Targeted Computational Probabilistic Corroboration of Experimental Knee Wear Simulator: The Importance of Accounting for Variability

M.A.Strickland[1], M.R.Dressler[2], T.Render[2], M.Browne[1], M.Taylor[1]

[1] Bioengineering Sciences Research Group, University of Southampton, Hampshire, SO17 1BJ (UK)

[2] DePuy Orthopaedics, 700 Orthopaedic Drive, Warsaw, IN 46581-0988 (USA)

Abstract

Experimental testing is widely used to predict wear of total knee replacement (TKR) devices. Computational models cannot replace this essential *in-vitro* testing, but they do have complementary strengths and capabilities, which make *in-silico* models a valuable support tool for experimental wear investigations. For effective exploitation, these two separate domains should be closely corroborated together; this requires extensive data-sharing and cross-checking at every stage of simulation and testing.

However, isolated deterministic corroborations provide only a partial perspective; *in-vitro* testing is inherently variable, and relatively small changes in the environmental and kinematic conditions at the articulating interface can account for considerable variation in the reported wear rates. Understanding these variations will be key to managing uncertainty in the tests, resulting in a 'cleaner' investigation environment for further refining current theories of wear.

This study demonstrates the value of probabilistic *in-silico* methods by describing a specific, targeted corroboration of the AMTI knee wear simulator, using rigid body dynamics software models. A deterministic model of the simulator under displacement-control was created for investigation. Firstly, a large sample of experimental data (N>100) was collated, and a probabilistic computational study (N>1000 trials) was used to compare the kinematic performance envelopes for *in-vitro* and *in-silico* models, to more fully corroborate the mechanical model. Secondly, corresponding theoretical wear-rate predictions were compared to the experimentally reported wear data, to assess the robustness of current wear theories to uncertainty (as distinct from the mechanical variability).

The results reveal a good corroboration for the physical mechanics of the wear test rig; however they demonstrate that the distributions for wear are *not* currently well-predicted. The probabilistic domain is found to be far more sensitive at distinguishing between different wear theories. As such we recommend that in future, researchers move towards probabilistic studies as a preferred framework for investigations into implant wear.

Introduction

Pre-clinical evaluation of implants is an essential part of orthopaedic design, and historically experimental (*in-vitro*) methods have been the mainstay of this testing. Subsequently, computational (*in-silico*) methods have been widely adopted to analyse a range of factors traditionally investigated experimentally. At first consideration it may seem an attractive option to substitute, rather than supplement, *in-vitro* tests with less resource-intensive *in-silico* testing. Computational models can provide a fast, high-volume simulation capability for interpolating within a known design-space (especially if some areas are difficult to explore experimentally), or extrapolating where the physics of the test domain are well understood. However, when the test domain is poorly understood, or phenomenologically novel effects are being investigated, experimental testing provides an essential grounding in real-world physics (as the governing laws are implicit, so need not be explicitly defined). These two different approaches are therefore *complementary*, and not *competitive*.

A pertinent example is wear-testing. Ultrahigh molecular weight polyethylene (UHMWPE) wear can result in osteolysis [1], so considerable effort is invested in wear testing (e.g. material tests [2, 3] and implant tests [4, 5]). The tribology of UHMWPE is not fully understood, so a purely

theoretical approach is not possible. Consequently, *in-vitro* testing is still essential; unfortunately, these tests are time-consuming and require manually-intensive measurements. It is therefore desirable to maximise the added value from experimental testing, and *in-silico* models can facilitate this; over the last two decades various *in-silico* wear models have emerged (e.g. [6-8]). Computational models can be used to ensure that the rig mechanics are fully characterised (so that inputs for the theoretical wear model correspond to the real-life test). Fast *in-silico* studies also provide a means to perform high-volume simulation, e.g. for probabilistic or 'design-of-experiment' screening studies where multiple input factors are investigated simultaneously.

Various theories of wear have been proposed, from the first-generation concepts presented by Archard [9] to the orientation-sensitive 'crossing-shear' (CS) models proposed by Wang [10], and confirmed by Turell et al [11], with further adjustments proposed by investigators including Vassiliou and Unsworth [12], Ernsberger et al [13], Willing and Kim [14], Hamilton et al [15] and Mazzucco and Spector [16]. These theories have been shown to have limited, but nonetheless useful, predictive power in deterministic comparison studies [17].

However, a theoretical wear model can only yield useful results if the correct kinematics and kinetics can be reproduced, necessitating a well-corroborated mechanical model. Early *in-silico* finite-element studies were often based on limited experimental data, with little or no attempt to reproduce the dynamics of the system beyond the implant interface (e.g. [18, 19]). This does not provide adequate 'corroboration' of the true complexities of a typical servo-hydraulic or pneumatic knee simulator. The dynamics of the rig (friction, damping, inertia, pliancy, etc) must be quantitatively measured and explicitly modelled if results are to be meaningful. The 'target' *in-vitro* platform in this study has been simulated by various researchers, including Zhao et al [20], and more recently Lanovaz et al [21, 22]. The more recent reported models make use of

‘force-feedback’ from the simulator’s six degree-of-freedom load-cell, giving greater confidence in the corroborated results. This can also reveal important details about the simulator mechanics (e.g. the femoral flexion arm cannot be treated as a strictly rigid rotating hinge, due to pliancy in the assembly). With careful modelling, any given individual experimental test may be reproduced.

However, a cursory analysis of any cohort of experimental data reveals too much variability between stations and between machines for such individual deterministic comparisons to ‘corroborate’ a model. At best, a single deterministic trial might be found to lie within the bounds of a ‘performance envelope’ of the various results. Just as a ‘one-on-one’ deterministic corroboration is not adequate, so this ‘one-on-many’ form of probabilistic corroboration (e.g. [23]) is inadequate to describe an *in-silico* model as ‘validated’. Instead, variability effects for *both* the experimental and computational models must be compared: i.e. a ‘many-on-many’ corroboration, comparing the respective performance envelopes *in-vitro* and *in-silico*.

Therefore the present study has two parts:

1. Firstly, a probabilistic model of the simulator mechanics will be corroborated for displacement-driven operation, using large samples of both experimental and computational data. This will give confidence in the mechanical model performance, laying the foundations for subsequent wear-modelling, and allowing a distinction to be made between uncertainties associated with mechanical modelling, and uncertainties associated with wear prediction.

2. Next, this model will be used as the baseline for an assessment of the performance of contemporary wear prediction theories, to assess how well these algorithms capture not only the basic ‘mean’ wear rates, but also the associated distribution of wear values.

The resulting analysis gives a more holistic perspective of the wear algorithms' performance, thereby informing the ongoing debate regarding the most appropriate mathematical description of UHMWPE wear. It was hypothesised that, although a reasonable correlation could be achieved for the mechanics, there are greater uncertainties associated with probabilistic wear analysis, and this would result in a reduced predictive capability.

Part 1. Mechanical Corroboration

Experimental

Kinematic and kinetic data were collated from five long term (5MCycles or more) multi-station wear tests with nominally equivalent test conditions (N=179 intervals total). All tests used a four-axis, six-station knee wear simulator (AMTI, Watertown, MA). Femoral components (CoCr posterior cruciate retaining medium-sized right) were secured in a consistent position by using bespoke fixturing and jigs, using custom Delrin adapters and a thin mantle of acrylic cement, with the centre of flexion located according to ISO 14243-3 [24]. Tibial trays (CoCr cruciate retaining, fixed bearing, medium sized) were first modified to expose the insert locking mechanism to facilitate disassembly of the insert for subsequent cleaning and analysis (with no perceivable deleterious effect to function) and were then mounted to a lightweight platen also using acrylic cement. The trays were secured in the center of a platen using bone cement while held in the proper location using a detachable custom jig that was removed prior to testing - the observed variability was on the order of ≈1mm / ≈1° for translational / rotational positioning (dependent upon the operator and the components under test). Tibial inserts (cruciate retaining, fixed bearing, medium size and thickness, comprised of UHMWPE GUR1020 gamma-irradiated

(40 KGy nominal) in vacuum) were soaked in reverse osmosis filtered water for a minimum of 28 days prior to testing.

Identical inputs for vertical load, flexion-extension, internal-external (IE) rotation and anterior-posterior (AP) displacement were employed in all tests using profiles similar to those reported by Barnett et al [25] to produce a 'high kinematics' gait cycle at 1Hz. Motions in the medial-lateral (ML) and varus-valgus (VV) directions were uncontrolled.

Force-feedback from the load cell was logged automatically for each individual station (F_X, F_Y, F_Z and M_X, M_Y, M_Z components, defined as shown in Figure 1, sampled at 1kHz through the gait cycle midway through each testing interval).

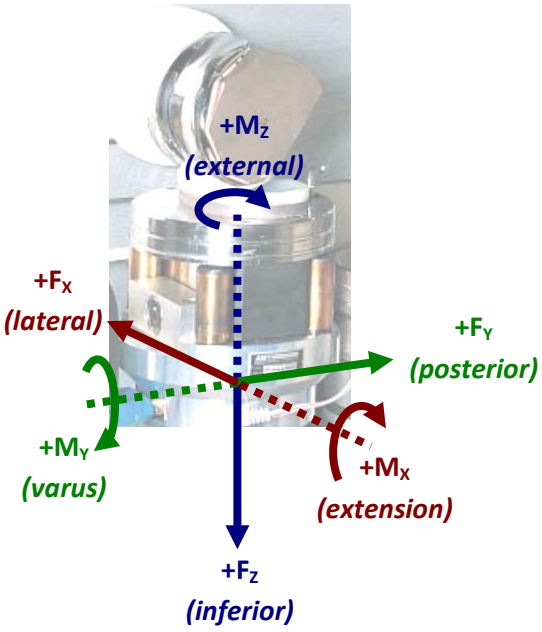

Figure 1. Polarity of applied force-components reported for the tibial load-cell on the AMTI simulator (nearest equivalent anatomic directions shown for 'right' knee under test).

Computational

The computational model was based on a modified version of extant TKR models developed using MSC.ADAMS (MSC Software, CA), and previously tested against other experimental data (extensive details on the model used are included in this earlier study [26]). The models use rigid-body dynamics, with 'spring bed' elements to simulate 'soft' contact (contact is modelled as non-linear, but elastic; creep-deformation was excluded from this variant of the model, to avoid the need for adaptive iterations) – spring properties were 'tuned' based on previous studies [18]. A single non-adaptive cycle was used for modelling, although two cycles were simulated so that any 'transient' settling effects in the first cycle could be discarded. For this study the model was specifically updated to represent the AMTI *in-vitro* knee simulator (Figure 2). The model scope was extended to include relevant rig fixtures, component positioning was updated with respect to the axes of rotation and limited pliancy was included for both the femoral axes (similar to the work of Lanovaz et al [21, 22]) and tibial axes. Appropriate degrees of freedom were established using combinations of primitive motion constraints (rotational and translational). The dimensions of parts were measured directly from the experimental rig, and inertial properties were based on geometry and appropriate assigned material properties, with representative values for dynamics (friction, damping), derived by estimates from the rig, and historical data in the literature [18]. Because so many trials are needed for a stochastic corroboration, the model is necessarily of lower integrity (than for example a fully deformable, non-linear FEA simulation); this performance-accuracy compromise is often an important limitation of such stochastic studies.

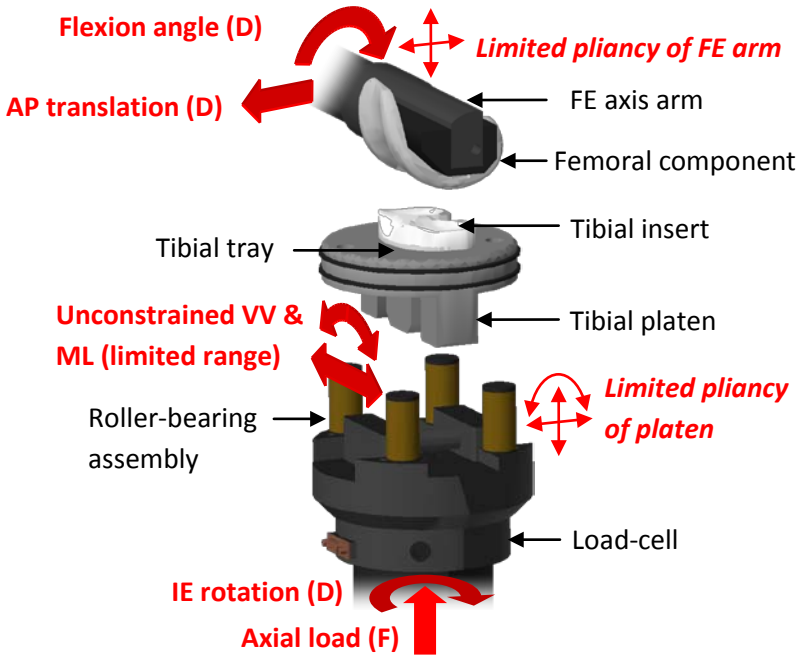

Figure 2. Schematic of the AMTI model *in-silico*: (D) designates displacement-driven axis; (F) designates force-driven axis. Note the explicit inclusion of pliancy within both the 'femoral' and 'tibial' assembly.

For corroboration purposes, the load cell measures (F_X, F_Y, F_Z and M_X, M_Y, M_Z) were recreated *in-silico*, to compare with experimental load-cell data. Force-feedback for a handful of individual trials was checked to ensure that the fixed flexion axis and the AP 'dwell' position of the femur with respect to the tibia matched the experimental tests (even small errors <1mm can otherwise produce significant deviations in the measured force-feedback).

Probabilistics

In this pilot study, seven factors were selected for stochastic variation (Table 1). This is far from an exhaustive list of possible variables; many others could be included (material properties,

controller performance, etc), but this sub-set of factors had been found to be influential in internal developmental sensitivity studies. These included mal-positioning, friction and dwell-point errors. The variables are assumed to be independent, as in previous studies [23] (currently unpublished internal data supports this modelling assumption). In the absence of specific data, generic Gaussian (Normal) distributions were used for most factors (cropped at ±3σ); except those for which negative values are invalid - Lognormal distributions were used instead (cropped at +3σ). The 'mean' values represent the intended (ideal) component positioning. Mean friction values are based on tuned levels from deterministic studies. The SD for positioning is an empirical estimate, based on engineering judgement and verified by inspection of force-feedback from a limited number of trials. Monte-Carlo analysis with 1,200 trials was used to produce variability 'envelopes' for force-feedback.

Factor	**Description**	**Distribution**	**Mean (μ)**	**SD (σ)**
AP_Dwell	Initial A-P dwell offset	Normal	12.5mm	0.5mm
Fem_FE	Femoral F-E mal-rotation	Normal	0°	0.5°
Fem_IE	Femoral I-E mal-rotation	Normal	0°	0.5°
Fem_VV	Femoral V-V mal-rotation	Normal	0°	1°
Tib_ML	M-L offset of insert on platen	Normal	0mm	0.5mm
TF_μ	Tibiofemoral contact friction	Lognormal	0.01	0.02
Roll_μ	Roller-bearing friction	Lognormal	0.02	0.01

Table 1. Variables used for the probabilistic study, with distribution parameters.

Results

The experimental data revealed considerable variability, even with no deliberate alteration made to any test parameters. The variability in experimental force-feedback is illustrated in Figure 3, with peak standard deviation (SD) for some axes as high as 45% of the peak absolute cyclic range (e.g. for the ML force, F_X). This emphasises the fact that deterministic studies are inadequate for accurate corroboration work; the observed variability must be reflected within *in-silico* modelling. Figure 3 also shows the simulated *in-silico* model results. The force-feedback corresponds well between the *in-vitro* and *in-silico* model, particularly on the controlled axes. It is clear, however, that there are still discrepancies; in particular, the experimental variability never approaches zero, whereas for the computational data, the controlled channels (in particular F_Z, the axial force) exhibit almost zero variability. Investigation revealed that this was not due to controller tracking; some load-cells would intermittently exhibit a fixed 'offset' throughout the cycle. This appeared to be an irregular 'fault' behaviour, rather than standard operating variability, but was within the 5% full scale tolerance specified by the manufacturer – therefore no attempt was made to replicate this effect *in-silico*. (This apparent mal-tracking would only be influential if it occurred on the 'master' station used for controller force-feedback – but even then, the offset would be applied consistently across the 'slaved' stations). Nonetheless, even accounting for this effect it is equally clear from the other channels (especially the transverse forces F_X and F_Y) that the *in-vitro* variability envelope thickness changes considerably across the gait cycle, and as such cannot be adequately explained by a simple 'measurement offset' (i.e. the other sources of variability - which *have* been explicitly modelled - are dominating). A sensitivity analysis of the variables under test suggests that, while results vary for different measures, certain parameters are consistently more influential. As a result, we identify two

important elements in reducing the experimental variability: accurate AP ‘dwell’ positioning of components, and careful control of friction effects within the tibial roller-bearing assembly.

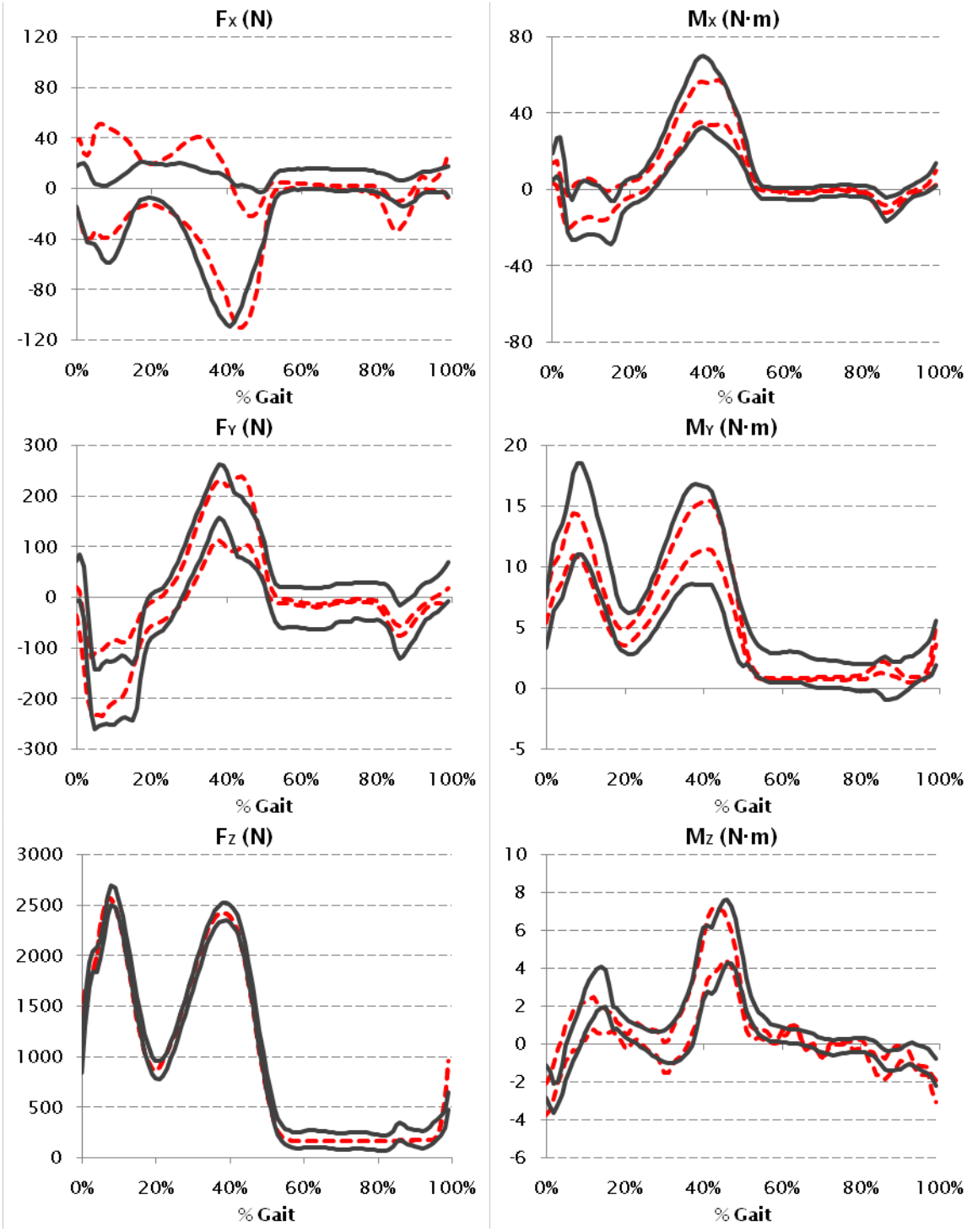

Figure 3. Force feedback for *in-vitro* (solid) and *in-silico* (dashed) probabilistic results. Envelopes are for mean value ± 1 SD.

Part 2. Wear Corroboration

Experimental

Tibial inserts were fully submerged in bovine calf serum (Hyclone Laboratories, Logan, UT) maintained at 37±2°C via recirculation. The serum was treated with sodium azide at a concentration of 0.2% mass fraction to retard bacterial growth and with EDTA at a concentration of 20mM (7.45 g/L) to prevent calcium precipitates. Femoral/tibial component pairs remained the same for the duration of the test but switched banks every 1.5Mcycles.

Wear was determined gravimetrically using methods similar to ISO 14243-2 [27]. Tibial inserts were weighed on an analytical balance (0.01 mg resolution, XP205, Mettler-Toledo, Columbus, OH) prior to testing and then every 0.5Mcycles, corresponding to the lubricant change intervals. Loaded soak controls were used to account for apparent weight gain while in serum. Wear rates were calculated using linear regression of the compensated wear.

Computational

All wear post-processing was performed internally within MSC.ADAMS. Wear algorithms were included based on a number of standardised extant theories: first generation ‘Archard’ and second generation ‘cross-shear’ models, and proposed alternative models discarding the contact pressure (CP) term in the original algorithms. Three key alternative models are defined in Table 2. Note that, owing to the computational overhead of a probabilistic study, multi-step ‘adaptive’

algorithms (as in [28]) were not used for this study. This is an important limitation, since over longer time-scales creep and surface adaptations would impact on both test mechanics and wear rates; however, this is considered a tolerable performance-accuracy compromise for this pilot corroboration study.

Wear model	***Formulation***	***Wear factor, k (from [17])***
Archard	$W = k_1.CP.s$	2.0×10^{-7} mm³.N^{-1}.m^{-1}
'Cross-shear' (A/A+B)	$W = k_2(CS).CP.s$ $k(CS) = k_2 \cdot \frac{\sum\lvert A\rvert}{\sum\lvert A\rvert + \sum\lvert B\rvert}$	3.3×10^{-6} mm³.N^{-1}.m^{-1}
Modified cross-shear (A/A+B)*	$W = k_3(CS).s$ $k(CS) = k_3 \cdot \frac{\sum\lvert A\rvert}{\sum\lvert A\rvert + \sum\lvert B\rvert}$	1.8×10^{-5} mm³.m^{-1}

Table 2. Summary of alternative wear algorithms: W is wear-depth per unit contact area; CP is contact pressure; s is sliding distance. 'Wear factor', *k*, varies for different models (indicated by different subscripts). For 2nd generation models, *k* is a function of cross-shear (CS); for the example shown, 'B' is principle sliding direction; 'A' is perpendicular direction.

Importantly, the values for '*k*' in this study are deliberately given fixed deterministic values, based on 'generic' values in the literature [17], and are *not* tuned to this particular study, or varied as part of the probabilistic analysis. This means that any observed discrepancy between

variability *in-silico* (based on mechanical variability) and variability *in-vitro* will provide an indication of how much variability is not captured by the model – i.e. the inherent uncertainties in the prospective (rather than retrospective) *wear prediction* capability of the model, as distinct from the *kinetics and kinematics* of the underlying mechanical model.

The results are therefore expected to be different, since the *in-silico* model is *not* attempting to capture additional variability in the wear-assessment procedure. This difference between the computational and experimental model might then be interpreted in one of two ways:

- In the short-term, and most simplistically, it may be interpreted as representing the SD that should be applied to the wear-factor '*k*' values when using these 1st and 2nd generation models in probabilistic studies. In other words, the observed shortcomings are simply 'bundled' into a variability term in the wear constant. This would require no further modelling, but offers no detailed insight into the nature of this uncertainty.

- Ultimately, it is better to interpret this difference as representing the limitations of the predictive power of current models (i.e. the potential 'room for improvement' in future next-generation algorithms). Efforts could then be made to increase this predictive power, by including specific mechanistic or phenomenological terms in the wear model which represent known physical effects, hence improving predictive power in a more causally explicable manner.

Results

Figure 4 shows a comparison of wear rate distributions. The first pertinent observation is that the experimental results show a very wide distribution. The SD is ≈36% of the mean value; although this might seem to imply that gravimetric wear assessment is a relatively poorly conditioned metric, it should be remembered that this SD encompasses time-variations in the test conditions (bedding-in and long-term surface adaptation). The second point of note is the apparent distribution shape; the distribution appears to follow a relatively ‘Gaussian’ profile; however, there is limited evidence of a degree of skewness, which may imply that an asymmetric distribution type (e.g. lognormal, Weibull, etc) might be a better descriptor of the variability. Unfortunately, the data source here is too coarse to be more specific; a higher ‘N’ value (i.e. more experimental trials) would be needed to select a specific distribution & associated parameters with any statistical confidence.

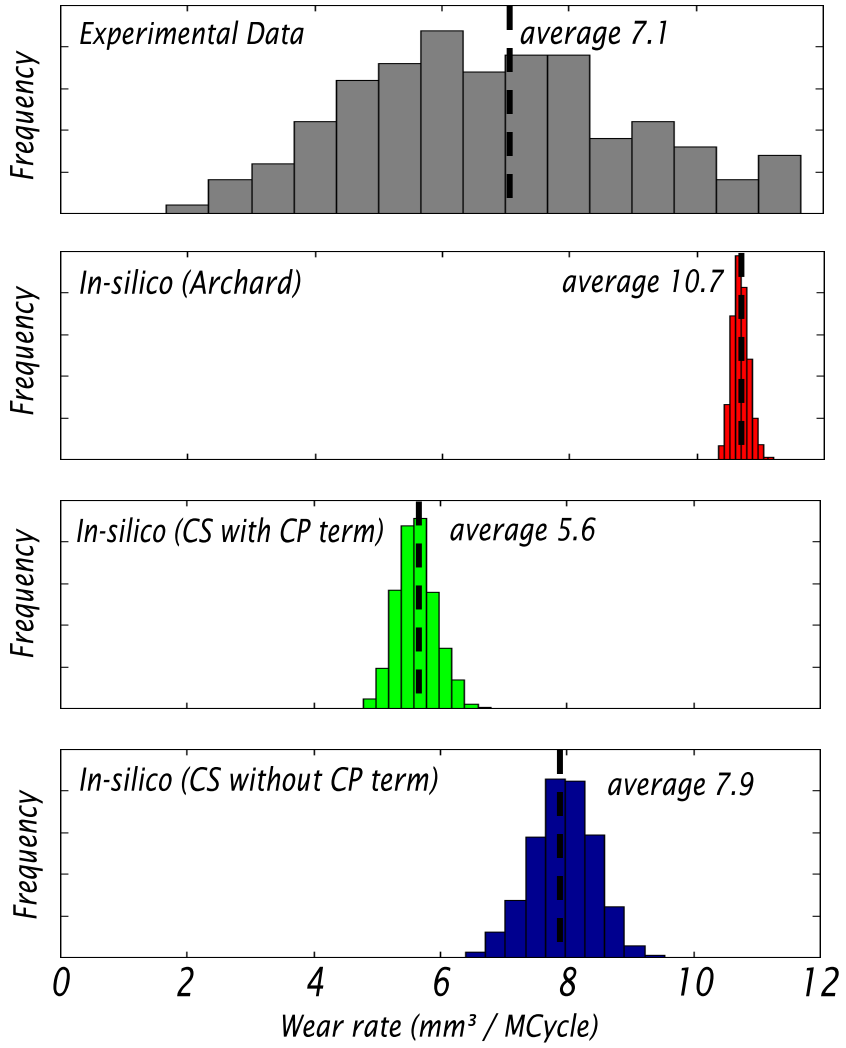

Figure 4. Wear rates for *in-vitro* data (top), and three different variations of *in-silico* algorithm (below).

Nonetheless, even with the limited number of trials sampled here, it is clear that there is a large disparity between the observed variability, and the variations predicted by current *in-silico* methods. The first-generation Archard model is highly insensitive to the permutations, with SD only ≈1% of the mean wear rate. The second-generation models with some form of 'cross-shear' metric predict more variability (SD ≈6% of the mean). However, this is still much lower than the experimental value. It would, of course, now be possible to assign a SD to the wear factor, *k*,

based on the results of this study. (Based on this data-set, a value of approximately 30% of the ‘mean’ value would seem to be necessary - more than an order of magnitude higher than values used in previous studies based on engineering judgement [29]). However, there are two problems with this. Firstly, it is not prudent to suggest a value based on this ‘training set’, until it has been independently tested using a separate unrelated data-set (this would require further probabilistic corroboration work). Secondly, combining all this variability into one ‘catch-all’ SD term on the wear factor fails to recognise that this is a complex combination of different effects; the overall variability is probably partly due to genuine variability in wear rate (e.g. due to local variations in lubrication, surface morphology, etc), partly due to variability in the manual gravimetric assessment process, and partly due to the predictive-power shortcomings of the (imperfect) algorithms presently used for *in-silico* wear prediction. It is better to treat these factors separately, and attempt to reduce variability where possible, rather than accepting this aggregate uncertainty margin as an inevitable ‘constant’ SD.

The different models predict different shapes of distribution; for example, the ‘cross-shear’ model with CP included has a more negative skew, whereas this skew is less evident if the CP term is excluded. (This example is intuitive, since mal-positioning will lead to decreased conformity (and hence contact area) but increased contact-pressure; if wear is modelled as independent of CP then the reduced area is enough to reduce wear, if however wear increases with CP the two effects will be antagonistic and lower wear outcomes will be less prevalent).

Discussion

This study aimed to demonstrate two key principles of pre-clinical analysis; firstly, the importance of closely-corroborated experimental and computational methods, and secondly the necessity of applying probabilistic methods to account for variability.

Historically, little attention has been paid to achieving high-integrity corroboration of *in-silico* models. Often, no attempt has been made to compare multiple sources of data for the tests. This 'multi-dimensional' corroboration is important, since with *in-silico* models it may always be possible to inadvertently 'tune' internal model parameters to match any single observed measurements. This is less likely if there are more 'known' data-points to compare to (i.e. more different independent measurement channels) than there are 'unknown' internal variables.

Therefore, every effort should be made to design *in-silico* models targeted at specific *in-vitro* test platforms, for which there is ready access to multiple streams of experimental data. This closer corroboration is fundamental to the utility of the model for any other subsequent research purposes.

Equally, experimental tests 'targeted' for corroboration must be of the highest quality and provide meaningful insight into *in-vivo* performance, with every effort made to avoid unintentional 'artefacts' influencing the *in-vitro* results - tandem *in-silico* simulation can be a useful investigatory tool for better understanding the 'true' behaviour of experimental tests.

It is clear from the present experimental data that even though this is a well-controlled *in-vitro* wear test, nonetheless there is a large degree of innate variability. The complexity of TKR mechanics, the magnitude of input loads and the positional accuracy coupled with implant sensitivity to mal-positioning all contribute to this envelope of uncertainty; whilst every effort should be made to reduce such factors, they cannot be eliminated. Therefore this variability must

be accounted for by any *in-silico* model attempting to reproduce experimental behaviours. This makes stochastic methods an essential part of future pre-clinical development.

In this pilot study, the displacement-driven mechanical forces of the AMTI simulator could be predicted successfully. This is similar to the work of Lanovaz et al [21, 22] and confirms findings from that study (e.g. flexion axis pliancy must be modelled), but the degree of corroboration is advanced by predicting complete variability envelopes (rather than one individual trace within the envelope bounds). This demonstrates that although the AMTI simulator is not entirely deterministic due to inherent uncertainties, the sources of *mechanical* variability can be identified and quantified with reasonable confidence. As such they could be addressed, or at least accounted for in any analysis. Of the residual variability which was not accounted for, errors in the load-cell values are believed to predominate. Feedback data from individual stations occasionally featured a notable 'offset' deviation; this offset was often constant throughout the gait cycle and could be as high as ≈400N (for the axial load channel, F_Z; the offset was proportionately lower on the other channels with reduced magnitudes). These 'sensor errors' may not be relevant to the mechanical model since they are effectively 'downstream' of the actual mechanical articulation, unless there is a closed feedback loop using the sensor data (e.g. for force-controlled axes) which may then affect the mechanics at the interface. These sensor errors could be included within the *in-silico* model; however the errors were sporadic, not following a continuous distribution, and occurred disproportionately on certain stations. This implies that they may be specific to the test machine used, and so including them may make the *in-silico* model specific to this one rig. There is an important trade-off to be made between specialisation and generalisation with these models; specialisation offers greater accuracy at the cost of broader applicability. The appropriate trade-off will depend on the

intended applications of the model (in this case, this ‘fault’ behaviour should perhaps be addressed experimentally, rather than incorporated in the computational domain). A trade-off must also be made between level of detail, and the overall resources invested into the modelling effort. In the present case, where uncertainty in the wear processes dominates greatly over uncertainty in the actual mechanics, there is little benefit in further refining the underlying mechanics at this stage, until our understanding of wear has improved. Finally, a trade-off is made between generality and specificity when choosing wear factors; a particular mean and SD could be chosen to better fit the observed distributions in this study, but conversely tuning to one set-up may compromise the general predictive power of the model against other unrelated tests.

These results illustrate the limits of current wear prediction capabilities. If only deterministic results are considered, it is easy to tune a wear factor ‘constant’ to any given result, or small set of results. However, the probabilistic approach reveals that the tuning of these factors may be ‘masking’ true performance; the first- and second-generation wear theories fail to predict anything like the reported variability distribution. The first-generation Archard model is the most deficient, adding weight to the consensus that this model has limited predictive power for UHMWPE. However, second-generation models also heavily under-predict the distribution. There were differences between the distributions predicted by different second-generation algorithms (e.g. with and without CP). Theoretically, these differences could be used to discriminate between the different algorithms and rank them in terms of how well they match the experimental distributions. However, at present this is not possible due to the large disparity between the *in-vitro* and *in-silico* SD magnitude. It would require a much more robust set of models, accurately accounting for the true stochastic nature of wear, to make such comparisons justifiable.

Even if these stochastic wear rates were accurately predicted, they represent only a simplified ‘aggregate’ metric; in reality, wear depth is a function of surface location, resulting in a complex 2D mapping of wear rate. Ultimately, experimental surface scanning would need to be compared to computational contour maps. However, the additional resolution this more complex comparison would provide is not yet necessary, since even with the simple ‘wear volume’ metric the shortcomings of the *in-silico* methods are readily apparent.

It is important to recognise that the reported *in-vitro* wear variability is a product of two distinct sources: actual variability in the wear rate, and the compound effect of errors in the gravimetric wear measurement process (e.g. calibration errors, load-soak control errors, manual-process uncertainty). This second class of ‘measurement’ errors presently has no equivalent in the *in-silico* model domain; however it is known that the wear measurement process is difficult and prone to inaccuracies, due to the proportionately small amount of debris produced. This uncertainty is a potential obstacle to obtaining better data for refining present wear theories. Whilst every effort should be made to minimise sources of error, this important error source may need to be factored into future *in-silico* predictions when comparing to *in-vitro* data. Again, a trade-off must be made; should the *in-silico* model reflect the measurement errors from one specific source of *in-vitro* data, or attempt to generalise to reflect more broadly applicable ‘typical’ experimental error? Once again, the answer to this question depends upon the envisaged model applications.

The present study has raised a number of important issues regarding variability in knee-wear testing, and the role of *in-silico* corroboration models. Whilst there are undeniably many challenges remaining for accurate pre-clinical wear prediction, this study raises the bar for computational-experimental corroboration. We have demonstrated that by effective collaboration

and data-rich multi-channel corroboration, a more robust computational model can be created leading to a better understanding of the subtleties and sensitivities of the experimental rig – in turn enhancing both *in-silico* and *in-vitro* capabilities. We have used probabilistic methods to provide a more comprehensive corroboration than has been demonstrated previously, illustrating that our understanding of knee simulator mechanics is good, but our understanding of wear is less mature. Future studies should combine this strong emphasis on corroboration with a probabilistic study framework to begin addressing this knowledge-gap, with the ultimate aim of delivering superior pre-clinical wear prediction capabilities.

Acknowledgements

Support for this research was provided by the UK Engineering & Physical Sciences Research Council (EPSRC), and DePuy, a Johnson & Johnson company.

References

1. Casey, D., et al., *PFC Knee Replacement: Osteolytic Failures From Extreme Polyethylene Degradation.* Clin Orthop Relat Res, 2007.
2. Besong, A.A., et al., *Quantitative comparison of wear debris from UHMWPE that has and has not been sterilised by gamma irradiation.* J Bone Joint Surg Br, 1998. **80**(2): p. 340-4.
3. Walker, P.S., *Friction of Internal Artificial Joints.* Engineering in Medicine, 1972. **1**(4): p. 91-92.
4. Barnett, P.I., et al., *Investigation of wear of knee prostheses in a new displacement/force-controlled simulator.* Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2002. **216**(1): p. 51-61.
5. Walker, P.S., et al., *A knee simulating machine for performance evaluation of total knee replacements.* J Biomech, 1997. **30**(1): p. 83-9.
6. Maxian, T.A., et al., *3-Dimensional sliding/contact computational simulation of total hip wear.* Clin Orthop Relat Res, 1996(333): p. 41-50.
7. Knight, L.A., et al. *Influence of cross shear on the wear of TKA under various kinematic conditions*. in *52nd Annual Meeting of the Orthopaedic Research Society*. 2006. Chicago, USA: Orthopaedic Research Society.

8. Kang, L., et al., *Quantification of the effect of cross-shear on the wear of conventional and highly cross-linked UHMWPE.* J Biomech, 2008. **41**(2): p. 340-6.
9. Archard, J.F., *Contact and Rubbing of Flat Surfaces.* Journal of Applied Physics, 1953. **24**(8): p. 981-988.
10. Wang, A., *A unified theory of wear for ultra-high molecular weight polyethylene in multi-directional sliding.* Wear, 2001. **248**(1-2): p. 38-47.
11. Turell, M., A. Wang, and A. Bellare, *Quantification of the effect of cross-path motion on the wear rate of ultra-high molecular weight polyethylene.* Wear, 2003. **255**(7-12): p. 1034-1039.
12. Vassiliou, K. and A. Unsworth, *Is the wear factor in total joint replacements dependent on the nominal contact stress in ultra-high molecular weight polyethylene contacts?* Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2004. **218**(2): p. 101-107.
13. Ernsberger, C., D. Whitaker, and J. Chavarria, *UHMWPE Wear Rate As A Function Of Contact Area And Stress*, in *Transactions of the 53rd Annual Meeting, Orthopaedic Research Society* 2007: San Diego, CA.
14. Willing, R.T. and I.Y. Kim, *A Pseudo-Qualitative Method for Measuring Cross-Shearing motions in Total Knee Replacements*, in *Transactions of the 54th Annual Meeting, Orthopaedic Research Society* 2008: San Francisco, CA.
15. Hamilton, M.A., et al., *Quantifying multidirectional sliding motions in total knee replacements.* ASME. Journal of Tribology, 2005. **127**(2): p. 280-6.
16. Mazzucco, D. and M. Spector, *Effects of contact area and stress on the volumetric wear of ultrahigh molecular weight polyethylene.* Wear, 2003. **254**(5-6): p. 514-522.
17. Strickland, M.A. and M. Taylor, *In-silico wear prediction for knee replacements - methodology and corroboration.* Journal of Biomechanics, 2009. **42**(10): p. 1469-1474.
18. Godest, A.C., et al., *Simulation of a knee joint replacement during a gait cycle using explicit finite element analysis.* J Biomech, 2002. **35**(2): p. 267-75.
19. Halloran, J.P., A.J. Petrella, and P.J. Rullkoetter, *Explicit finite element modeling of total knee replacement mechanics.* J Biomech, 2005. **38**(2): p. 323-31.
20. Zhao, D., et al. *Can Pin-on-Plate tests be used to predict joint-level wear in Knee Replacements?* in *2006 Summer Bioengineering Conference, June 21-25 2006*. 2006. Amelia Island Plantation, Amelia Island, Florida, USA: The American Society of Mechanical Engineers, New York.
21. Lanovaz and Ellis, *Dynamic simulation of a displacement-controlled total knee replacement wear tester.* Proceedings of the I MECH E Part H Journal of Engineering in Medicine, 2008. **222**(5): p. 669-681.
22. Lanovaz, J., et al., *A Dynamic FE Model to Generate Input Data for a TKR Wear Tester*, in *Transactions of the 52nd Annual Meeting, Orthopaedic Research Society* 2006: Chicago, IL.
23. Laz, P.J., et al., *Probabilistic finite element prediction of knee wear simulator mechanics.* J Biomech, 2006. **39**(12): p. 2303-10.
24. *ISO 14243-3:2004 Implants for surgery. Wear of total knee joint prostheses. Loading and displacement parameters for wear-testing machines with displacement control and corresponding environmental conditions for test*. 2004, The International Organization for Standardization.
25. Barnett, P.I., et al., *Comparison of wear in a total knee replacement under different kinematic conditions.* J Mater Sci Mater Med, 2001. **12**(10-12): p. 1039-42.
26. Strickland, M.A., M. Browne, and M. Taylor, *Could passive knee laxity be related to active gait mechanics? An exploratory computational biomechanical study using probabilistic methods.* Computer Methods in Biomechanics and Biomedical Engineering, 2009. **12**(6): p. 709-20.

27. *ISO 14243-2:2000 Implants for surgery. Wear of total knee joint prostheses. Part 2: Methods of Measurement*. 2000, The International Organization for Standardization.
28. Knight, L.A., et al., *Comparison of long-term numerical and experimental total knee replacement wear during simulated gait loading.* J Biomech, 2007. **40**(7): p. 1550-8.
29. Pal, S., et al., *Probabilistic computational modeling of total knee replacement wear.* Wear, 2008. **264**(7-8): p. 701-707.