

Research Article

Assessing Concordance in Post-operative Knee Implant Position Analysis Between Automated Image Analysis Algorithm and Human Radiologists

Arman Motesharei^{1a}, Kevin de Souza^{1b}, Benjamin Harder^{1c}, Regis Pailhe^{2d}

¹ Stryker (Netherlands), ² Department of Orthopaedic Surgery and Sport Traumatology, Université Grenoble Alpes

Keywords: Bayesian analysis, Total Knee Arthroplasty, Implant detection, Knee alignment

<https://doi.org/10.60118/001c.124659>

Journal of Orthopaedic Experience & Innovation

Vol. 6, Issue 1, 2025

Aim

The authors have developed an automated algorithm-based technique to identify implant position based on post-operative CT scans for total knee arthroplasty. To analyse the performance of this algorithm we compared the implant orientation against two trained radiologists. This was an exploratory secondary analysis from CT scans collected as part of an RCT.

Methods

To assess this algorithm we used Bayesian hierarchical modeling, which has the unique benefits of being able to accept the null hypothesis and allow interpretation of concordance between the radiologists and our algorithm.

Results

Results show this automated algorithm is effective for deriving post-operative implant locations from CT scans. Lack of concordance in certain parameters are likely driven by the challenges associated with locating certain implant landmarks due to the geometry of the components. Such limitations will not impact the accuracy of our automated algorithm. With the increase in 3D pre-surgical planning and robotic assistance, these techniques can be used in future studies, such as assessing implant alignment philosophies.

Conclusion

Overall, this automated algorithmic approach for locating post-operative implant position provides good accuracy, valuable insights in surgical execution and saves resources by reducing the time demands on trained medical specialists.

-
- a Arman Motesharei is a Data Scientist for Stryker Research and Development. He has worked in orthopaedics for nearly a decade in both academic and industrial capacities, focusing as well as publishing work in Robotic-assisted Orthopaedic surgery.

[Connect with Arman Motesharei on LinkedIn](#)

[Conflicts of Interest Statement for Arman Motesharei](#)

- b [Conflicts of Interest Statement for Kevin de Souza](#)

- c [Conflicts of Interest Statement for Benjamin Harder](#)

- d Dr. Regis Pailhe is a Professor of Orthopedic Surgery.

[Visit Dr. Regis Pailhe](#)

[Conflicts of Interest Statement for Dr. Regis Pailhe](#)

This article, like all JOEI articles, is available for continuing education

Reflect Now!

Click here : <https://joeipub.com/learning>

INTRODUCTION

Identifying post-operative implant position is essential for the analysis of orthopedic implants in knee replacement surgery. These measurements can provide quantitative data on surgical execution and can provide insights on different surgical methods, such as conventional, navigation and robotic-assisted.

Unfortunately, making these measurements (for example, by identifying landmarks in CT images) is time-consuming and requires trained medical specialists. Even with trained professionals we observe differences in intra- and inter-observer reliability for certain implant position parameters (Miura et al. 2018; Yoshino et al. 2019).

Most post-operative assessments rely on plain radiographs, their accessibility, low cost and low radiation dose makes it a mainstay in routine clinical practice. However many factors such as leg rotation, flexion and magnification errors can affect these measurements and can lead to inaccuracies (Brouwer et al. 2007; Holme et al. 2011a; Radtke, Becher, Noll, et al. 2010; Jamali et al. 2017).

Additionally, it is not possible to measure the 3D position & orientation of an implant from a 2D radiograph. Implant rotation can be assessed using two-dimensional computed tomography (2D-CT) (Berger et al. 1998; Jazrawi et al. 2000). While this method provides 3D data, 2D slices may lead to inaccuracies due to leg orientation during scanning influencing the location of anatomical landmarks (Slevin et al. 2018).

To assess post-surgical implant position accuracy, it is essential to establish a reproducible and accurate evaluation method. Unfortunately, the reproducibility and evaluation method for post-operative alignment is often not validated (De Valk, Noorduyn, and Mutsaerts 2016). Recently the use of automated 3D measurement based on CT scans shows promise in deriving implant position values more reproducibly (Slevin et al. 2018; Watanabe, Akagi, Shiko, et al. 2021; Hirschmann et al. 2011; Cerquiglini, Henckel, Hothi, et al. 2018).

Stryker has developed a methodology for locating implants in CT images where we automate all stages of the process, thus eliminating human error and variability. Using a novel approach of Bayesian hierarchical modeling, the authors aimed to determine if concordance exists (defined as a three-way agreement between the two radiologists and the algorithm). Therefore, the model will evaluate corroboration between both the radiologists and the algorithm.

MATERIAL AND METHODS

PATIENTS AND SURGICAL TECHNIQUE

We collected CT scans from patients undergoing TKA surgery in the RCT (NCT03566875) who received primary TKA between March and December 2018. In total 60 patients were enrolled in this study, and we received consent from 48 patients to use their data. Patient demographic data, including sex and age, were not provided to the authors since they were considered unnecessary for this study.

The surgeons used the Triathlon™ Knee (Stryker, Kalamazoo, Michigan, USA) for all cases with cement fixation. Our dataset has a mixture of robotic-assisted and manual TKA. All patients received a pre-operative CT scan and underwent a post-operative CT scan 7 days after surgery. Both scans were subject to the Stryker Mako CT protocol, which took 1 mm slices at the knee, and lower resolution images at 2 mm slices at the hip and ankle.

AUTOMATED IMAGE ANALYSIS

CT scans were analysed to compare post-operative implant position to the pre-operative planned position. The analysis workflow (see [Figure 1](#)) was performed with a custom automatic segmentation and automatic registration process. Step 1 involved automatic segmentation of the femur and tibia in the pre-operative CT image. Secondly, the bones were automatically segmented in the post-operative CT image. Segmentation of bones can be difficult in a post-operative image for three reasons: (1) portions of the patient's bone have been removed during surgery; (2) identifying bright voxels and edges corresponding to the implant itself; and (3) scattering artefacts due to the presence of metal in the image. A specialised version of the automatic segmentation process was developed to overcome all three challenges. After segmenting the bones, the automated segmentation tool was then used to segment the implant. Segmentation of the implant can also be difficult, due to scatter and other image artefacts resulting from the metal of the implants. To overcome these challenges, a database of implant designs was used to select a 3D model of the exact design and size of the implant used in each surgical case, and a specialised automatic segmentation model based on that implant was employed.

After segmentation was completed, an automated registration process was used to align the post-operative bone with the pre-operative bone. The executed implant placement was then compared with the intra-operative planned

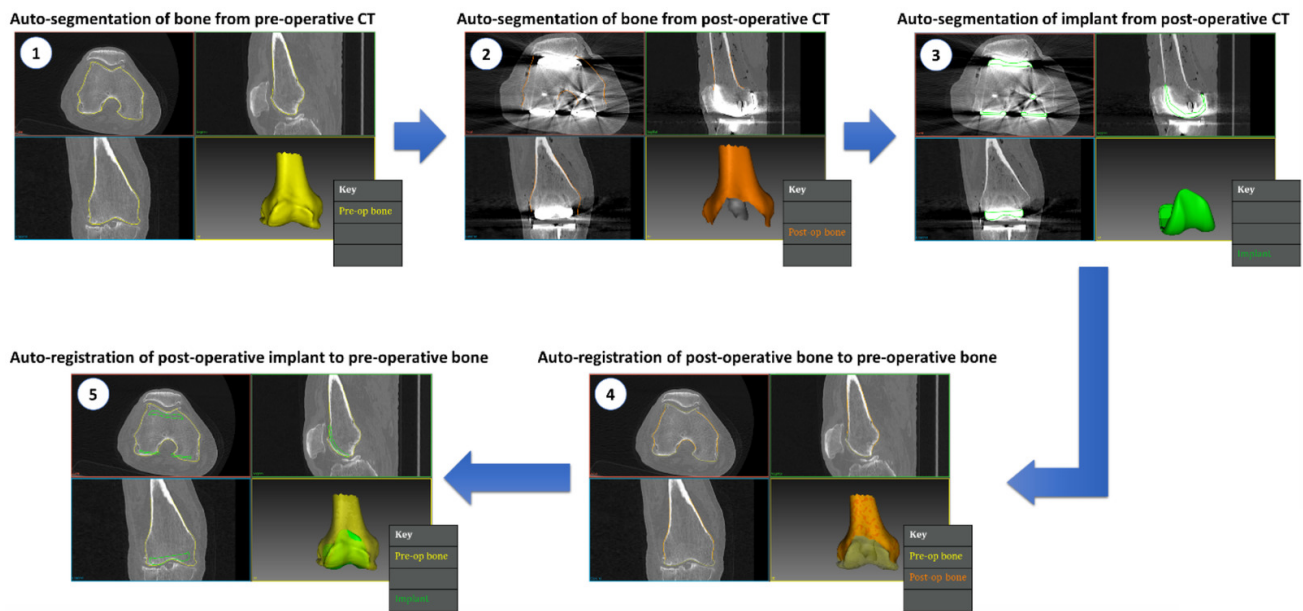


Figure 1. Description of workflow to analyze surgical implant placement based on pre-operative and post-operative CT scans. An automated segmentation and automated registration process was developed and specialized for this study.

implant position and the difference between these two plans was calculated to provide the error in implant placement.

MANUAL IMAGE ANALYSIS

Two consultant radiologists independently evaluated all 48 patients and repeated their measurements twice. For each patient the radiologists marked several landmarks from CT images. The landmarks used to compute implant axes in 3D space are shown in [Table 1](#). A 3D illustration of these landmarks on the femoral component is shown in [Figure 2](#), where each colour represents a different radiologist.

The medial-lateral femoral axis (x-axis) was defined as the average between the lines generated by landmarks 1 & 2, 3 & 4, and 5 & 6. Rotation about the x-axis will be the flexion/extension angle. The superior-inferior axis (z-axis) was defined as the average between lines generated by landmarks 3 & 5, and 4 & 6. Rotation about the z-axis is the internal/external angle. The anterior-posterior axis (y-axis) was the cross product of the two previously generated axes and the rotation about the y-axis is the varus/valgus angle.

Similarly the tibial medial-lateral axis was defined as the average lines generated by landmarks 7 & 8, and 9 & 10. The anterior-posterior line was defined from the landmarks 11 & 12, and the superior-inferior axes were the cross product of the two previously generated axes.

THEORY/CALCULATION

BAYESIAN STATISTICAL ANALYSIS

We employed a hierarchical Bayesian model to address the concordance dilemma, aiming to determine whether the

landmarks identified by two radiologists were indistinguishable from each other and from our algorithm (Gelman et al. 2013).

To explain concordance, consider the graph in [Figure 3](#) where each node represents an agent in the experiment. We want to test if:

1. The landmarks identified by two radiologists (nodes A and B) are indistinguishable.
2. Radiologist's landmarks match those of an algorithm (node C).
3. All three agents (both radiologists and the algorithm) produce indistinguishable results.

We're testing whether there's any detectable difference among all agents in the network (Linde et al. 2021).

Bayesian hierarchical models, also known as multilevel models, are probabilistic frameworks used to estimate parameters across different groups. These models account for both the individual characteristics of each group and the similarities between them, enabling more accurate predictions. By sharing information across groups, they help prevent overfitting (overly complex models) or underfitting (overly simple models). A key feature of hierarchical models is shrinkage, which naturally regularises estimates by pulling extreme values or results from small groups toward the overall average, thereby reducing the influence of outliers or sparse data (Lustig et al. 2021).

For this research, the groups were stratified into two levels: patient and radiologist. Where $j=2$ for radiologists and $k=48$ for total knees. The hierarchical model architecture lends itself well to the data, given that there are multiple, repeated observations per doctor per patient, which would violate the assumption of independence for many frequentist statistical tests. The full model structure and equations are detailed in Appendix A.

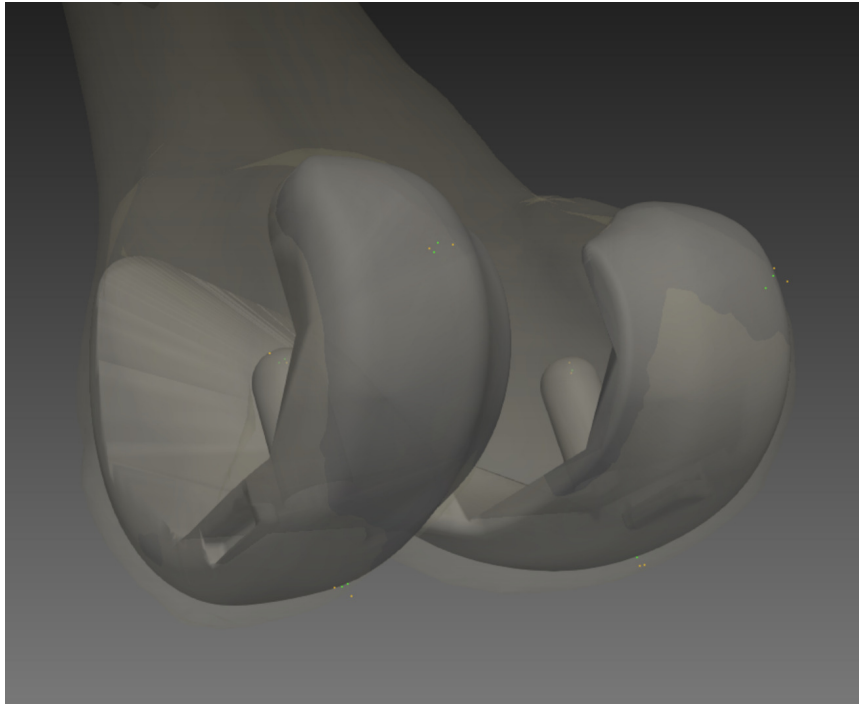


Figure 2. Illustration of femoral landmarks taken by two radiologists (green: radiologist 1, orange: radiologist 2) in 3D.

Table 1. Landmarks taken manually by the radiologists for femur and tibia

Femur	
1	Posterior medial implant
2	Posterior lateral implant
3	Distal medial implant
4	Distal lateral implant
5	Medial femoral peg
6	Lateral femoral peg
Tibia	
7	Posterior medial implant
8	Posterior lateral implant
9	Medial baseplate
10	Lateral baseplate
11	Posterior baseplate
12	Anterior baseplate

One of the advantages of Bayesian methods is that they afford the researcher multiple options in interpreting results that extend and, in many cases, surpass the null hypothesis significance testing framework. By comparison, Bayesian methods allow the decision criteria to be created to reject or accept the null hypothesis. These options map well to this research objective, as the goal is to accept the null hypothesis per the concordance research objective and aligns with Bayesian principles by:

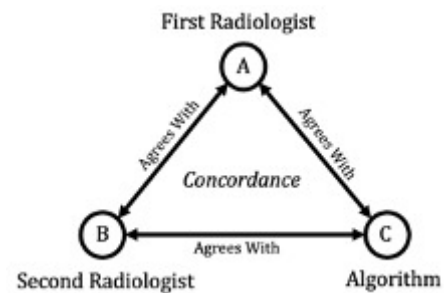


Figure 3. Graph network demonstrating the problem of concordance

1. **Incorporating uncertainty** explicitly in the analysis, providing a richer understanding of the parameter estimates.
2. **Facilitating hypothesis testing** in a Bayesian framework through the use of probabilistic statements about parameters rather than relying solely on p-values.
3. **Enhancing interpretability** of the results by examining the relationship between parameters and their joint effects on outcomes.

To evaluate concordance, we applied a Region of Practical Equivalence (ROPE) (Kruschke 2018), establishing a boundary of ± 0.5 degrees. If more than 95% of the joint posterior distribution falls within this boundary, concordance is achieved. Otherwise, concordance is rejected. [Figure 4](#) illustrates this decision process.

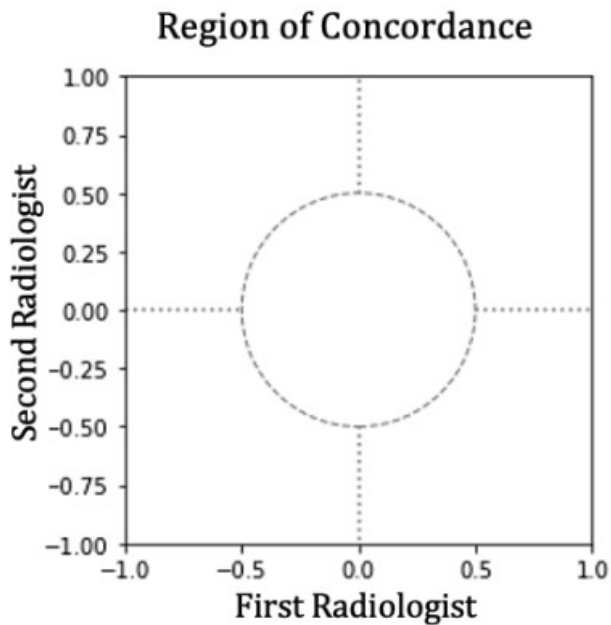


Figure 4. Demonstration of the Region of Practical Equivalence or ROPE of concordance. The ROPE is essentially mapped from a one-dimensional to two-dimensional space.

RESULTS

The models were fit independently for the flexion/extension, varus/valgus, and internal/external angles for both the tibia and femur using the Metropolis-Hastings algorithm, with 3000 samples generated and 1000 subsequently discarded during the initial sampling process to promote chain convergence and optimal mixing. All models displayed good mixing and convergence. [Table 2](#) represents the model output related to the radiologists' common effects. Figures [5](#) and [6](#) show the joint distribution plots of the posteriors for the common effects mapped to the concordance decision rule.

In this paper, positive values of flexion/extension indicate more flexed, positive values of varus/valgus more varus, and positive values of internal/external rotation more internal.

The above figures (5, 6) show the 95% HDI for the joint posterior distributions mapped to the concordance decision rule. The plots demonstrate concordance for the femur of varus/valgus and internal external. By comparison, concordance only exists for varus/valgus for the tibia. When reviewing the posterior distributions for the tibia, it is apparent that the distributions are much broader when compared to the femur. Given that the data points are the same for both anatomical regions, the broader joint posterior distributions are driven by larger variances within the data and between the radiologists. Much of this variance can also be assumed to be driven by the inherent challenges of identifying landmarks for the tibia, which will be addressed in the Discussion.

To gain intuition as to how the model functions and how the concordance decision rule can be interpreted, we can draw an example from the flexion/extension of the femur. Reviewing [Figure 5](#), it can be concluded that concordance was not achieved for this angle. However, further insight can be illuminated upon closer inspection of the individual posterior distributions for both radiologists, as shown in [Figure 7](#), which displays the individual posterior plots for the femur flexion/extension for the fixed effects of both radiologists. The plot demonstrates the posterior distributions, the mean of the posterior distribution labelled in black, the percentage of the distribution falling above or below the posterior mean in orange, and the percentage of the posterior falling within the ROPE in green.

The posterior distribution on the left demonstrates that the first radiologist disagreed with the algorithm, as shown by the entire distribution excluding the ROPE boundaries. By comparison, the distribution on the right shows that approximately 39% of the posterior contains the ROPE, meaning that a conclusion must be withheld for this radiologist. It is, therefore, acceptable to conclude that in this instance, not only did the radiologists disagree with the algorithm but with each other as well as neither posterior cross the

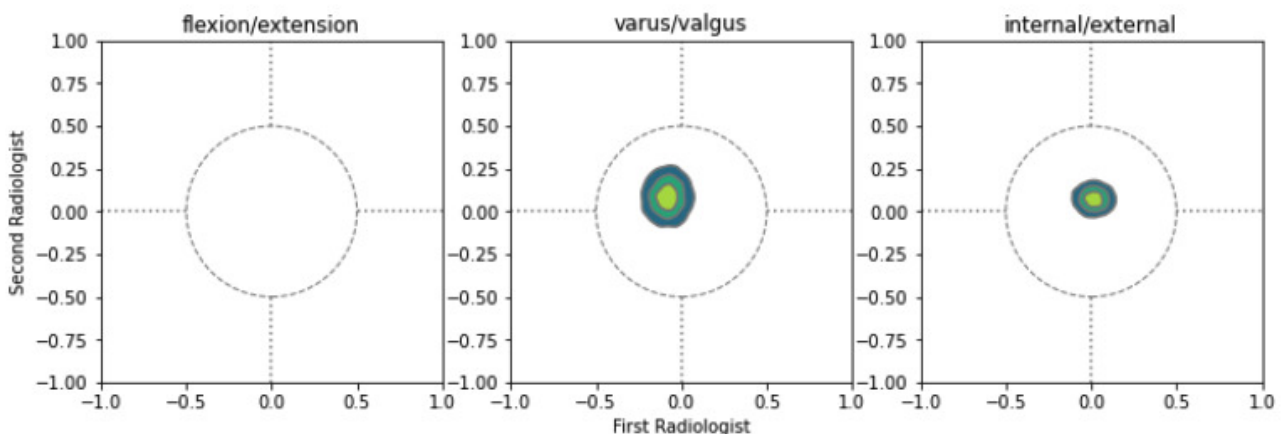


Figure 5. Posterior plots for the femur, the contours on the plot represent one standard deviation from the mean of the posterior.

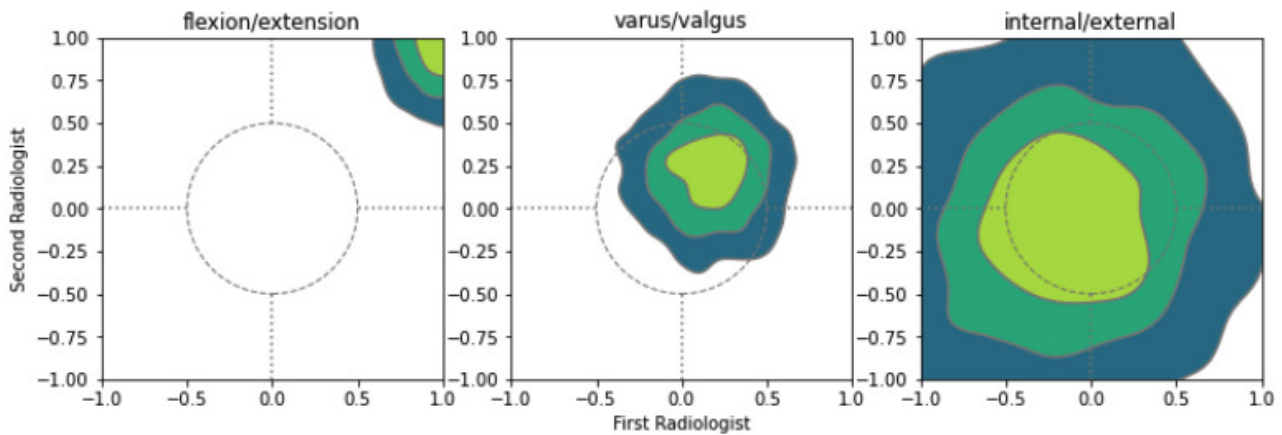


Figure 6. Posterior plots for the tibia, the contours on the plot represent one standard deviation from the mean of the posterior.

Table 2. Coefficients for the fixed effects of the mean error by radiologist. The values in parentheses represent the upper and lower 95% highest density interval of the posterior.

Anatomy	Angle	First Radiologist	Second Radiologist	Concordance %	Notes
Femur	flexion/extension	1.63 (1.11, 2.14)	0.57 (0.09, 1.06)	No < %1	Both radiologists disagree with algorithm and each other see figure.
	varus/valgus	-0.08 (-0.26, 0.10)	0.08 (-0.14, 0.29)	Yes 100%	Concordance achieved
	internal/external	-0.02 (-0.13, 0.17)	0.07 (-0.05, 0.20)	Yes 100%	Concordance achieved
Tibia	flexion/extension	1.01 (0.50, 1.5)	1.01 (0.41, 1.6)	No 5%	Both radiologists disagree with algorithm
	varus/valgus	0.16 (-0.39, 0.66)	0.24 (-0.39, 0.82)	Yes 98%	Concordance achieved
	internal/external	-0.28 (-1.46, 0.85)	0.01 (-1.46, 1.30)	No* 59%	Results inconclusive

barrier of the concordance decision rule as shown in [Figure 5](#).

DISCUSSION

There has been regular and controversial discussion within the orthopaedic community regarding optimal implant alignment philosophies (Rivière et al. 2018; Winnock de Grave et al. 2022). From initial mechanical alignment approaches (Rivière, Vigdorchik, and Vendittoli 2019) – albeit to prolong implant life (Lotke and Ecker 1977; Moreland, n.d.) – to the more modern kinematic (Hiranaka et al. 2022) and functional alignment techniques (Lustig et al. 2021). In recent years the accuracy of robotic-assisted surgery (Mahoney et al. 2022; Bell et al. 2016; Scholl et al. 2022) has enabled these novel alignment approaches. Surgeons are no longer restricted by cutting jigs, and the advent of robotic systems can assist their desired implant position and execute plans that manual surgery cannot (Shatrov et al. 2022).

To evaluate the differences in implant philosophies we need a robust and reproducible measurement tool. While radiographs are routine practice for post-operative assessment due to their low cost and low radiation dose, they are susceptible to limb flexion and rotation impacting the alignment measurements at the time of radiographic assessment. Indeed when investigating post-operative implant positioning Radtke et al. found a range of 3.83 degrees in mechanical lateral distal femoral angle (mLDFA) measurements based on rotation and 8.36 degrees in the mechanical lateral distal femoral angle (mLDTA) (Radtke, Becher, Noll, et al. 2010). Similarly when Nisar et al (Nisar et al. 2020). studied kinematic alignment in TKA they noted that radiographic measurements of limb alignment are prone to error due to the rotation of the lower limb and magnification errors. As little as 3° of rotation can lead to a significant difference in the measured alignment (Jamali et al. 2017).

Therefore for a study of this nature where maximising measurement accuracy is fundamental, 2D imaging cannot be relied on for assessment of implant position (Slevin et

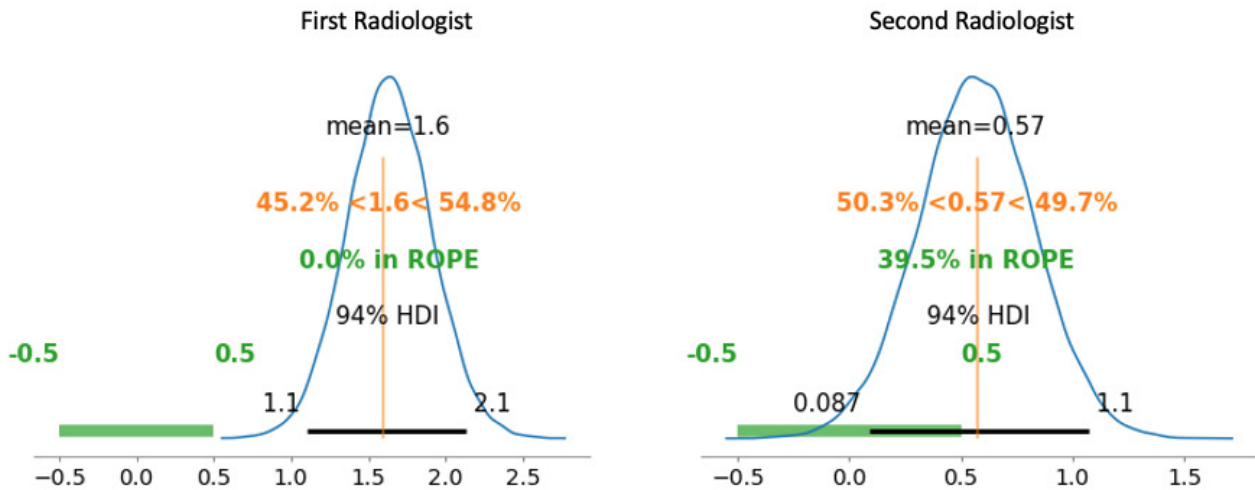


Figure 7. Individual posterior plots for the femur flexion/extension for the fixed effects of both radiologists. The plot displays the mean of the posterior in black, the observed mean value, and the percentage of the distribution above and below the mean in orange, and the ROPE in green, along with the percentage of the distribution that includes the ROPE boundary.

al. 2018), while CT has been shown to be superior in assessing limb alignment and component position (Holme et al. 2011b). In addition x-rays cannot determine implant rotation due to their 2D nature, whereas this tool is able to automate the assessment of rotational implant alignment.

While our results have successfully identified key points and some agreement between human radiologists and the algorithm, several limitations are found with this study. The femoral component has shown excellent concordance for the varus/valgus and internal/external implant. However, we didn't observe this level of concordance in flexion/extension. Neither radiologist agreed with the algorithm, and were inconclusive against each other.

The landmarks used for generating the flexion/extension angle relied heavily on the femoral pegs and curved surfaces. The tip of these points can be difficult to locate accurately on 2D slices and is impacted by patient position within the CT scanner (Slevin et al. 2018). Indeed, a similar dynamic was observed by Watanabe et al. Where their largest error was seen in the sagittal alignment of the femoral component at 1.7° - similar to this study.

We saw less concordance in the tibial component. While concordance was achieved for tibial varus/valgus, the other planes were less conclusive. This is to be expected as the tibial landmarks used in this study (Table 1) are difficult for a human to locate accurately. In contrast to the femoral component where the overall shape has more distinct points that a human can locate due to sharp edges and vertices, the tibial component mostly consists of shallow curves. It is difficult to locate the edges of these points in 2D slices, and as previously mentioned can be impacted by patient position within the CT scanner (Slevin et al. 2018). Furthermore, the tibial landmarks are largely coplanar and therefore the tibial pose may have greater uncertainty than the femoral component. Indeed this is borne out in the data where the standard deviation of the posterior for tibial

landmarks is larger than the femoral landmarks. This variance is not caused by lack of data as the number of observed data points are identical for both femoral and tibial implants as discussed in the Results section (Figure 7). Therefore, we can conclude that the larger variance of the posterior is driven by the challenges associated with the geometry of the tibial component. The implant shape does not limit the accuracy of the automated algorithm process. However even though there are benefits to the automated algorithmic approach, we still don't know the ground truth, therefore cannot conclude this approach is superior. We can only conclude there is no concordance in certain parameters of measurement, however the differences in concordance can be hypothesised by the constraints discussed above.

One limitation of this study was that we were restricted to the landmarks taken by the radiologists for the original surgical accuracy study (NCT03566875) which we compared against our algorithm. However, other measurements could potentially be made upon the CT images to improve the accuracy of manual markup. For example, marking lines and planes in the image, especially in 3D rather than limited to individual slices, could reduce the potential errors.

This is not a limitation of the technology, rather a limitation of the current method for measuring post-op CTs. Measuring landmarks is the limitation. This algorithm looks at the entire implant in the whole image, whereas humans pick a few points and assess them in limited slices.

Due to this limitation in manual markup, we only looked at the axes defined by the reference points obtained on the implant itself, thus this study is essentially evaluating the radiologists analysis of implant landmark accuracy against our algorithm. For future studies if we want to observe alignment philosophies we should take implant positions relative to patient anatomical axes to give more clinical relevance and context. This comes with its own challenges

such as locating anatomical axes in a post-operative CT scan due to implant scatter. Stryker has a technique to address this issue if the same patient's pre-operative CT scan is available and has been used in previous studies (Scholl et al. 2022). While this solves the scatter issue on the algorithm side, it will still be difficult to validate against human radiologists as they still must locate anatomical landmarks in post-operative CT scans – which will contain implant scatter and bone resections.

Stryker has also developed techniques, using a combination of image processing and machine learning, to reduce metal artefacts and noise in CT images, which could improve the accuracy of measurements made in CT. Finally, this algorithm has been designed for a specific family of Stryker implants and can only be applied to other prostheses where the CAD models are available.

Despite these limitations, the algorithm has proved effective, and can be used for future work. With the increase in robotic-assisted surgery and 3D pre-surgical planning, such algorithms can be used to automate and better assess the execution of planned implant positioning in three dimensions in post-operative CT scans.

This technology enables several avenues for future research. One direction is the potential use of our algorithm to monitor implant loosening over time. Although the current study focuses on assessing the congruence of implant positioning immediately post-operation, the algorithm could be adapted for longitudinal studies. By analysing repeated CT scans over time, the algorithm could detect subtle changes in implant position, potentially serving as an early indicator of loosening. This would greatly enhance the long-term monitoring of implant stability and patient outcomes.

Another potential area for future work is the generalisation of this technology to other surgical procedures and anatomical regions. It has already been successfully used in hip replacement surgeries to evaluate implant positioning accuracy (Rodriguez et al., n.d.). Expanding the use of this technology could provide a standardised method for assessing implantation accuracy across various surgical fields, thereby improving the consistency and quality of post-operative assessments and relieving repetitive work.

CONCLUSION

The automated algorithm shows considerable success in identifying key points and valuable landmarks in CT scans. While there are opportunities for further refinement, the algorithm demonstrates strong performance in post-operative implant assessment. Discrepancies between radiolo-

gists and the algorithm are primarily due to technical challenges in landmark identification rather than limitations of the algorithm itself. Overall, this approach provides good accuracy and valuable insights into surgical execution, reducing the need for time-consuming manual evaluations by trained specialists.

Benefits for:

1. Surgeons and Clinicians: This automated approach offers a more efficient and potentially more consistent method for post-operative evaluation, reducing the time and resource burden of manual assessments. It could enable more routine and comprehensive post-operative analyses.
2. Researchers: The algorithm provides a standardised, reproducible method for assessing implant position, which is crucial for comparing different surgical philosophies and techniques, including the evaluation of novel alignment approaches enabled by robotic-assisted surgery.
3. Hospital Administrators: Implementing an automated system leads to more efficient use of radiologists' time and potentially reduces costs associated with post-operative assessments.

Beyond assessing knee implants, this algorithm holds promise for expanding to other joints, and has already been successfully applied to hip implants, where accurate post-operative evaluation is equally important. With further development the algorithm could be adapted to monitor implant loosening over time through longitudinal CT scan analysis. This would allow for early detection of loosening, potentially improving long-term patient outcomes by enabling timely intervention.

Overall, this automated tool offers valuable advantages over manual methods in post-operative evaluations as it supports researchers and clinicians in assessing surgical decisions about implant positioning. As robotic surgery becomes more common, such tools will be essential for precise evaluation, contributing to better patient outcomes. This methodology is also adaptable for various joints and long-term monitoring, making it a versatile asset in surgical research and practice.

ACKNOWLEDGMENTS AND AFFILIATIONS

Arman Moteshareei, Kevin de Souza and Benjamin Harder are employees of Stryker. Kevin de Souza has shares in Stryker.

Submitted: June 03, 2024 EST. Accepted: October 12, 2024 EST. Published: January 25, 2025 EST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.

REFERENCES

- Bell, S. W., I. Anthony, B. Jones, A. MacLean, P. Rowe, and M. Blyth. 2016. "Improved Accuracy of Component Positioning with Robotic-Assisted Unicompartamental Knee Arthroplasty: Data from a Prospective, Randomized Controlled Study." *J Bone Joint Surg Am* 98 (8): 627–35. <https://doi.org/10.2106/JBJS.15.00664>.
- Berger, R. A., L. S. Crossett, J. J. Jacobs, and H. E. Rubash. 1998. "Malrotation Causing Patellofemoral Complications after Total Knee Arthroplasty." *Clin Orthop Relat Res* 356:144–53.
- Brouwer, R. W., T. S. Jakma, K. H. Brouwer, and J. A. Verhaar. 2007. "Pitfalls in Determining Knee Alignment: A Radiographic Cadaver Study." *J Knee Surg* 20 (3): 210–15. <https://doi.org/10.1055/s-0030-1248045>.
- Cerquiglini, A., J. Henckel, H. Hothi, et al. 2018. "3D Patient Imaging and Retrieval Analysis Help Understand the Clinical Importance of Rotation in Knee Replacements." *Knee Surg Sports Traumatol Arthrosc* 26:3351–61. <https://doi.org/10.1007/s00167-018-4891-9>.
- De Valk, E. J., J. C. A. Noorduyn, and E. L. A. R. Mutsaerts. 2016. "How to Assess Femoral and Tibial Component Rotation after Total Knee Arthroplasty with Computed Tomography: A Systematic Review." *Knee Surg Sport Traumatol Arthrosc* 24:3517–28. <https://doi.org/10.1007/s00167-016-4325-5>.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. CRC press. <https://doi.org/10.1201/b16018>.
- Hiranaka, T., Y. Suda, A. Saitoh, M. Koide, A. Tanaka, A. Arimoto, T. Fujishiro, and K. Okamoto. 2022. "Infographic: Three Key Elements of Kinematic Alignment Total Knee Arthroplasty for Clarified Understanding of Its Approaches." *Bone Joint Res* 11 (4): 226–28. <https://doi.org/10.1302/2046-3758.114.BJR-2021-0543.R1>.
- Hirschmann, M. T., P. Konala, F. Amsler, F. Iranpour, N. F. Friederich, and J. P. Cobb. 2011. "The Position and Orientation of Total Knee Replacement Components: A Comparison of Conventional Radiographs, Transverse 2D-CT Slices and 3D-CT Reconstruction." *J Bone Joint Surg Br* 93 (5): 629–33. <https://doi.org/10.1302/0301-620X.93B5.25893>.
- Holme, T. J., J. Henckel, J. Cobb, and A. J. Hart. 2011a. "Quantification of the Difference between 3D CT and Plain Radiograph for Measurement of the Position of Medial Unicompartamental Knee Replacements." *Knee* 18 (5): 300–305. <https://doi.org/10.1016/j.knee.2010.07.010>.
- . 2011b. "Quantification of the Difference between 3D CT and Plain Radiograph for Measurement of the Position of Medial Unicompartamental Knee Replacements." *Knee* 18 (5): 300–305. <https://doi.org/10.1016/j.knee.2010.07.010>.
- Jamali, A. A., J. P. Meehan, N. M. Moroski, M. J. Anderson, R. Lamba, and C. Parise. 2017. "Do Small Changes in Rotation Affect Measurements of Lower Extremity Limb Alignment?" *J Orthop Surg Res* 12 (1): 77. <https://doi.org/10.1186/s13018-017-0571-6>.
- Jazrawi, Laith M., Luther Birdzell, Frederick J. Kummer, and Paul E. Di Cesare. 2000. "The Accuracy of Computed Tomography for Determining Femoral and Tibial Total Knee Arthroplasty Component Rotation." *The Journal of Arthroplasty* 15 (6): 761–66. <https://doi.org/10.1054/arth.2000.8193>.
- Kruschke, J. K. 2018. "Rejecting or Accepting Parameter Values in Bayesian Estimation." *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245918771304>.
- Linde, M., J. N. Tendeiro, R. Selker, E. J. Wagenmakers, and D. van Ravenzwaaij. 2021. "Decisions about Equivalence: A Comparison of TOST, HDI-ROPE, and the Bayes Factor." *Psychol Methods*, November. <https://doi.org/10.1037/met0000402>.
- Lotke, P. A., and M. L. Ecker. 1977. "Influence of Positioning of Prosthesis in Total Knee Replacement." *The Journal of Bone and Joint Surgery. American Volume* 59 (1): 77–79. <https://doi.org/10.2106/00004623-197759010-00013>.
- Lustig, S., E. Sappey-Marini, C. Fary, E. Servien, S. Parratte, and C. Batailler. 2021. "Personalized Alignment in Total Knee Arthroplasty: Current Concepts." *SICOT J* 7:19. <https://doi.org/10.1051/sicotj/2021021>.
- Mahoney, O., T. Kinsey, N. Sodhi, M. A. Mont, A. F. Chen, F. Orozco, and W. Hozack. 2022. "Improved Component Placement Accuracy with Robotic-Arm Assisted Total Knee Arthroplasty." *J Knee Surg* 35 (3): 337–44. <https://doi.org/10.1055/s-0040-1715571>.
- Miura, M., S. Hagiwara, J. Nakamura, Y. Wako, Y. Kawarai, and S. Ohtori. 2018. "Interobserver and Intra-Observer Reliability of Computed Tomography-Based Three-Dimensional Pre-Operative Planning for Primary Total Knee Arthroplasty." *J Arthroplasty* 33 (5): 1572–78. <https://doi.org/10.1016/j.arth.2017.12.035>.
- Moreland, J. R. n.d. "Mechanisms of Failure in Total Knee Arthroplasty." *Clinical Orthopaedics and Related Research* 226 (January 1988): 49–64. <https://doi.org/10.1097/00003086-198801000-00010>.
- Nisar, S., J. Palan, C. Rivière, M. Emerton, and H. Pandit. 2020. "Kinematic Alignment in Total Knee Arthroplasty." *EFORT Open Rev* 5 (7): 380–90. <https://doi.org/10.1302/2058-5241.5.200010>.
- Radtke, K., C. Becher, Y. Noll, et al. 2010. "Effect of Limb Rotation on Radiographic Alignment in Total Knee Arthroplasties." *Arch Orthop Trauma Surg* 130:451–57. <https://doi.org/10.1007/s00402-009-0999-1>.

- Rivière, C., S. Lazic, O. Boughton, Y. Wiart, L. Villet, and J. Cobb. 2018. "Current Concepts for Aligning Knee Implants: Patient-Specific or Systematic?" *EFORT Open Rev* 3 (1): 1–6. <https://doi.org/10.1302/2058-5241.3.170021>.
- Rivière, C., J. M. Vigdorchik, and P. A. Vendittoli. 2019. "Mechanical Alignment: The End of an Era!" *Orthop Traumatol Surg Res* 105 (7): 1223–26. <https://doi.org/10.1016/j.otsr.2019.07.005>.
- Rodriguez, R., L. Scholl, M. Thompson, and K. de Souza. n.d. "Automation of Segmentation Using Computer Vision and Deep Learning Algorithms." In .
- Scholl, L. Y., E. L. Hampp, K. M. de Souza, T. C. Chang, M. Deren, Z. C. Yenna, N. Sodhi, M. A. Mont, and G. H. Westrich. 2022. "How Does Robotic-Arm Assisted Technology Influence Total Knee Arthroplasty Implant Placement for Surgeons in Fellowship Training?" *J Knee Surg* 35 (2): 198–203. <https://doi.org/10.1055/s-0040-1716983>.
- Shatrov, J., C. Battelier, E. Sappey-Mariniér, S. Gunst, E. Servien, and S. Lustig. 2022. "Functional Alignment Philosophy in Total Knee Arthroplasty - Rationale and Technique for the Varus Morphotype Using a CT Based Robotic Platform and Individualized Planning." *SICOT J* 8:11. <https://doi.org/10.1051/sicotj/2022010>.
- Slevin, O., A. Hirschmann, F. F. Schiapparelli, F. Amsler, R. W. Huegli, and M. T. Hirschman. 2018. "Neutral Alignment Leads to Higher Knee Society Scores after Total Knee Arthroplasty in Pre-Operatively Non-Varus Patients: A Prospective Clinical Study Using 3D-CT." *Knee Surg Sports Traumatol Arthrosc* 26:1602–9.
- Watanabe, S., R. Akagi, Y. Shiko, et al. 2021. "Intra- and Inter-Observer Reliability of Implant Positioning Evaluation on a CT-Based Three-Dimensional Post-Operative Matching System for Total Knee Arthroplasty." *BMC Musculoskeletal Disord* 22:363. <https://doi.org/10.1186/s12891-021-04228-2>.
- Winnock de Grave, P., T. Luyckx, K. Claeys, T. Tampere, J. Kellens, J. Müller, and P. Gunst. 2022. "Higher Satisfaction after Total Knee Arthroplasty Using Restricted Inverse Kinematic Alignment Compared to Adjusted Mechanical Alignment." *Knee Surg Sports Traumatol Arthrosc* 30 (2): 488–99. <https://doi.org/10.1007/s00167-020-06165-4>.
- Yoshino, K., S. Hagiwara, J. Nakamura, T. Tsukeoka, Y. Tsuneizumi, and S. Ohtori. 2019. "Intra- and Interobserver Reliability and Agreement in Three-Dimensional Computed Tomography Measurements of Component Positions after Total Knee Arthroplasty." *Knee* 26 (5): 1102–10. <https://doi.org/10.1016/j.knee.2019.07.001>.

$$\epsilon = \text{Normal}(0, \sigma)$$

APPENDIX A

In this appendix we outline the mathematical details of the Bayesian hierarchical model used in the analysis.

$$\sigma = \text{Exponential}(\lambda)$$

$$\lambda = 1$$

1. Model Structure

The hierarchical model can be described by the following equation:

$$y = \beta_j + b_{jk} + \epsilon$$

Where:

y = Response for the difference in degrees from radiologist-identified landmarks to the algorithm.

β_j = Coefficient for the radiologist level

b_{jk} = Coefficient for the radiologist level j specific to patient

ϵ = Residual random error

2. Prior Distributions

- For the radiologist-level effect:

$$\beta_j = \text{Normal}(0, \sigma_{\beta j})$$

- For the group-specific effect:

$$\beta_{jk} = \text{Normal}(0, \sigma_{\beta j})$$

$$\sigma_{\beta j} = \text{Exponential}(\tau)$$

$$\tau = 1$$

- For the standard error:

In this model b_{jk} are radiologist and patient random effects, and the coefficient of interest is β_j as it is the common slope or fixed effect for the radiologists. Since a global intercept is omitted in this model, both radiologists have an individual intercept and slope. Due to the hyperprior, the group-specific effects aren't independent due to partial pooling.

3. Region of Practical Equivalence (ROPE)

A ROPE of ± 0.5 degrees centered on zero was employed. The decision criterion was as follows:

- If more than 95% of the joint posterior distribution falls within the circumference of the ROPE, concordance is achieved.
- If more than 95% lies outside, concordance is rejected.
- If the joint distributions overlap the ROPE boundary, a decision is withheld.