Improved radiograph measurement inter-observer reliability by use of statistical shape models

Elise C. Pegg[1], Stephen J Mellon[1], Gabriella Salmon[1], Abtin Alvand[1], Hemant G Pandit[1], David W. Murray[1], Harinderjit S. Gill[1]

[1] University of Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7LD, UK.

Abstract

Pre- and post-operative radiographs of patients undergoing joint arthroplasty are often examined for a variety of purposes including preoperative planning and patient assessment. This work examines the feasibility of using active shape models (ASM) to semi-automate measurements from post-operative radiographs for the specific case of the Oxford™ Unicompartmental Knee. Measurements of the proximal tibia and the position of the tibial tray were made using the ASM model and manually. Data were obtained by four observers and one observer took four sets of measurements to allow assessment of the inter- and intra-observer reliability, respectively. The parameters measured were the tibial tray angle, the tray overhang, the tray size, the sagittal cut position, the resection level and the tibial width. Results demonstrated improved reliability (average of 27% and 11.2% increase for intra- and inter-reliability, respectively) and equivalent accuracy ($p>0.05$ for compared data values) for all of the measurements using the ASM model, with the exception of the tray overhang ($p=0.0001$). Less time (15 s) was required to take measurements using the ASM model compared with manual measurements, which was significant. These encouraging results indicate that semi-automated measurement techniques could improve the reliability of radiographic measurements.

Keywords: Radiograph; Measurement; Shape; Unicompartmental; Knee

Introduction

Throughout the field of orthopaedics, measurements taken from radiographs play an important role in patient assessment. With regards to lower limb arthroplasty, pre-operative measurements are essential to accurately plan the optimal component size and position. Post-operative measurements are often used for surveillance; either to monitor the patient recovery, surgical performance or for research. Clinical decisions are made based upon these measured parameters [1, 2] and it is therefore important that these data are correct and reliable.

Radiographic measurements of distances or angles are often made manually; these can be taken using a pencil, or more commonly in recent times with digital image processing software (such as Picture Archiving and Communications System, PACS [3]). With both these methods, anatomic points need to be selected by the user and requires clinical knowledge; this process can be subjective and therefore result in error. Manual measurements can also be time consuming. Another technique often used for pre-operative planning is templating. Often the orthopaedic component manufacturer will provide a series of templates for each size component, these can be overlaid on the radiograph to help decide which size to implant; as components are discrete sizes, this is therefore a discrete measurement method [4].

With the advent of digital radiographs, two main methods have been used to attempt to automate assessments of radiographs. The first is edge detection, where edges are defined by an abrupt change in brightness. This method works well provided it is used for the application it is written; this is not a particularly versatile technique and does not cope well with images of varying content or brightness. Nevertheless, this method has been used successfully to measure knee joint space [5, 6]. The second method is statistical shape

modelling; this is where an average shape is defined by manually selected landmarks from a collection of standard images. Once the information is gathered, this model can then be applied to numerous new images to search for the shape [7]. This technique has been primarily used as a segmentation tool for medical images or to assess natural variations in anatomical shapes rather than for measurements. However, some studies have successfully used this technique for measurements of both vertebral [8, 9] and hip radiographs [10-13], and have shown the measurement reliability to be equivalent to manual measurements [8]. This is referred to as ASM (Active Shape Modelling).

Due to the semi-automatic nature of measurement using ASMs, the subjectivity of the user is minimised and also the necessity for clinical knowledge is removed. The current study investigates the hypothesis that this could result in an improvement in measurement reliability while maintaining accuracy and reducing measurement time. The particular case of unicompartmental knee arthroplasty (UKA) was chosen; parameters were measured from the post-operative radiographs using both the ASM and manually, and the intra- and inter-observer reliability was assessed.

The current study investigates the feasibility of the use of active shape models to measure the positioning of implanted components after unicompartmental knee arthroplasty (UKA) from post-operative radiographs. We tested the null hypothesis that there is no difference in the measured values of radiographic parameters using an ASM based measurement compared to manual measurement. Specific parameters were measured using both the ASM and manual measurements, and the intra- and inter-observer reliability was assessed.

Materials and Methods

The ASM model was created using custom written routines within MATLAB (Version 7.10, MathWorks Inc., Massachusetts, USA); the model was based upon the technique reported by Cootes et al. [7]. First, the ASM model was trained using a set of 36 standard anterior-posterior (AP) radiographs. The model was then applied to a different set of 19 radiographs, which ranged in quality (from 9.5 to 0.9 pixels per mm), to find the shape and location of the proximal tibia and the implanted tibial tray within the images. Finally, calculations were performed on the recorded shapes to provide the measured parameters and these data were compared with manual measurements.

Training the model

Thirty-six post-operative radiographs of patients who had undergone UKA were chosen to train the model. Each radiograph was flipped or inverted if necessary, to ensure that the image represented a right knee, with the implanted component appearing white. Sixty-four points surrounding the tibia were manually selected in landmark locations (Figure 1a); 20 points were interpolated between each of the landmarks to give a total of 1,198 co-ordinates. The same process was performed for the tibial tray shape model using 53 landmark points (Figure 1b), which were interpolated to give a total of 989 co-ordinates. The shapes were aligned using the Procrustes method [14]; translation was removed by positioning the shape so that the centroid was at the origin, rotation was removed using the mean angle to the centroid as reference; scaling was removed by normalising using the width of the tibia; due to varied length of the tibial radiographs, the overall shape could not be scaled.

Equation 1: $$Translation = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Where x is the list of co-ordinates for a certain axis, N is the number of co-ordinates.

Equation 2: $$Rotation = \frac{1}{N}\sum_{i=1}^{N} \tan^{-1}\frac{y_i}{x_i}$$

Where x and y are the co-ordinates and N is the number of points.

Equation 3: $$Scale = \frac{1}{TibialWidth}$$

Principal component analysis (PCA) was then performed on the 36 sets of co-ordinates, and the pixel profiles perpendicular to each point (12 pixels long) and greyscale differences in the pixel profiles (normalised by the average profile).

Application of the model

The individual assessing the radiograph positioned the average shape (from the training data of either tibia or tibial tray) over the image and if necessary the average shape could stretched in x- or y-directions or rotated to enable a rough fit. Once satisfied, a starting position was chosen for the model to run. Both the proximal tibia and tibial tray ASMs performed 40 iterations to find the final shape. At each iteration, the current pixel profiles were calculated and the points moved to a location which minimised the Mahalanobis distance (maximum movement of 6 pixels). The Mahalanobis distance is a way of describing how far a parameter is away from the desired value, whilst accounting for correlation and variance in the data [15]. The PCA of the co-ordinates, which was performed during the training of the model, provided a description of how the landmarks move together as the shape varies; this

information was stored as eigenvalues and eigenvectors. Using these, the movement of the points were limited to ensure that they remained within the shape variability defined. After this was complete the next iteration would be performed. This process is well described by Cootes *et al*.[7].

Calibration

Once the locations of the tibia and the tibial tray were known within the radiograph, the approximate location of the femoral component was inferred. The image was cropped to this region, and then transformed to a black and white image using a threshold definition. The threshold was set at 0.73 (determined empirically) for an image with greyscale values ranging from 0 to 1. The outline of the femoral component was then found and a circle fitted to regions around the femoral component using a least-squares method. Each region was the same length (60% of the femoral width) and at started at set positions (intervals of 16^{th} of the total perimeter) around the outline. The circle fitted with the smallest residuals was chosen as the femoral diameter and the user verified this visually. The implanted femoral component sizes were known for each subject and for each radiograph and these, therefore, enabled calibration of the pixel to millimeter relationship. For this study, calibration using the femoral component was possible due to the spherical design of the Oxford™ UKA (Biomet UK Ltd., Bridgend, UK); however, other implants may require a calibration marker.

Measurements

Prior to calculating the measurements, the mechanical axis of the tibia was determined and the co-ordinates rotated so that the mechanical axis was in line with the image y-axis. The mechanical axis was defined as line between; the two tibial spines, the most medial and

lateral points on the tibia and the furthest distal medial and lateral points. Six measurements were taken (Figure 2); (1) the angle of the tray with respect to a line perpendicular to the mechanical axis (positive if the tray wall was more superior than the edge, negative if not), (2) the degree of tray overhang, defined as the distance from the outermost point of the tibial tray to the tibia underneath (if the tray protruded over the tibia this was a positive value, negative if not), (3) the width of the tibial tray, (4) the distance from the sagittal cut to the mechanical axis, (5) the resection level, from the tip of the tibial spines to the base of the tray, (6) the width of the tibia, defined as the distance between the most medial and lateral points of the tibia.

These parameters were chosen for a variety of reasons. The risk of periprosthetic fracture of the tibia has been shown to be related to the sagittal saw cut [16, 17] therefore the position of this was assessed. The resection level was also measured because there are some indications that this may increase stresses within tibia [18, 19]. The degree of tray overhang has been shown to influence clinical outcome after UKA [20] and therefore was measured. The angle of the tray relates indirectly to the angular alignment, and the width of the tibial tray when compared with the implanted size can be an indication of radiographic alignment.

Model Validation

In order to validate the model, both the reliability and accuracy of the measurements was compared to manual measurements. Manual measurements were taken using the same point definitions as described for the ASM; these were performed using Image J software (National Institutes of Health, Bethesda, Maryland, USA). Calibration for the manual measurements was performed in the same way as the ASM model; however, the circle fit was based on 10

points surrounding the femoral component selected by the user rather than the larger number used by the edge detection method. Measurements were taken, using both the ASM and the manual method, by four observers (one clinician, two engineers and one novice); one observer (engineer) took four sets of measurements. The time spent making the measurements was recorded for one observer using a stop watch.

Statistical Analysis

The intra-class correlation coefficient (ICC) was calculated for each dataset to assess the reliability; a two-way mixed model was used with statistical significance set at 95%. Single measures were reported for the intra-observer calculations and average measures for the inter-observer calculations. The average ICC values for different methods were compared using a Mann-Whitney (non-parametric) test for significance. Bland-Altman plots [21] and paired t-tests were used to examine the correlation between the measurement values using the two measurement methods; this was examined using the inter-observer data. To compare the time taken using each measurement technique a Mann-Whitney test was performed. Values were defined as statistically significant for all tests if $p<0.05$ (*). All statistical analyses were performed using PASW Statistics (version 18.0.0, SPSS Inc, Chicago, USA).

Results

The intra-observer reliability ICC results were on average 27% higher (range 4.3% to 65.2%) for all the ASM datasets compared with the manual data (Figure 3, Table 2) and the difference was statistically significant (Mann-Whitney t-test, $p=0.018$). The confidence intervals (CI) were smaller for all ASM measurements compared with the manual measurements; the average ASM CI was 50% that of the manual CI (range 8.5% to 90.3%).

Out of the ASM measurements, the degree of tray overhang had the lowest ICC value at 0.725; the size of the tray had the lowest ICC value for the manual measurements, this was 0.494. The calibration value (pixel size) was the most reliable parameter measured by both methods (0.988 for ASM, 0.855 for manual).

The inter-observer reliability ICC results (Figure 4) showed similar trends to the intra-observer results, but there were some important differences. The inter-observer ASM measurements for the degree of tray overhang were worse overall compared with the manual measurements, both in terms of the ICC value (0.81 ASM, 0.9 manual) and the confidence interval range (0.232 ASM, 0.125 manual). All other parameters had significantly (Mann-Whitney t-test, $p=0.037$) greater ICC values compared with the manual measurements (average 11.2% increase, range 0.2% to 25.1%). As before, the ICC for the manual measurement of tray size was relatively low (0.783), but the lowest ICC value measured was for the tray angle (0.738); both of these were considerably higher than the lowest intra-reliability ICC data (0.494).

The results from the paired t-tests for each (Table 1) demonstrated good correlation between the measurements values from the two methods, with the exception of the degree of tray overhang. This showed a very significant difference (Paired t-test, $p=0.0001$) between the two measurement methods. The ASM model underestimated the degree of tray overhang by 1.49 mm on average, compared with the manual measurements. The Bland-Altman plot (Figure 5) of the tray overhang supported this result, illustrating a greater number of points below the zero line on the y-axis. The plots for all the other parameters were evenly spread around zero, demonstrating a good correlation between the measurement methods.

On average, the shape model technique took 116 seconds to take the measurements (n=38, standard deviation 14.8 s), the manual measurements took 131 seconds on average (n=59, standard deviation 17.5 s). This was a difference of 15 seconds and was significant (Mann-Whitney, p=0.001) (Figure 6).

Discussion

This study demonstrates that the ASM model provides a feasible and accurate method of assessing implant positioning for the UKA tibial component on post-operative radiographs. The reliability analysis for both the intra- and inter-observer tests showed better ICC values for the ASM measurements compared with the manual measurements, except for the measurement of tray overhang.

The tray overhang measurements using the ASM were not only less reliable compared with the manual measurements, the measured values were consistently lower (Figure 5) by approximately 1.5 mm. The manual measurements are assumed to be accurate as this is the current gold standard; it therefore appears that the model has a systematic error in this region. One possibility is that the error is purely due to the scale of the parameter; this is the smallest value measured and it could be that the model fit has a lower limit for measurement. Another possibility is that the model itself is at fault; if so this could be either due to the tray or the tibial fit. The shape of the tray is relatively consistent due to the manufactured shape, whereas variation in tibial shape is far greater; therefore the tibial model is the most likely to be at error. Increasing the number of training radiographs may improve the model so that it better represents the anatomical variation in tibia shape; this work is currently underway.

All other parameters showed an improvement in reliability of measurement using the shape model compared with the manual technique (Figures 3 and 4) and the confidence interval ranges for the ICC values were also smaller for the shape model on average. It was important to check the reliability of the calibration technique, because error in this value would affect the other data. The pixel calibration was the most reliable parameter for both techniques; this is therefore thought to be a minimal cause of error, with measurement error being the primary cause. The improvement in reliability observed could prove particularly useful for analysis of serial radiographs from the same patient over time, for situations such as component loosening.

Previous studies have shown the reliability of shape models to be equivalent to manual measurements [8], but an improvement in reliability of radiograph assessment by the use of a shape model has not been reported before. It is thought that this improvement in reliability is due to the increased level of automation when using shape models, because this reduces the influence of the user. However, it is important to ensure that the improvement in reliability is not at the cost of the accuracy of the measurements.

The average measurement values of from both techniques for all parameters, except for the tray overhang, correlated well (Table 1) and no significant differences between the data were found. In addition, the Bland-Altman plots (Figure 5) all centre around the zero line supporting this. It is also worth noting that the quality of the radiographs which were assessed varied; many were not in line with the tibial tray and had different resolutions and

contrast (Figure 7). Therefore, we believe this validates the accuracy of the ASM measurements.

It is important to determine the amount of error which is acceptable for each parameter. The purpose of this study was to address the null hypothesis that there is no difference between manual radiograph measurements and those using an ASM method. The standard error (SEM) results show the ASM method to have equivalent error for all parameters with the exception of tray overhang (Table 1). However, this assumes that the manual measurement error is acceptable, which is relatively undetermined. The only parameter for which clinically relevant values are known is the degree of tray overhang, which has been reported to affect clinical outcome when it is >3 mm [20]. The SEM for the manual measurement is 0.1 mm (Table 1) therefore is sufficient to detect the parameter. For the other parameters, if any future conclusions are drawn then the values will need to take into account the error.

As well as decreasing measurement subjectivity and the required expertise, another of the main aims of creating a semi-automated measurement technique was to decrease the labour involved in measurement. For this reason, the time spent measuring using each technique was recorded for one observer. The results did show a significant reduction in time taken using the shape model, but only by 15 seconds. However, it is worth considering the number of measurements which were taken. With the shape model the time taken to assess the radiograph is not dependent on the number of measurements; because once the shape of the tibia and tibial tray are known, any number of measurements can be subsequently made. However, the time taken for manual measurements is dependent on the number of

parameters. Therefore, the shape model will result in the greatest time reduction compared to manual measures when large numbers of parameters are examined.

It is thought that the main time during the ASM application is not spent on measurement, but on using the graphical user interface (i.e. opening files, clicking through menus, deciding where to position the average shape); this is not an issue for the manual measurements. Therefore it is possible that some time improvements could be made through adapting the user interface.

The use of active shape models to take measurements from radiographs should be treated with caution. It is important to validate each model to ensure that the accuracy and reliability of measurements are sufficient for the specific application it is to be used. This study is therefore limited to this one type of radiograph and the specific measurements chosen. Another important factor is the image resolution, the measurement accuracy can only be as accurate as the image resolution which can limit data; in this study the images varied in resolution and this may have affected the resulting data; however, the scale of the measurement parameters chosen were greater than the resolution in all cases and is therefore not thought to have affected the results significantly. The relatively small size of the training data set (36) may have affected the tray overhang measurements and ideally a greater number would have been used.

Conclusions

This study has shown the null hypothesis, that there is no difference in the measured values of radiographic parameters using an ASM based measurement compared to manual

measurements, to be correct for almost all parameters tested, with the exception of the degree of tray overhang. With further refinement and training of the model it is hoped that measurement of this parameter can be improved. The ASM model improved the reliability and reduced measurement time for almost all the parameters measured for the specific case of the Oxford UKA, while maintaining the accuracy of the measurement. Provided the limitations of the model are observed, this technique has the potential to be a useful tool in routine radiographic assessment.

Acknowledgements

The authors would like to thank Mrs B. Marks (Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, UK) for her assistance with this study. The study was supported by the NIHR Biomedical Research Unit into Musculoskeletal Disease, Nuffield Orthopaedic Centre and the University of Oxford. Some of the authors have received funding from a commercial party, but this was unrelated to the present study.

References

[1] Cowell HR. Radiographic measurements and clinical decisions. J Bone Joint Surg Am 1990;72(3):319.

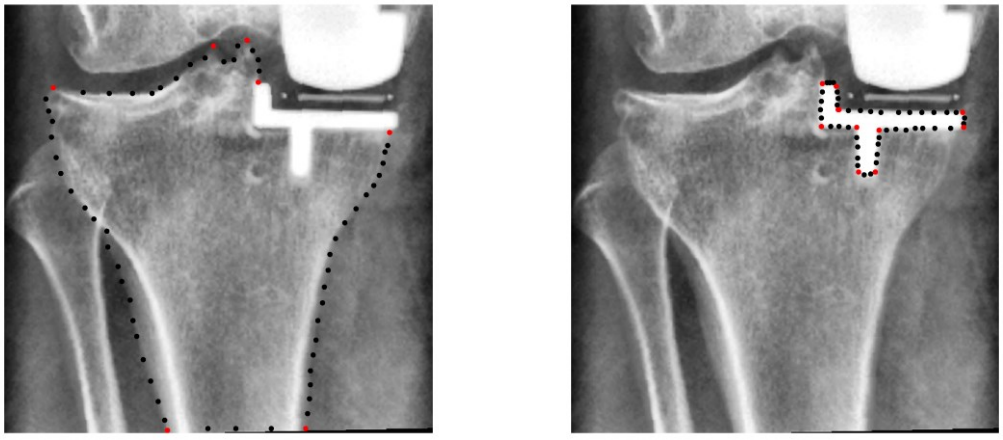
[2] Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. J Bone Joint Surg Br 1992;74-B(2):287-91.

[3] Segev E, Hemo Y, Wientroub S, et al. Intra- and interobserver reliability analysis of digital radiographic measurements for pediatric orthopedic parameters using a novel PACS integrated computer software program. Journal of Children's Orthopaedics 2010;4(4):331-41.

[4] Viceconti M, Lattanzi R, Antonietti B, et al. CT-based surgical planning software improves the accuracy of total hip replacement preoperative planning. Medical Engineering & Physics 2003;25(5):371-7.

[5] Lynch JA, Buckland-Wright JC, Macfarlane DG. Precision of joint space width measurement in knee osteoarthritis from digital image analysis of high definition macroradiographs. Osteoarthritis Cartilage 1993;1(4):209-18.
[6] Dacre JE, Huskisson EC. The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. Rheumatology 1989;28(6):506-10.
[7] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. Computer Vision and Image Understanding 1995;61(1):38-59.
[8] Smyth PP, Taylor CJ, Adams JE. Vertebral Shape: Automatic Measurement with Active Shape Models. Radiology 1999;211(2):571-8.
[9] Roberts MG, Cootes TF, Adams JE. Vertebral shape: automatic measurement with dynamically sequenced active appearance models. In. Proceedings of the 8th international conference on Medical image computing and computer-assisted intervention. Palm Springs, CA: Springer-Verlag, 2005: 733-40
[10] Gregory J, Testi D, Stewart A, Undrill P, Reid D, Aspden R. A method for assessment of the shape of the proximal femur and its relationship to osteoporotic hip fracture. Osteoporosis International 2004;15(1):5-11.
[11] Gregory JS, Waarsing JH, Day J, et al. Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features: Can hip shape tell us anything about the progression of osteoarthritis? Arthritis Rheum 2007;56(11):3634-43.
[12] Behiels G, Vandermeulen D, Maes F, Suetens P, Dewaele P. Active Shape Model-Based Segmentation of Digital X-ray Images. In: Taylor C, Colchester A, editors. Lecture Notes in Computer Science. Berlin Heidelberg New York: Springer, 1999: 128-37
[13] Boukala N, Favier E, Laget B, Radeva P. Active shape model based segmentation of bone structures in hip radiographs. In. Industrial Technology, 2004: 1682-7
[14] Goodall C. Procrustes Methods in the Statistical Analysis of Shape. Journal of the Royal Statistical Society Series B (Methodological) 1991;53(2):285-339.
[15] De Maesschalck R, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems 2000;50(1):1-18.
[16] Seeger J, Haas D, Jäger S, Röhner E, Tohtz S, Clarius M. Extended sagittal saw cut significantly reduces fracture load in cementless unicompartmental knee arthroplasty compared to cemented tibia plateaus: an experimental cadaver study. Knee Surgery, Sports Traumatology, Arthroscopy 2011:1-5.
[17] Clarius M, Haas D, Aldinger PR, Jaeger S, Jakubowitz E, Seeger JB. Periprosthetic tibial fractures in unicompartmental knee arthroplasty as a function of extended sagittal saw cuts: An experimental study. The Knee 2010;17(1):57-60.
[18] Simpson DJ, Price AJ, Gulati A, Murray DW, Gill HS. Elevated proximal tibial strains following unicompartmental knee replacement—A possible cause of pain. Medical Engineering & Physics 2009;31(7):752-7.
[19] Whiteside L, A. Making Your Next Unicompartmental Knee Arthroplasty Last: Three Keys to Success. The Journal of Arthroplasty 2005;20, Supplement 2(0):2-3.
[20] Chau R, Gulati A, Pandit H, et al. Tibial component overhang following unicompartmental knee replacement—Does it matter? The Knee 2009;16(5):310-3.
[21] Altman DG, Bland JM. Measurement in Medicine: The Analysis of Method Comparison Studies. Journal of the Royal Statistical Society Series D (The Statistician) 1983;32(3):307-17.

Figure Legends



(a) (b)

Figure 1. Position of landmark points selected surrounding the tibia (a) and the tibial tray (b) on proximal post-operative UKA radiographs. Red markers represent fixed anatomical landmarks and black dots represent perimeter points in-between.

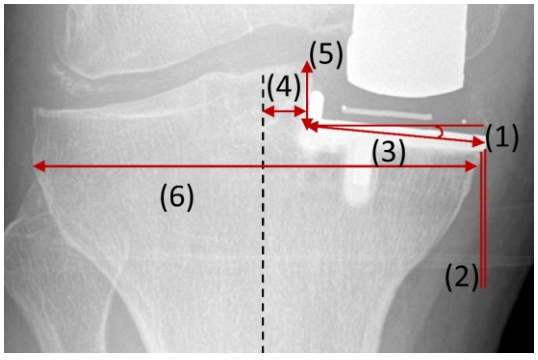


Figure 2. Parameters measured from post-operative UKA radiographs, (1) the tray angle, (2) the degree of tray overhang (3) size of the tray, (4) the sagittal cut position, (5) the resection level and (6) the width of the tibia.

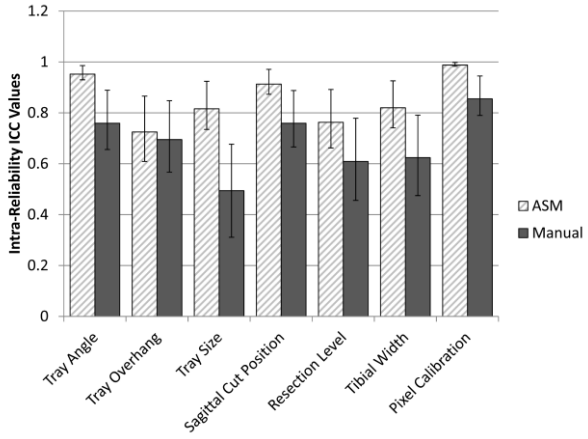


Figure 3. Intra-observer Intra-class correlation (ICC) values for each of the measurements taken using the manual and the shape model method. Error bars represent 95% confidence intervals.

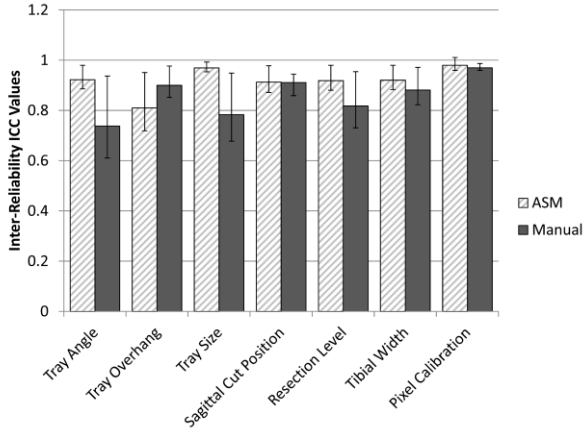


Figure 4. Inter- observer Intra-class correlation (ICC) values for each of the measurements taken using the manual and the shape model method. Error bars represent 95% confidence intervals.

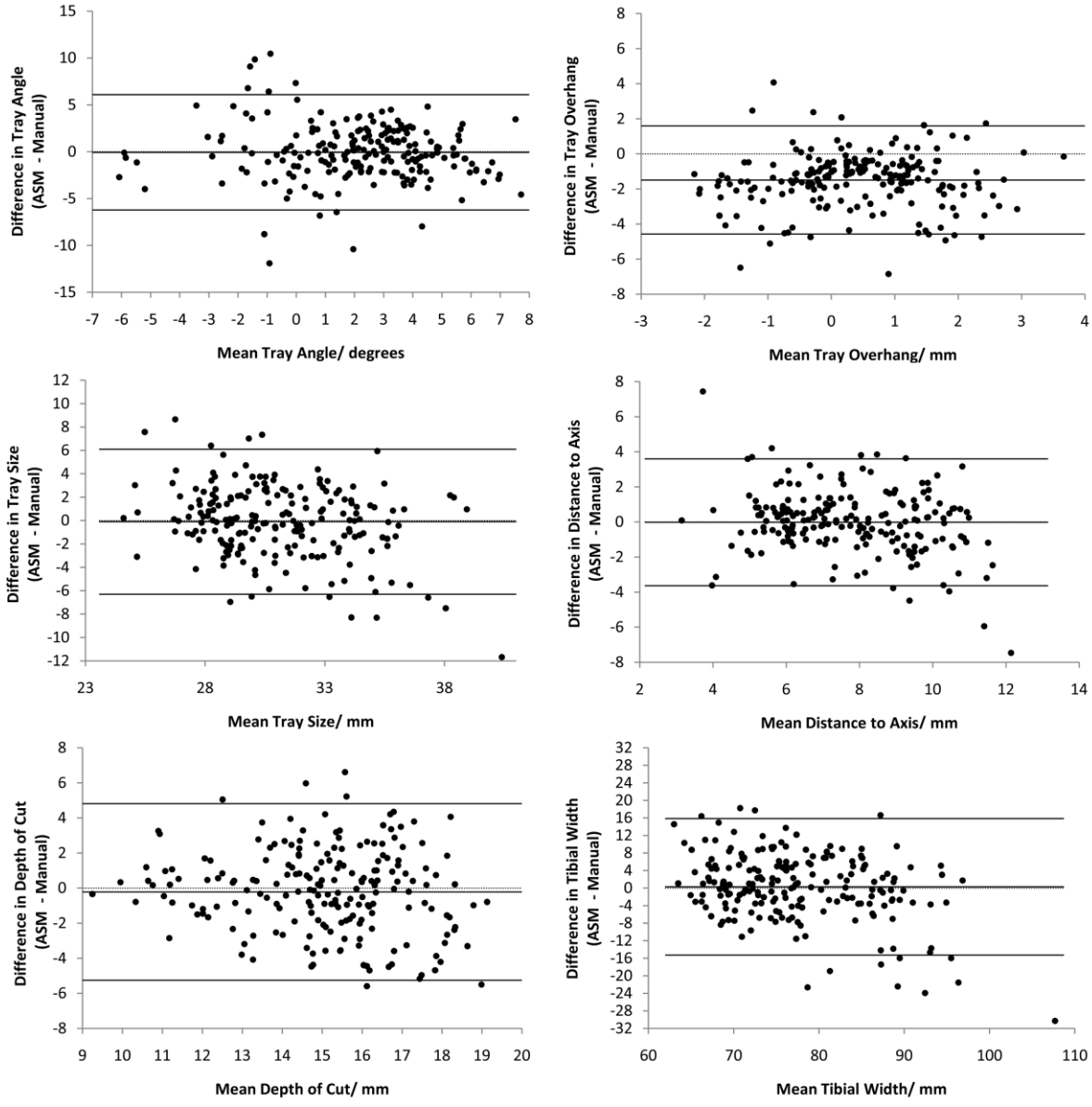


Figure 5. Bland-Altman plots of the different measured parameters comparing the shape model results to the manual measurements. The dotted line represents y=0; the upper solid line represents the mean difference + 2 standard deviations, the lower solid line represents the mean difference -2 standard deviations and the middle solid line represents the mean difference value.

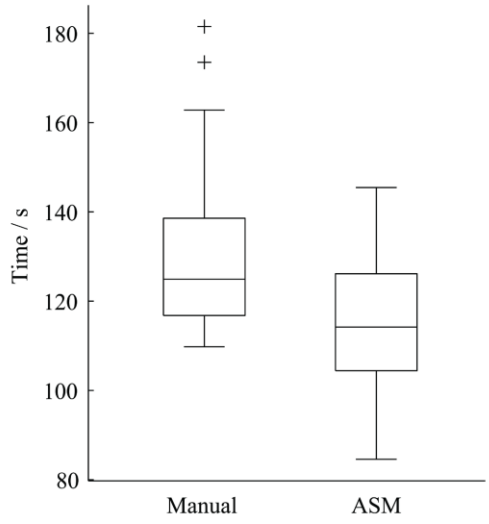


Figure 6. Box plot illustrating change in time (minutes) taken to perform measurements manually and with the shape model (ASM).

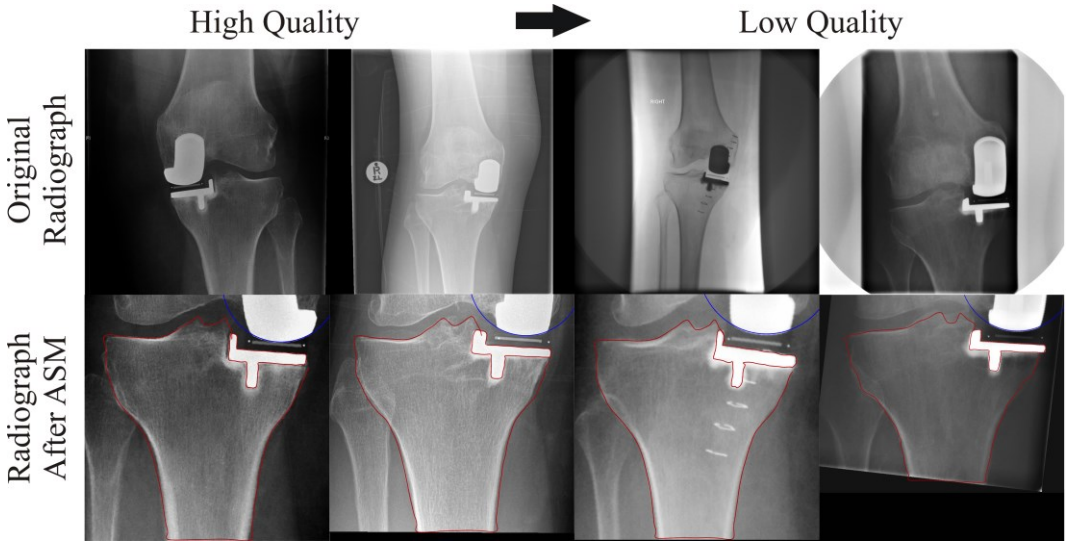


Figure 7. Illustration of four radiographs ranging in quality; the original radiograph is shown and the same radiograph after flipping/ inverting (if necessary) then fitting of the ASM model (red lines) and alignment with the mechanical axis. The circle fit to the femoral component is also illustrated (blue lines).

Tables

		ASM			Manual		
	Parameter	ICC	Upper CI	Lower CI	ICC	Upper CI	Lower CI
Intra-Reliability	Tray Angle	0.952	0.918	0.974	0.759	0.629	0.862
	Tray Overhang	0.725	0.584	0.841	0.695	0.542	0.823
	Tray Size	0.816	0.708	0.897	0.494	0.311	0.677
	Sagittal Cut Position	0.913	0.855	0.953	0.759	0.630	0.852
	Resection Level	0.763	0.634	0.864	0.609	0.439	0.762
	Tibial Width	0.820	0.714	0.899	0.624	0.457	0.773
	Pixel Calibration	0.988	0.978	0.993	0.855	0.765	0.920
Inter-Reliability	Tray Angle	0.923	0.866	0.960	0.738	0.539	0.865
	Tray Overhang	0.810	0.669	0.901	0.900	0.823	0.948
	Tray Size	0.969	0.945	0.984	0.783	0.617	0.888
	Sagittal Cut Position	0.913	0.848	0.954	0.911	0.877	0.963
	Resection Level	0.919	0.858	0.957	0.818	0.681	0.905
	Tibial Width	0.921	0.862	0.959	0.881	0.790	0.939
	Pixel Calibration	0.980	0.927	0.978	0.970	0.961	0.988

Table 1. Measurement inter-observer data summary for each parameter using the two methods. Standard error in the mean (SEM, n=205) reported and the difference between the data (Manual-ASM). Paired t-test results comparing the shape model measurements of the parameters to the manual measurements. Results deemed statistically different if $p<0.05$ (*).

Measured Parameter	Manual Data		ASM Data		Difference	Paired t-test Results		
	Mean	SEM	Mean	SEM		p	Upper 95% CI	Lower 95% CI
Tray Angle / °	2.1822	0.2203	2.0865	0.2033	0.09569	0.6590	-0.33161	0.52298
Tray Overhang / mm	1.1588	0.0944	-0.3314	0.0978	1.49021	0.0001*	1.27866	1.70176
Size Tray / mm	31.1660	0.2616	31.0493	0.2068	0.1164	0.5900	-0.30899	0.54226
Sagittal Cut Position / mm	7.7215	0.1606	7.7274	0.1320	-0.00588	0.9630	0.12692	-0.25613
Resection Level / mm	15.1917	0.1782	15.0058	0.1533	0.18596	0.2950	0.17701	-0.16304
Tibial Width / mm	77.4562	0.7143	77.5611	0.5740	-0.10498	0.8550	-1.23909	1.02914
Pixel Size / mm	0.1688	0.0047	0.1660	0.0039	0.0028	0.0820	-0.00036	-0.00597

Table 2. Summary of the intraclass correlation coefficient (ICC) values and confidence intervals (CI) for the intra- and inter-reliability analysis of parameters measured with the shape model (ASM) and manually.