editing in about 20% of cases (calculated from data in Fig. 2). The rectum was clinically acceptable in ~50% of the time. On this volume, ABAS tended to contour further superiorly and inferiorly than the protocol guideline, however, around the prostate and area of interest the generated structure was reasonable.

The percentages of ABAS contours categorised as clinically acceptable or only requiring minor edits for each atlas test group are shown in more detail in Figures 1 and 2. As discussed earlier, it shows that the femoral heads were generally classified as clinically acceptable even though volumetrically they did not always match the volumes in the departmental protocols. The patient contour and bladder were also contoured reasonably well with the higher number of stapled atlases. These contours are also ones where required edits are unlikely to have a detrimental effect on patient outcomes. The rectum contours fell into each editing category as shown in Figure 2. This structure is one in which differences from protocol could be consequential; the rectum is usually a structure that is close to tolerance doses. If it is over contoured, then it could give misleading information and lead to increased patient

Table 1. Categorisation of agreement between ABAS-generated contours and the 'as treated' reference.

Category	Descriptor	Definition
1	Good agreement	Structure acceptable to treat 'as is'
2	Very minor differences	Edits to approximately 10% or less of the CT slices required. Still clinically acceptable.
3	Minor differences	Modest noncritical edits required 10–20% of the volume predominately outside the clinically relevant portion of the structure
4	Edits required	Modest edits required 10–20% of the volume in areas of the volume likely to affect clinical outcome
5	Moderate edits required	Moderate changes required 20–50% of slices require a manual edit to meet clinical standards.
6	Major edits required	Significant changes to meet typical clinical standards in >50% of slices need to be manually edited
7	Gross error	No resemblance to the clinical structure or >75% slices needing edit

Table 2. Overall qualitative evaluation: percentage of ABAS contours meeting clinical requirements.

Target	Acceptable or minor edits required ¹ (%)
Patient contour ($n = 129$) ²	76.74
Bladder ($n = 129$)	77.52
Prostate ($n = 119$)	21.01
Rectum ($n = 121$)	48.76
Right femoral head ($n = 134$)	94.03
Left femoral head ($n = 133$)	93.98
Right pelvis ($n = 114$)	95.61
Left pelvis ($n = 116$)	95.69

¹Categories 1 and 2 (Table 1).²Contours pooled from all atlas combinations.

morbidity. The prostate was another contour in which edits were required more often than not to get contours that were similar to the ‘as treated’ volumes. This difference can be at least partially explained by the fact that the reference contours for each patient were contoured based on MRI instead of CT imaging.

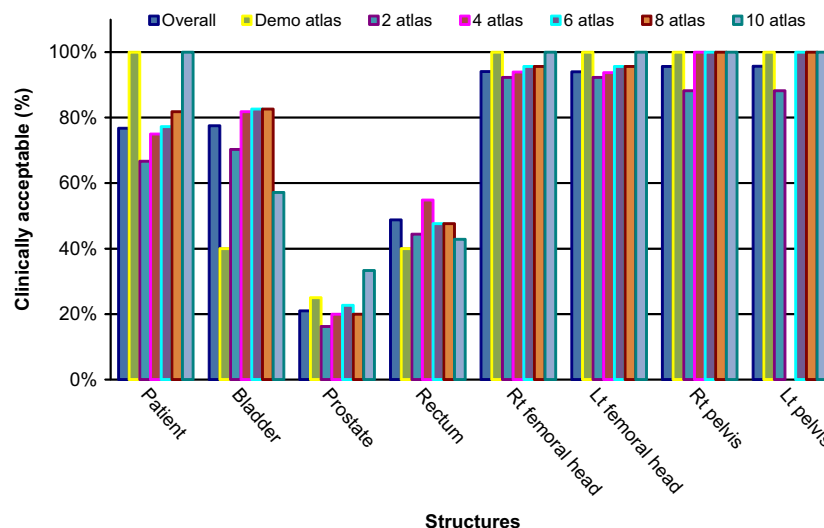
The degree of agreement between the ‘as treated’ reference and the ABAS-generated volumes are summarised in Table 3. An example of a correlation (for the bladder) is shown in Figures 3 and 4. In general, the analysis shows that there is a high degree of correlation between the two procedures. The closest correlations were noted when the demonstration atlas provided by the vendor was used. In terms of the atlas staples, the best correlation with the reference was noted for the bladder; moreover, the number of atlases used did not appear to change the reliability of this result (Table 3). The patient

contour was generally reasonable, but the correlations were weakest when six atlases were stapled together. The rectum was quite inconsistent with correlations becoming poorer after the inclusion of four or more atlases in the STAPLE algorithm. Correlations for the prostate were variably poor (apart from the demo atlas), which aligns with earlier discussions regarding the different imaging modalities used for prostate contouring. For the femoral heads, the best results were obtained when six atlases were used.

Discussion

The present study investigated the utility of a computer-based software program to generate clinically relevant contours in the assessment of prostate patients. Overall, ABAS generally produced structures that when compared with clinician contours were essentially the correct volume, similar in shape and in the correct place. In general, the software generated volumes that were in close agreement with manually drawn ones, particularly for hard tissues like bone and where there was clear delineation between the structure and its surrounding tissue. This correlation was supported by the qualitative analysis which demonstrated that high percentages of all contours could be used with little or no manual editing. For example, over three quarters of all bladder contours were clinically acceptable, or required very minor editing (Table 2). For the femoral heads and pelvis, nine in ten scans met clinical requirements.

The clinical utility of ABAS was supported by the quantitative analysis. Although the closest correlations

**Figure 1.** Percentage of ABAS volumes categorised as clinically acceptable (categories 1 and 2) versus the number of atlas groups scanned. ABAS, atlas-based auto-segmentation; rt, right; lt, left.

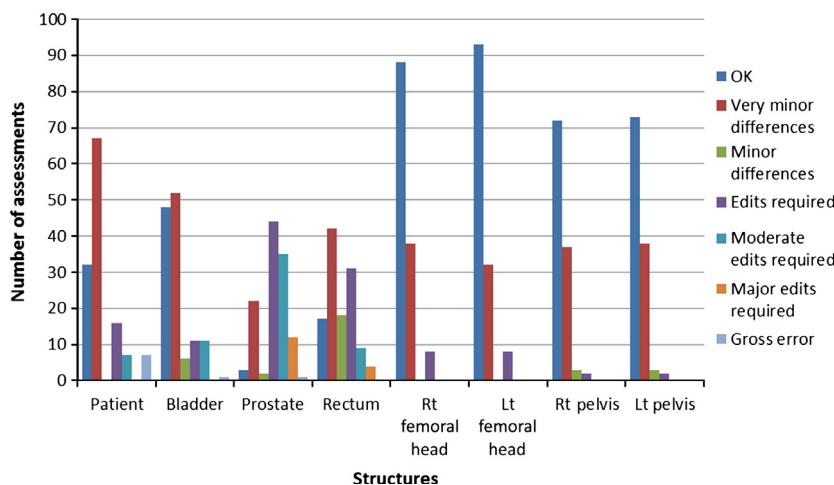


Figure 2. Qualitative evaluation: subjective assessment of contours/volumes generated by ABAS (categories 1–7). Number of occasions each criterion met for each structure and category is shown. ABAS, atlas-based auto-segmentation; rt, right; lt, left.

Table 3. Correlation between ABAS-generated volumes and ‘as treated’ reference volumes for a range of stapled atlases.

Structures	Demo	Number of stapled atlas					Pooled
		2	4	6	8	10	
Patient contour	0.9995 ¹ ***	0.8368***	0.8461***	0.6155**	0.8511***	0.7952*	0.8426***
Bladder	0.9932***	0.9535***	0.9882***	0.9767***	0.9852***	0.9922***	0.9780***
Prostate	0.9994***	0.4016*	0.4007*	0.4795*	0.1098	0.6748	0.2000*
Rectum	0.9686*	0.6843***	0.7385***	0.5458**	0.5273*	0.5613	0.7881***
Right femoral head	0.8531	0.3204	0.2760	0.8265***	0.3865	−0.1733	0.3342***
Left femoral head	0.9885*	0.5774***	0.5603***	0.8999***	0.5975**	0.1095	0.6559***

¹Correlation coefficient, *r*.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

were noted when the demonstration atlas provided by the vendor was used, correlations between the ‘as treated’ reference and the ABAS-generated volumes were generally significant. The bladder and patient contour proved to be the most reliable contours based on the volumetric assessment; correlations between ABAS and the ‘as treated’ volumes were highly significant, correlation coefficients for the bladder exceeded 0.95 for all the atlas staples ($P < 0.001$). The bladder and patient contour are the two largest structures in this region and small variations in volume are less significant than these same differences would be for a smaller structure. The volumetric assessment also revealed some limitations of ABAS. The femoral heads were the most clinically acceptable contour in the qualitative analysis, but occasionally showed poor correlations with the reference in the volumetric analyses. The poor correlations for some atlases were attributed to the fact that ABAS contoured inferiorly to the lowest extent of the data set, whereas the ‘as treated’ contours were cut short at the

level of the lower trochanter, as per clinical protocol. The prostate presented a different set of challenges. The difficulties in contouring the prostate are widely recognised and have been discussed by others.^{11,12} The ABAS-contoured prostate was typically smaller than the reference contour and clinical target volume (CTV) was not clinically acceptable (as a general rule). The ability of ABAS to adequately delineate the prostate was compromised in part due to the fact that the radiation oncologist standardly defined the target based on registered MRI images which ABAS was unable to use. Based on our data, ABAS-generated volumes were on average of 25% lower than manually drawn ones based on MRI images.

In theory, it is possible that a single atlas is all that is required to generate clinically acceptable contours for any patient if the appropriate atlas is selected for each case. In clinical practice, however, the male pelvic anatomy can vary quite dramatically due to factors such as bladder and rectal filling,³ thus potentially requiring numerous atlases

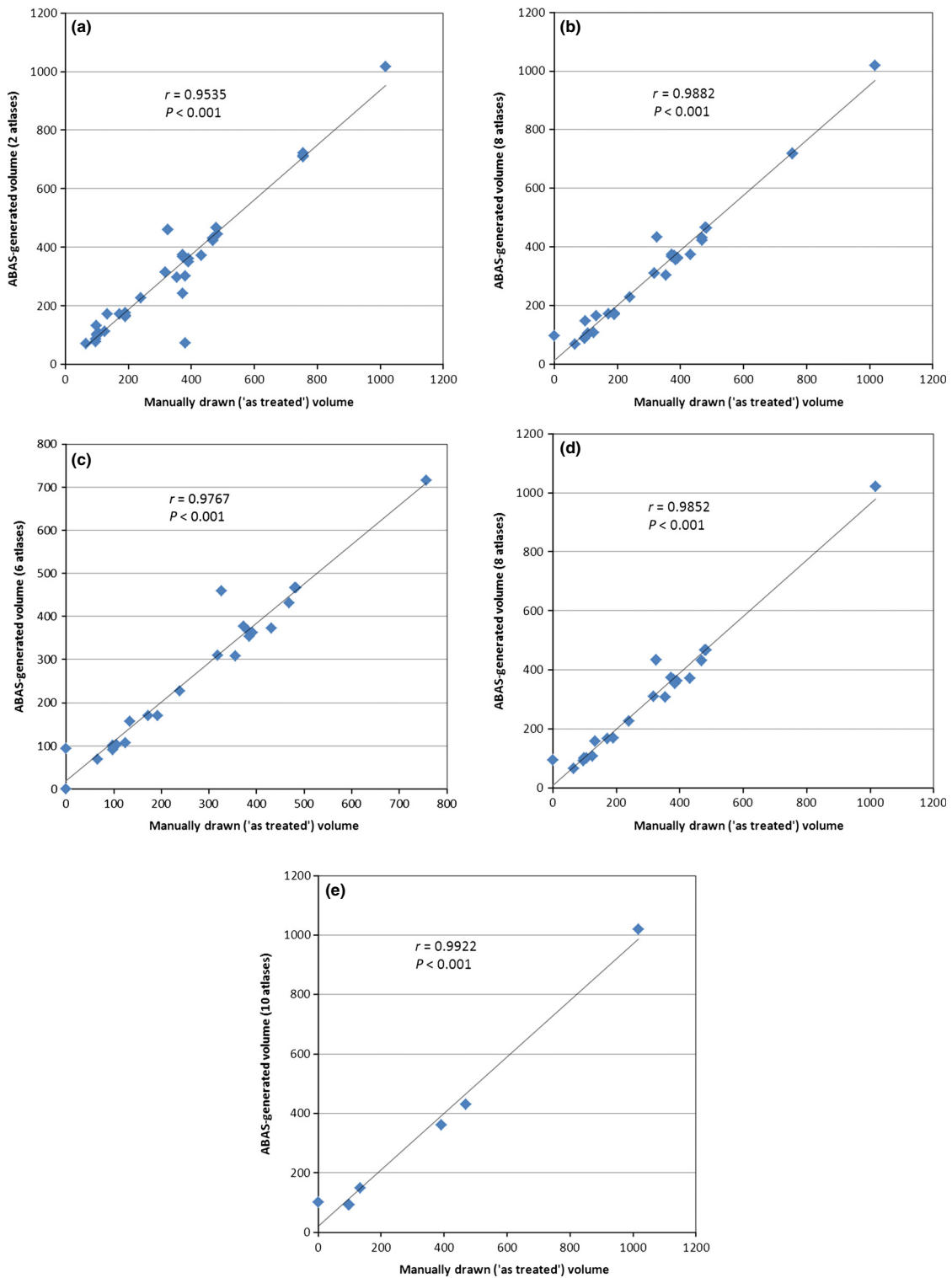


Figure 3. Representation of the correlation between the manually drawn ('as treated') reference and ABAS-generated volumes for the bladder for 2 (A), 4 (B), 6 (C), 8 (D) and 10 stapled atlases (E). The correlations were highly significant ($P < 0.001$).

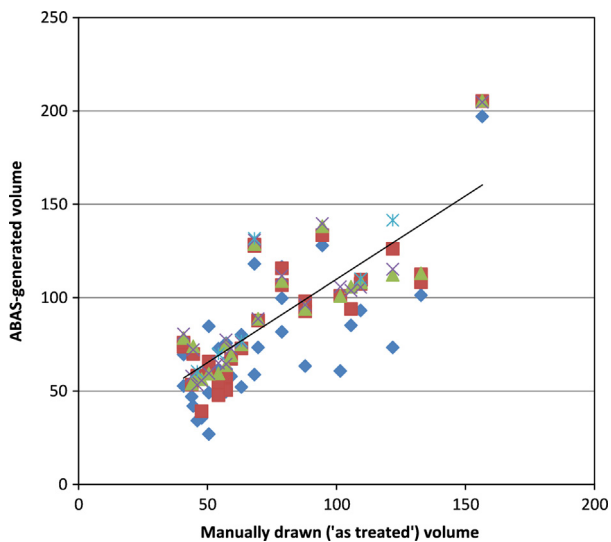


Figure 4. Representation of the correlation between manually drawn ('as treated') reference and ABAS-generated volumes for 2 (red square), 4 (blue diamond), 6 (green triangle), 8 (purple cross), 10 (blue cross) stapled atlases for the rectum. The trendline for four atlas staples is shown. ABAS, atlas-based auto-segmentation.

to cover the range of natural variation found in patients. It then becomes a challenge to select the most appropriate atlas for a particular patient. Investigation is therefore justified into the use of multiple atlases which may allow the software (ABAS) to accommodate a wider range of anatomical variation and select features from any atlas that best match the case in question and provide for a better fit with the patient. In practice, selection of appropriate atlases is not straight forward, nor is the optimum number of atlases required for the best fit. One study suggested that a data set of 15 patients may be required.¹¹ Further research is required in this respect.

Conclusion

This analysis of ABAS has provided a logical examination of the capability and performance of this software to produce contours of anatomical structures in the male pelvic region. Findings from this study suggest that ABAS has the potential to produce clinically acceptable volumes that will lead to a reduction in planning workload. In this respect, our study is in agreement with other recent studies involving the male pelvis. Manual review and refinement of contours is presently required with this software, however, as part of a structured approach to contouring, ABAS provides a good foundation on which to achieve more consistent contouring while improving efficiency over the fully manual process. As the process for ABAS has been developed and fine-tuned it has

become an established part of the workflow for all male pelvic patients receiving radiotherapy at our cancer institute.

Acknowledgements

Elekta Inc., for the provision of a Beta version of the ABAS software.

Conflict of Interest

The authors declare no conflict of interest.

References

1. eviQCancer Treatments Online a service of cancer institute NSW. Urogenital prostate protocols [updated 27 July 2014; cited March 2013]. Available from: <https://www.eviq.org.au/Protocol/tabid/66/id/235/view/FullView/Radiation+Oncology%2c+Prostate%2c+High+Risk%2c+EBRT%2c+Definitive.aspx>.
2. Voet PW, Dirkx ML, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis *Radiother Oncol* 2011; **98**: 373–7.
3. Huyskens DP, Maingon P, Vanuytsel L, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiother Oncol* 2009; **90**: 337–45.
4. Speight R, Sykes J, Lindsay R, Franks K, Thwaites R. The evaluation of a deformable image registration segmentation technique for semi-automating internal target volume (ITV) production from 4DCT images of lung stereotactic body radiotherapy (SBRT) patients. *Radiother Oncol* 2011; **98**: 277–83.
5. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004; **23**: 903–21.
6. Hoogeman M, Han X, Teguh D, et al. Atlas-based auto-segmentation of CT images in head and neck cancer; what is the best approach? Proceedings of the 50th Annual ASTRO meeting, ASTRO, Boston, MA, 21–25 September 2008.
7. Levendag PC, Al-Mamgani A, Teguh DN, Voet P, Heijmen B, Han X, Hoogeman M. Clinical validation of atlas-based auto-contours of target tissues and critical normal tissue structures in the head and neck. Proceedings of the 51st Annual ASTRO meeting, ASTRO, Chicago, IL, 1–5 November 2009.
8. Isambert A, Dhermain F, Bidault F, et al. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiother Oncol* 2008; **87**: 93–9.

9. Skala M, Berry M, Duchesne G, et al. Australian and New Zealand three-dimensional conformal radiation therapy consensus guidelines for prostate cancer. *Australas Radiol* 2004; **48**: 493–501.
10. Teguh DN, Levendag PC, Voet PW, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys* 2011; **81**: 95–7.
11. Hwee J, Louie AV, Gaede S, et al. Technology assessment of automated atlas based segmentation in prostate bed contouring. *Radiat Oncol* 2011; **6**: 110. doi: 10.1186/1748-717X-6-110
12. Gambacorta MA, Valentini C, Dinapoli N, et al. Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumors using auto-segmentation computed system. *Acta Oncol* 2013; **52**: 1676–81. doi: 10.3109/0284186X.2012.754989