



Universiteit
Leiden
The Netherlands

Pinning down loosened prostheses : imaging and planning of percutaneous hip refixation

Malan, D.F.

Citation

Malan, D. F. (2015, October 29). *Pinning down loosened prostheses : imaging and planning of percutaneous hip refixation*. Retrieved from <https://hdl.handle.net/1887/36019>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/36019>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/36019> holds various files of this Leiden University dissertation.

Author: Malan, Daniel Francois

Title: Pinning down loosened prostheses : imaging and planning of percutaneous hip refixation

Issue Date: 2015-10-29

2

Measuring femoral lesions despite CT metal artefacts

Daniel F. Malan, Charl P Botha, Gert Kraaij,
Raoul M.S. Joemai, Huub J.L. van der Heide,
Rob G.H.H. Nelissen, Edward R. Valstar

Adapted from:
Skeletal Radiology,
Volume 41, Issue 5, pp 547–555, May 2012.
doi: 10.1007/s00256-011-1223-2

Abstract

Objective

Computed tomography is the modality of choice for measuring osteolysis but suffers from metal-induced artefacts obscuring periprosthetic tissues. Previous papers on metal artefact reduction (MAR) show qualitative improvements, but their algorithms have not found acceptance for clinical applications. We investigated to what extent metal artefacts interfere with the segmentation of lesions adjacent to a metal femoral implant and whether metal artefact reduction improves the manual segmentation of such lesions.

Materials and methods

We manually created 27 periprosthetic lesions in 10 human cadaver femora. We filled the lesions with a fibrotic interface tissue substitute. Each femur was fitted with a polished tapered cobalt-chrome prosthesis and imaged twice – once with the metal, and once with a substitute resin prosthesis inserted. Metal-affected CTs were processed using standard back-projection as well as projection interpolation (PI) MAR. Two experienced users segmented all lesions and compared segmentation accuracy.

Results

We achieved accurate delineation of periprosthetic lesions in the metal-free images. The presence of a metal implant led us to underestimate lesion volume and introduced geometrical errors in segmentation boundaries. Although PI MAR reduced streak artefacts, it led to greater underestimation of lesion volume and greater geometrical errors than without its application.

Conclusion

CT metal artefacts impair image segmentation. PI MAR can improve subjective image appearance but causes loss of detail and lower image contrast adjacent to prostheses. Our experiments showed that PI MAR is counterproductive for manual segmentation of periprosthetic lesions and should be used with care.

2.1 Introduction

Aseptic loosening caused by osteolysis is one of the foremost problems limiting the survival of hip prostheses [Agarwal, 2004]. Plain radiographs (Fig. 2.1) are the default modality for evaluating osteolysis [Garcia-Cimbrello et al., 2007] but tend to underestimate lesion volume [Agarwal, 2004]. [Claus et al., 2003] even refers to “the lack of any relationship between the two-dimensional lesion size and the actual three-dimensional lesion volume.”

In recent years CT has gained popularity for quantifying periprosthetic osteolysis. There seems to be consensus that CT has superior sensitivity and measurement accuracy for the detection and measurement of osteolysis compared to traditional radiographs [Cahir et al., 2007, Looney et al., 2002, Puri et al., 2002, Schwarz et al., 2003, Walde et al., 2005]. Unfortunately CT images suffer from metal-induced artefacts in the vicinity of metal prostheses [Liu et al., 2009, Watzke and Kalender, 2004]. These most notably arise due to beam hardening and photon starvation [Kalender, 2005, Lee et al., 2007].

Although there still exists no general solution for removing metal-induced artefacts [Hsieh, 2003], several approaches have been offered. [Glover and Pelc, 1981] and [Kalender et al., 1987] first proposed replacing metal sinogram projections with interpolations of adjacent data – referred to here as projection interpolation (PI) metal artefact reduction (MAR). To our knowledge all MAR techniques that have found clinical application are based on PI. A notable example is the algorithm [Watzke and Kalender, 2004] implemented on the Siemens SOMATOM from 1987 to 1990, and which is still undergoing further development [Liu et al., 2009]. Commercial software such as ScanIP (Simpleware, Exeter, U.K.) offers PI as an image preprocessing tool. These MAR techniques can lead to lowering of detail and cause unpredictable secondary artefacts, such as that described by [Mahnken et al., 2003] as a “ground glass like fan-shaped artifact”.

In comparison, non-PI algorithms are computationally expensive and remain confined to academic papers [Watzke and Kalender, 2004, Bal and Spies, 2006, Man et al., 2001, Lemmens et al., 2009, Wang et al., 2000]. “Extended CT scale” techniques [Link et al., 2000] have been made redundant by the 16-bit quantization used in modern scanners such as the Toshiba scanner used in this study.

The aim of this study was to examine the extent to which the presence of a metal hip prosthesis, and the subsequent application of PI MAR, affect the segmentation of periprosthetic fibrous lesions. Does the presence of metal

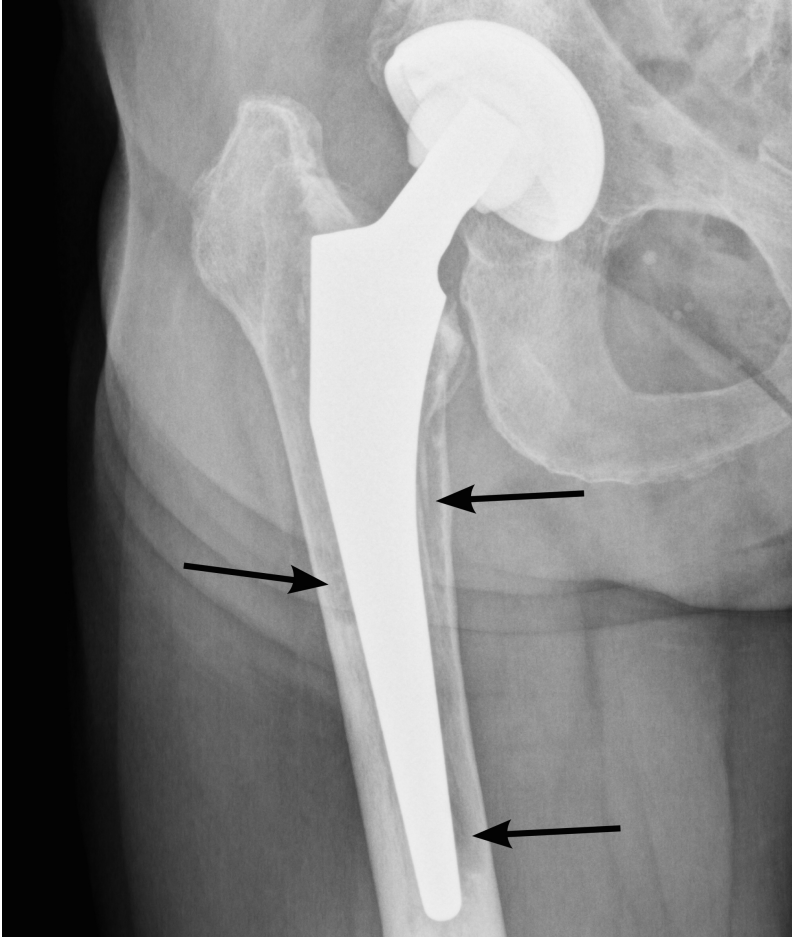


Figure 2.1 – In a clinical radiograph, osteolysis is indicated by radiolucent regions adjacent to the prosthesis (arrows) .

decrease the manual segmentation performance of such lesions? Does MAR improve the manual segmentation compared to the metal-degraded CT images? To answer these questions we first compare the segmented lesion volumes to ground truthed volumes obtained by filling each lesion with water. Second, we compare the segmentation boundaries between scans acquired under optimal metal-free scanning to those found in metal-degraded images, both before and after the application of MAR. Contrast and image intensity gradients are measured across segmentation boundaries to help explain the results. This enables us to either recommend or warn against MAR as a pre-processing step in assessing periprosthetic lesions.

2.2 Materials and Methods

Fig. 2.2 shows a flow chart of the complete experimental work flow. Ten human femora were retrieved post-mortem from seven donors. These comprised three female and seven male femora with a mean age of 80.7 years (range 67-98). Dual energy X-ray absorptiometry (DXA) measurements performed prior to preparation yielded a median T-score of -0.7 (average -1.4) within a range of -4.9 to $+1.1$. A T-score of -1 or higher is considered normal, whereas clinical osteoporosis is defined by a T-score of -2.5 or lower [Beers, 2006]. All femora were preserved in formalin and surrounding soft-tissue removed. We fitted each femur with a polished tapered cobalt-chrome Exeter size 42-2 stem (Stryker, Limerick, Ireland). For fixation we used radiopaque contrast-enhanced bone cement (Palacos, Biomet, Warsaw, IN, USA). The prostheses were implanted under supervision of an experienced orthopedic surgeon (H.J.L.vdH.) while using standard cemented implantation protocol.

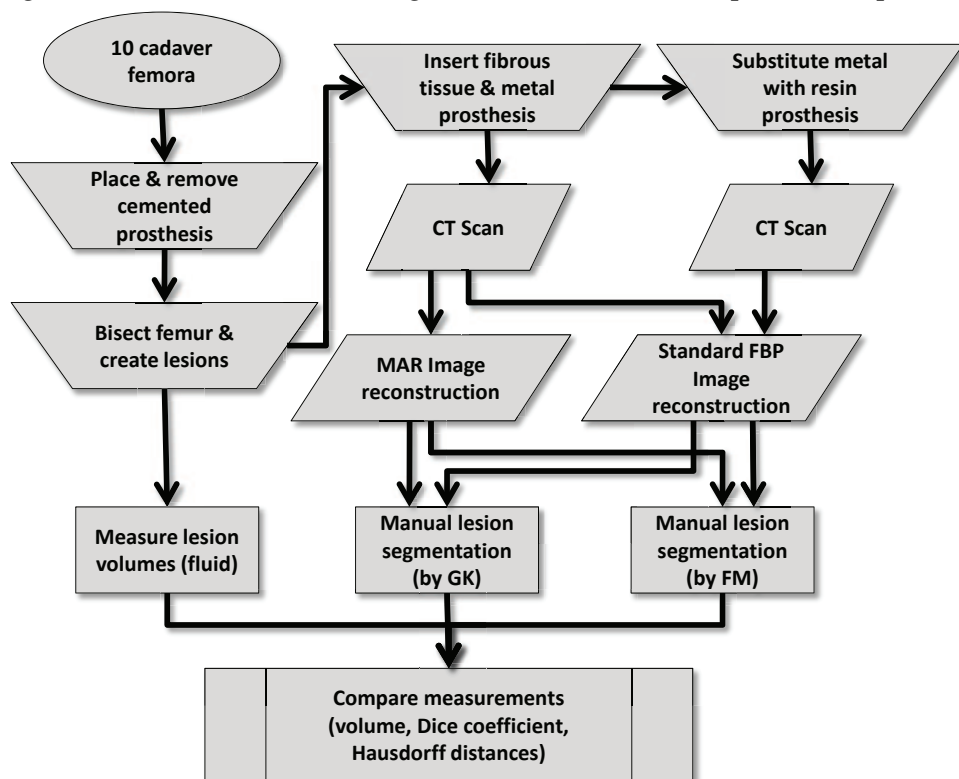


Figure 2.2 – Flow chart of the experiments performed in this study.

To enable scanning each femur with and without the metal prosthesis, we required removable prostheses. After each prosthesis was cemented it was mechanically removed from the femur, leaving the remaining cement mantle and femur intact (Figs. 2.3, 2.4). This was possible due to the Exeter prosthesis's smooth polished surface and tapered shape.



Figure 2.3 – The removable Exeter prosthesis is shown partly dislodged from one of the test femora.

Each femur was axially bisected so as to intersect the cement mantle. We subsequently created lesions both proximally and distally from the sawn-through interface and at varying locations along the circumference (Fig. 2.4) using a rotary burr (Dremel). In total, 27 cavities were created having a mean volume of 2.4 ml (range 1.1-5.0)ml. We measured lesion volumes by using a



Figure 2.4 – Femoral lesions mechanically created anterior and posterior of the cement mantle are shown with the prosthesis removed.

0.2ml graduated syringe to fill each cavity with water. The lesions were then drained and filled with a fibrous tissue substitute.

Previous studies used water [Stamenkov et al., 2003], lean beef mince [Walde et al., 2005, Weiland et al., 2005], or an unspecified “soft-tissue equivalent” material to fill artificially created lesions [Claus et al., 2003]. In this study we specifically chose radiologically compatible tissue to represent the fibrotic zones. On four occasions, real periprosthetic fibrotic tissue was retrieved during hip implant revision surgery and its CT opacity measured *ex vivo*. These tissues had a mean opacity of 72 Hounsfield Units (HU) with standard deviation of 10 HU. This differs substantially from water (mean 0 HU) and our measurements for lean beef mince (mean 50 HU). After evaluating several commercially available alternatives we chose chicken liver, which was considered sufficiently similar with a mean opacity of 77 HU and standard deviation of 6 HU.

During scanning each femur required an inserted prosthesis to hold the two bisected halves in place. When the metal prosthesis was removed we used a

mould-cast resin substitute. The resin had a measured CT opacity of 150 HU, placing it above the opacity of soft tissues and blood (≈ 50 HU), but less than bone (>300 HU) and much less than metal ($>3,000$ HU) [Mukherjee and Rajagopalan, 2007]. The resin prosthesis’s low radiopacity did not significantly contribute to beam hardening, the main source of metal-induced artefacts, and therefore enabled us to acquire optimal images for CT ground truthing.

Scans were performed on a helical CT scanner (Aquilion 16, Toshiba Medical Systems, Japan) at 135 kVp using a 200 mA tube current. The in-slice voxel spacing was 0.44×0.44 mm with a slice thickness of 0.5mm. Following the advice of [Lee et al., 2007] and [Douglas-Akinwande et al., 2006], we chose a standard smooth reconstruction filter (FC 12) to minimize metal artefacts.

For MAR we used the recent sinogram-interpolation method of [Veldkamp et al., 2010]. This algorithm has a lot in common with the original method of [Kalender et al., 1987] but uses raw sinogram data to interpolate metal traces. Adding a fraction of the original metal signal to the interpolation has a similar role as the nonzero “confidence parameter” of [Oehler et al., 2008] and makes the implant visible in the final reconstruction.

Each of the 27 fibrotic lesions was independently and manually segmented by each of two experienced users (F. M. and G.K.) using MITK, an interactive segmentation software tool [Maleike et al., 2009]. F.M. and G.K. independently segmented the resin prosthesis volumes as well as the metal prosthesis volumes with and without application of MAR. F.M. and G. K. segmented the volumes sequentially and in randomized order, with 2 weeks separating their segmentation work.

The volumes of the segmented lesions were compared to the physically measured ground-truthed fluid volumes. The metal-affected and MAR image segmentations were registered to their metal-free counterparts using a 3D iterative closest point (ICP) algorithm, correcting for translational and/or rotational offsets between scans. Geometric deviation in each segmented metal or MAR volume was compared to the corresponding metal-free resin prosthesis volume. To avoid interobserver bias when comparing segmentations performed with metal, MAR, or resin volumes, we always compared pairwise segmentations of the same lesion on a per-user basis. Measurements by F.M. and G.K. were treated as separate and not averaged.

The residual shape difference between each segmentation pair was computed by their Hausdorff distance, mean Hausdorff distance, and Dice coefficient. The Hausdorff distance is defined as the global maximum of all the minimum distances between two surfaces. The mean Hausdorff distance is the mean

minimum distance between the two surfaces. The Dice coefficient is a ratio between the volumes enclosed by the two surfaces, defined by $c = \frac{2|A \cap B|}{|A| + |B|}$ and has a value in the range $[0, 1]$ where 1 represents complete overlap between volumes and 0 represents completely disjoint volumes. A perfectly matched segmentation pair would have a zero Hausdorff distance and a Dice coefficient of one, whereas a bad match will have a high Hausdorff distance and Dice coefficient approaching zero. The Dice coefficient and Hausdorff distance are well suited to evaluating differences in 3D segmentation such as in [van der Lijn et al., 2008].

For each segmentation boundary we computed the median image gradient magnitude, as well as the Michelson contrast between the inner and outer region defined by this boundary. The Michelson contrast for each lesion is defined as $\frac{I_{out} - I_{in}}{I_{out} + I_{in}}$ where I_{in} and I_{out} represent the median image intensities in a 1 mm wide region symmetrically located inward and outward of the segmentation border.

Image registration, distance metrics, and contrast metrics were computed using the Insight Segmentation and Registration Toolkit (ITK), Visualization Toolkit (VTK) and the Python programming language. All computations were performed on the DeVIDE image processing and visualization platform [Botha and Post, 2008].

We did not assume normal distributions of the measured differences in volume, edge gradient magnitude, Michelson contrast, pairwise Hausdorff distances, or Dice coefficients. This decision was supported by the Shapiro-Wilk test for normality, indicating that the hypothesis of normality should be rejected for several of the measurement pairs, as is also visually evidenced in asymmetry in several of the measurement distributions (e.g. Figs. 2.6 and 2.8). Distributions of measurements and differences between measurement pairs are described by nonparametric measures such as median and interquartile range. Rather than the Student's t-test we therefore chose the Wilcoxon signed rank test to compare measurements of the same quantities under metal-free, metal-containing, and MAR acquisition. We furthermore chose not to assume linear relationships between variables when testing for correlation, choosing instead to use Spearman's rank correlation coefficient, which serves as a nonparametric analogue to Pearson's correlation.

2.3 Results

To answer whether the presence of metal degrades segmentation performance, we compared segmentations performed on the metal-free ground-truthed images to those of metal-affected images. In metal-free image segmentations we measured volumes that were not significantly different ($P = 0.65$) compared to the physically measured fluid volumes (Fig. 2.5), while metal-containing CT scans tended to significantly ($P = 0.002$) underestimate the physically measured volumes. The Hausdorff distances, mean Hausdorff distances, and Dice coefficients of metal-affected versus metal-free images show low dissimilarity albeit with several outliers (Figs. 2.6, 2.7, 2.8). Michelson contrast across segmentation boundaries is significantly lower ($P = 0.002$) than for metal-free scans (Fig. 2.9). Image gradient magnitudes on segmentation boundaries also have a lower median value compared to metal-free images (Fig. 2.10), although this difference is not significant ($P = 0.811$).

Parameter	Median difference	Significance
Volume	-0.1 mm	0.007
Hausdorff distance	-0.62 mm	0.400
Mean Hausdorff distance	-0.009 mm	0.474
Dice coefficient	0.0028	0.453
Median edge contrast	0.0046	0.885
Median edge gradient	-9 HU/mm	0.228

Table 2.1 – Interobserver differences calculated pair wise over all lesions according to the Wilcoxon signed rank test. The only statistically significant difference is a 0.1 ml bias in measured volume. (HU \equiv Hounsfield Units).

The second question is whether PI MAR improves segmentation performance relative to unprocessed metal-degraded CT. Unexpectedly, we found that volumes measured after application of PI MAR were even smaller than those measured in the metal-affected scans (Fig. 2.5), and significantly smaller than the ground-truthed volumes ($P < 0.001$). We see that the MAR segmentations exhibit significantly larger geometrical deviations ($P < 0.001$ in all three cases) from the ground-truthed results than unprocessed metal scans (Figs. 2.6, 2.7, 2.8). Michelson contrast across segmentation boundaries (Fig. 2.9) is significantly lower than for either resin scans ($P < 0.001$) or unprocessed metal ($P = 0.003$).

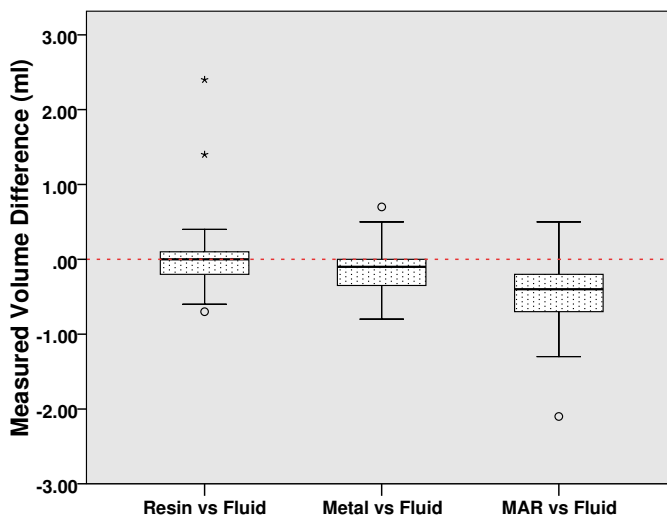


Figure 2.5 – Metal-free CT accurately estimates volume, whereas metal degradation causes volume underestimation. MAR causes even further volume underestimation.

Image gradient magnitudes on segmentation boundaries (Fig. 2.10) are significantly reduced compared to either metal-free ground-truthed or unprocessed metal scans ($P < 0.001$ in both cases).

We found no significant correlation between the lesion size or DXA T-score and any of the measured parameters using Spearman's rank correlation coefficient. Barring a statistically significant but small difference in segmentation volume we found observations between the two independent observers to agree well (Table 2.1).

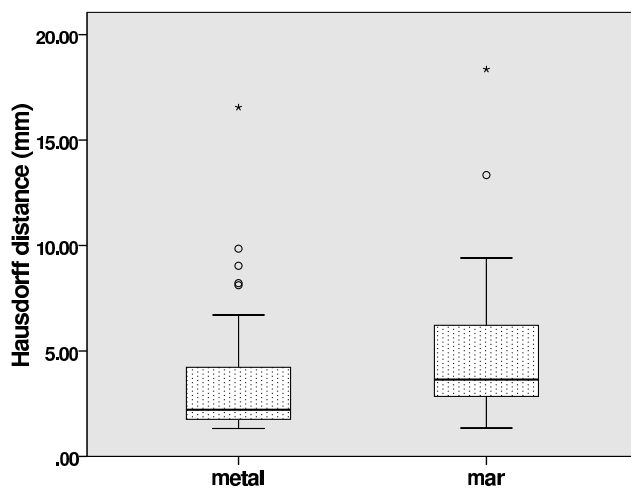


Figure 2.6 – Hausdorff distances compared to “resin” ground-truthed results show maximum local segmentation boundary errors.

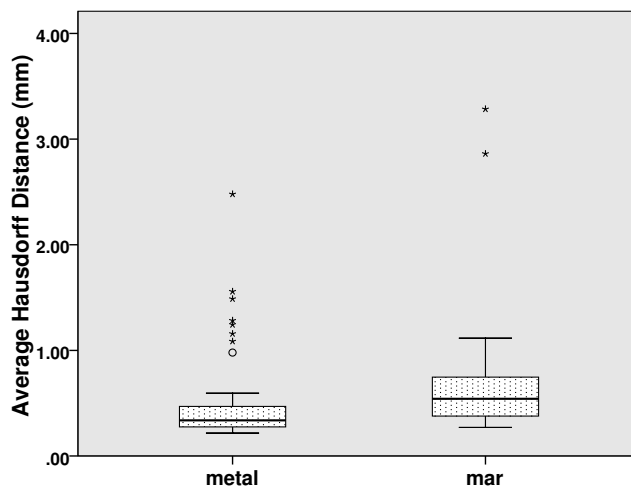


Figure 2.7 – Mean Hausdorff distances compared to “resin” ground-truthed results show average segmentation boundary errors.

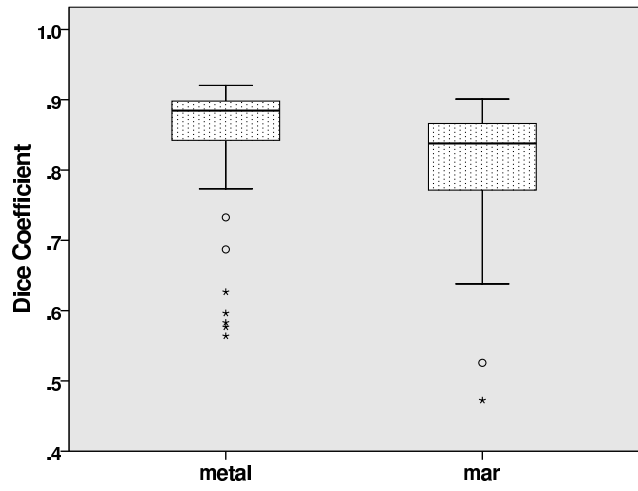


Figure 2.8 – Dice coefficients compared to “resin” ground-truthed results show volumetric agreement between segmentations.

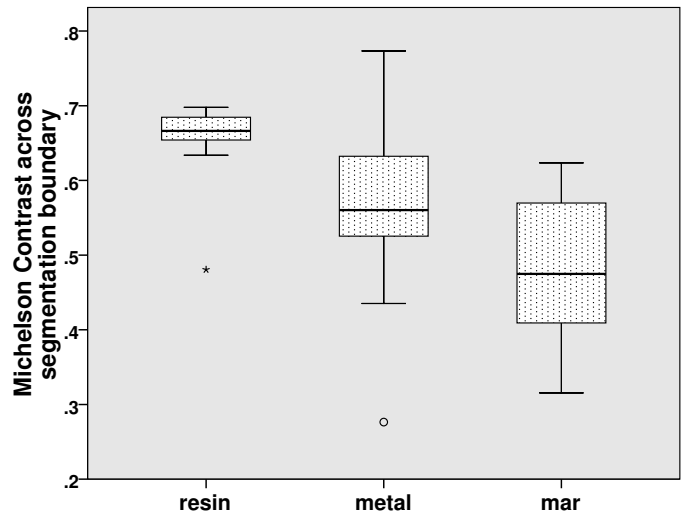


Figure 2.9 – The median Michelson contrast across each segmented lesion’s boundary.

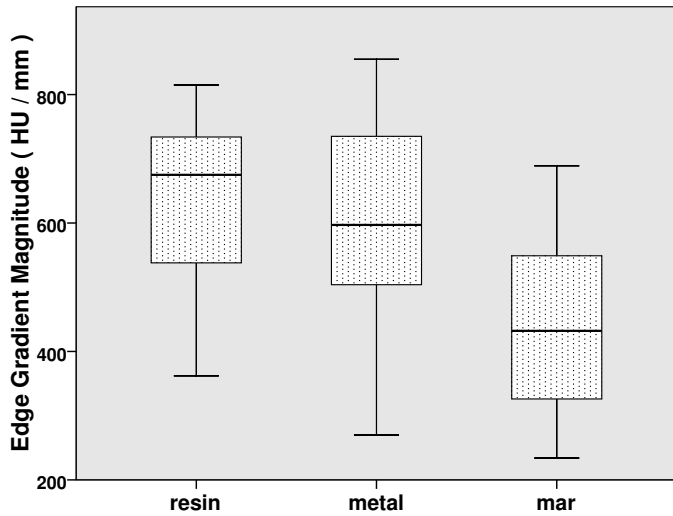


Figure 2.10 – The median edge gradient magnitude across each segmented lesion’s boundary.

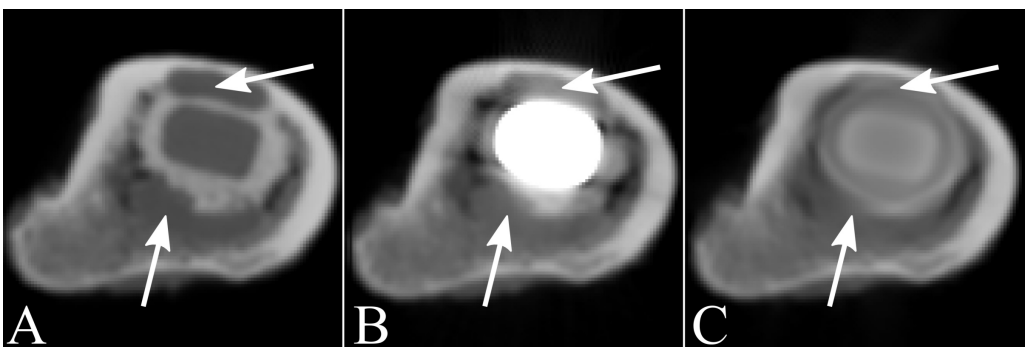


Figure 2.11 – A CT slice showing fibrous-tissue lesions (arrows). A: with resin prosthesis and B: metal prosthesis. C: after sinogram in-painting, metal artefacts are reduced at the cost of a loss in periprosthetic detail.

2.4 Discussion

We set out to determine whether the presence of a metal prosthesis and subsequent projection interpolation metal artefact reduction (PI MAR) affect the segmentation of periprosthetic lesions resembling osteolysis. We compared segmentation volume as well as geometrical deviation between segmentations performed with and without the presence of metal and after application of PI MAR.

We believe that the observed trend of lowered segmentation performance due to MAR is widely relevant to the diagnosis and quantification of periprosthetic tissues from CT. Our experimental data were obtained under optimal scanning conditions, with all soft-tissue removed from around the femora. In the clinical setting the image degrading effects of metal-induced beam hardening, as well as secondary artefacts created by MAR, are likely to present a greater obstacle to lesion detection and quantification than in the carefully controlled environment described in this paper. Through inspection we believe that the threshold we used for identifying metal prosthesis yielded a good segmentation of the metal boundary while still excluding all surrounding biological tissue. Using a different threshold affects the delineation of the interpolation region, and subsequently also the amount and the location of detail lost to the MAR algorithm. A detail-retaining compromise could involve decreasing the interpolation regions' size at the cost of artefact suppression.

We found no tendency for manual CT-based segmentation to either over- or underestimate lesion volume in the absence of metal hardware. When a metal prosthesis was introduced, however, lesion volume was underestimated. This agrees with [Walde et al., 2005] and [Leung et al., 2005] who found that CT neither consistently underestimated nor overestimated lesion volume, and [Stamenkov et al., 2003] who found that CT systematically underestimates lesion volume in the presence of metal artefacts. We explain this tendency by our measurements, which show that metal-induced artefacts cause lower contrast across lesion boundaries, which negatively influences their visibility. Contrary to expectation we found that lesion segmentation deteriorated even further after application of PI MAR, with larger associated underestimation of lesion volume and larger geometrical errors. PI MAR reduced image noise in homogenous regions, but this was achieved at the cost of a substantial loss of detail, evidenced by lowered edge gradients and image contrast across lesion boundaries. [Kalender et al., 1987] mentioned that PI MAR works best for objects with simple near-circular geometries, while [Watzke and Kalender, 2004] mentioned that PI MAR is well suited to larger implants consisting of

dense metal. This view is also echoed by [Liu et al., 2009] who wrote that MAR improved image quality in scans of large prostheses, whereas it had a negative effect on small metal objects due to image blurring. In this regard we expected PI MAR to be of benefit since the Exeter prosthesis chosen for this study meets the requirements mentioned above. Except for its smoother appearance (Fig. 2.11C), there is little to recommend the application of PI MAR above the original metal degraded image (Fig. 2.11B). Detail in the MAR image is noticeably blurred – especially in regions closest to the metal prosthesis. This is supported by a measured lowering of edge contrast and edge gradient magnitude after application of PI MAR (Figs. 2.9 and 2.10).

Papers showcasing MAR algorithms [Mahnken et al., 2003, Bal and Spies, 2006, Lemmens et al., 2009] emphasize “starburst” artefacts by choosing display windows that create the impression that these artefacts completely obliterate all image detail in their path. This study suggests that a human operator who has to manually delineate structures adjacent to a metal prosthesis might obtain better segmentations from unfiltered artefact-containing images than from images processed with PI MAR. This contrasts with the view that MAR invariably improves the appearance and usefulness of metal-affected clinical scans. However, in patients with bilateral prostheses, as often seen in practice, the beam hardening shadow connecting the two prostheses is much more pronounced than in this single prosthesis experiment. In this scenario PI MAR can improve the subjective appearance of radiographic cross sections by equalizing the shadow regions [Watzke and Kalender, 2004, Lemmens et al., 2009]. This improvement is often confirmed by radiologists’ subjective rating [Liu et al., 2009]. For our application of measuring periprosthetic lesions, however, there seems to be a net loss of quantifiable image information when applying PI MAR.

CT, in the absence of metal artefacts, is an accurate and unbiased tool for measuring the volume and geometry of periprosthetic lesions. When adding the presence of a metal prosthesis the result remains usable, albeit with degraded image quality, increased difficulty in discerning structures, and a tendency to underestimate lesion volume. Previous studies [Liu et al., 2009, Bal and Spies, 2006] investigating the merits of MAR used subjective rating scales to assess image quality and limited quantitative measurements to mean CT number and standard deviation within certain regions of interest. A strength of our study is its quantitative evaluation of segmentation performance, albeit for a small set of lesions.

A limitation of this study is the inclusion of only 27 fibrotic lesions from 10 hu-

man cadaver femora, all using the same type of metal prosthesis and scanned in the same CT scanner. We limited ourselves to evaluating a single software PI MAR implementation, and independent segmentations were performed by only two operators. Although the femora were harvested from older patients, only two of the 10 samples had DXA T-scores suggesting osteoporosis, whereas osteoporosis may be more common in patient populations. Our manually created lesions lacked the radio-dense sclerotic borders that may be found in clinical practice [Bauer and Schils, 1999, Sofka, 2007]. The current clinical significance of MAR algorithms is low, although it remains an active field of research. We suggest that in addition to our general observations, validation should be performed in any specific clinical setting whenever PI MAR is to be considered.

2.5 Conclusion

Despite its popularity in the literature and superficial improvements to image appearance, projection interpolation metal artefact reduction (PI MAR) was detrimental to the user-guided segmentation described in this paper. It remains to be seen whether other image-based metal artefact reduction techniques can improve quantitative segmentation results of such periprosthetic lesions.

