Can physical joint simulators be used to anticipate clinical wear problems of new joint replacement implants prior to market release?

John B Medley[1]

[1]Department of Mechanical and Mechatronics Engineering

University of Waterloo

Waterloo, ON

Canada

Tel: 001 519 888 4784 work

Tel: 001 519 616 7748 mobile

Abstract

One of the most important mandates of physical joint simulators is to provide test results that allow the implant manufacturer to anticipate and perhaps avoid clinical wear problems with their new products. This is best done before market release. The present study gives four steps to follow in conducting such wear simulator testing. Two major examples involving hip wear simulators are discussed in which attempts had been made to predict clinical wear performance prior to market release. The second one, involving the DePuy ASR implant systems, is chosen for more extensive treatment by making it an illustrative example to explore whether wear simulator testing can anticipate clinical wear problems. It is concluded that hip wear simulator testing did provide data in the academic literature that indicated some risk of clinical wear problems prior to market release of the ASR implant systems. This supports the idea that physical joint simulators have an important role in the pre-market testing of new joint replacement implants.

Keywords

Hip wear simulator, metal-on-metal hip implants, pre-market testing, correlation of wear rates, clinical wear

Introduction

"Physical joint simulators" is a rather broad topic for a special issue of the Journal of Engineering in Medicine. It could include the cadaveric simulators used to investigate joint loading, contact stress and risk of loosening such as the shoulder simulator described by Giles et al [1]. It could also include the whole joint natural knee simulator that may be used to assess cartilage or meniscal repair interventions as described by Liu et al [2]. However, most physical joint simulators are designed and used to investigate the wear of joint replacement implants that involve the articulation of various combinations of non-biological components. The surfaces of these components can be made from materials such as polyethylene, ceramic or metal. Therefore, the present study provides a focus on these simulators and their role in anticipating clinical wear problems.

Clinical wear problems are caused by the volume of wear particles and the reaction of the body to these particles. It is not enough to design components to achieve low levels of volumetric wear. The number, size, shape and biological reactions to the wear particles must be considered [3-6]. However, in the present study, only volumetric wear rate is considered and it is assumed that most wear simulator investigations try to address the issue of reducing volumetric wear rates. The reaction of the body to wear particles would eventually be considered when looking at the risk associated with a particular volumetric wear rate.

While it may seem that a progression to a rather narrow scientific field is occurring, the study of wear itself is quite complex. For example, when two hip or knee implant components articulate, nano-scale interactions occur over a large macro-scale contact zone. Depending on local conditions (geometry, contact stress, micro-hardness and so on) a wear particle may be produced. This local micro/nano-contact may or may not cause progressive damage that produces further particles. Thus, wear can be a cascading process. Furthermore, a second pair of nominally identical components may articulate without producing as extreme a local contact and may "cascade to" much lower, less progressive wear. As a result, volumetric wear rates are very scattered and abnormally high wear rates can occur unexpectedly for some component pairs.

To address the above issues, simulator wear testing involves the testing multiple specimen pairs of nominally identical features. Statistical comparisons are made and often design features

can be identified that cause higher wear rates. However, when two or more sample groups are compared and wear rate differences are not found, they are frequently interpreted as giving strong support for the idea that wear rate differences do not exist. This, of course, is wrong and leads us to the classic limitation of statistical analysis.

If statistically significant differences are not found, you are left with a conclusion that there is no statistically significant evidence to support either a difference or a lack of difference between the samples. Does the science of statistics have a solution to this problem? Yes, you need more "power" which essentially means you need to have large sample sizes to seek out differences and then, if you do not find differences, you can finally say (with the help of further statistical analysis) with some confidence that there are indeed no differences.

For simulator wear of joint replacement implants the sub-populations that the simulator results are trying to sample are massive and full of uncontrolled, often unknown variables that can influence wear. The only way to address this problem is to build large databases of wear simulator results so that findings of "lack of differences" can be supported by statistical theory that says they are likely to be valid. However, in this process, it may be found that there actually are differences.

This means that in the business of seeking out risks of high clinical wear rates, it is necessary to conduct extensive comparisons and explorations of simulator wear rates. It is never sufficient to test two groups of say six implants of different designs in one "standard" wear test protocol and finding no statistically significant differences in wear rates conclude that their design differences will not influence clinical wear rates.

Now, it is true that not all simulator studies are looking to identify risks. Often, they are explorations of wear behaviour to provide explanations for the in vivo performance of the implants with regards to wear. In other words, they are scientific investigations and they are essentially trying to advance our understanding of clinical wear phenomena. However, new joint replacement designs involving complex contact mechanics, lubrication and new materials are continually emerging from orthopaedic implant manufacturers and start-up companies. In the development of these new designs, simulator wear testing is performed to identify the risk of wear that might be high enough to cause clinical problems when the new product is released. It is this aspect of "physical joint simulators" that the present study seeks to explore. Can a wear simulator be used reliably in new product development to assess the risk of high volumetric wear

that can cause clinical wear problems prior to product release? This issue of pre-market testing is perhaps the most important philosophical question associated with all types of physical joint simulators. If an affirmative answer cannot be supported, the whole physical joint simulator "research industry" begins to collapse and one is left with relying on "trial-and-error" clinical trial approaches for new product development.

The Steps

It is recognized that specifying steps to identify the role of wear simulator testing in new product development is treading on "dangerous ground". However, the following steps are to be considered as general guidelines rather than exact recipes. An illustrative example is eventually presented in order to provide some more specific detail.

A good approach would involve multiple institutions performing a "suite" of very similar multi-implant tests to produce large and comprehensive data bases. Then, it would be necessary to compare data extensively. The whole procedure would probably have to be mandated by regulatory bodies with support from standards associations. This may seem impossible but a group of major automobile manufacturers (including Honda and General Motors) have collaborated in developing a "numerical human model" with finite element analysis and tissue properties provided by extensive multi-institution studies to be used to assess the various designs of safety devices in their automobiles under "vehicle crash" impact conditions [7, 8]. If this level of collaboration can occur in the automotive industry, why not in the orthopaedic industry?

In any case, the proposed steps are as follows.

1. Use wear simulators to establish the average wear rates under ideal (ID) conditions that are likely to be mild for wear. This should be done using prototype implants (or implants that are as close as possible to the new product in design). This was also called for by Fisher [9] in his stratified approach to pre-clinical tribological evaluation. Support for the average wear rates under these ID conditions can be obtained by showing that on average they have less wear than similarly designed implant retrievals. This identification of the average wear rates under ID conditions can be quite difficult and can involve simplifying assumptions, data exploration and correlations.

There is an obvious role for standard testing here because data comparisons are less meaningful when test conditions vary. However, such standard tests should not be "acceptance tests" for product release.

2. Select various clinically relevant (CR+) conditions that are likely to be adverse for wear (and thus a plus sign is included in the acronym). Apply them in simulator tests. These CR+ conditions could include stop-start, microseparation (or microlaterization as it is now called), poor surgical positioning, cup micromotion due to poor fixation and so on. Once again, this was called for by Fisher [9] and he suggested a series of these conditions for simulator wear testing. Sequential duty cycle testing can be very useful in rapidly exploring the wear under various CR+ conditions [10, 11]. Eventually some additional standard tests could be developed that apply CR+ conditions but just well-documented wear testing that avoids "averaging out" the more pertinent high wear sample pairs would be useful in this step.

    Some "device-specific" testing should also be performed in an attempt to anticipate high wear scenarios. For example, studies of cup deformation were performed for the ASR metal-on-metal implants [12, 13] because their original design had thin cup walls in comparison with other similar products but unfortunately no simulator wear studies were published for deformed ASR cups in the academic literature. It is unlikely that standard tests can be developed for device-specific testing and each new design feature of the implant product that is likely to influence wear should be specifically targeted.

3. Identify wear rates under CR+ conditions that suggest the risk of high clinical wear by comparing their wear rates with those under ID conditions and also comparing their wear rates and surface features with those of retrievals of similar implants. Previous simplifying assumptions and correlations may influence these comparisons. However, any identification of "risk" should elicit further study before product release.

4. If the risk levels are judged to be too high, new designs with fewer modifications compared with existing products should be considered. Besides modifications to implant design, surgical procedures and patient selection should be studied if they are likely to cause increases in clinical wear.

Clearly, it is easy to specify these steps and not so easy to follow them effectively. Some general advice is available. According to Fisher [9], the development of new products should move slowly with much more wide-ranging simulator test conditions and careful examination of the tissue response to wear particles of new implant materials. Zywiel et al [14] also called for a slow development with close monitoring of early clinical retrievals for new implant products. They also expressed some concern that innovation might be stifled by too rigorous an approval process while, at the same time, pointing out that the approval process for orthopaedic implants is much less rigorous than that faced by drug companies.

Reputation Issues

It is now interesting to consider the "track record" of wear simulator testing in the development of new hip implants over the last 25 years or so. Two major cases for hip wear simulator testing in pre-market new product development can be identified and considered. The first one is not considered in as much detail as the second one.

The first one, then, is the development of highly crosslinked polyethylene (HXPE) acetabular cup liners. This development involved considerable wear simulator testing at various institutions. However, when the first clinical retrievals were examined the articulating surfaces looked very different from those produced in simulator wear testing [15]. This was a distinct failure in wear simulation. It seemed that the CR+ condition of hard particle ingress into the joint space had not been included in the pre-market simulator testing (and it was more common than previously thought). Subsequent wear simulator testing suggested that, although considerable deformation occurred at the surface, the wear of the HXPE actually remained quite low in comparison with conventional polyethylene liners [16]. Thus, HXPE was not withdrawn from the market and they have apparently performed very well up to some 20 years in a wide variety of patients and hip implant designs. So, wear simulator testing had not identified the clinical surface appearance but, fortunately, crosslinked polyethylene liners still performed well in terms of low wear rates. However, there was some decline in the general reputation of physical joint simulator research in the orthopaedic community because the simulator tests did not simulate clinical conditions properly.

The next major case involving the efficacy of pre-market hip simulator testing was in the development of the DePuy ASR metal-on-metal hip implant systems. Without launching into the huge and controversial topic of the ASR problems, the present author thinks that there were some clinical wear issues with the ASR implant systems [17-19]. At a glance, it seems that wear simulator testing had not anticipated the various CR+ conditions that caused high wear in some patients. However, the ASR design was based on and apparently supported by quite extensive hip simulator testing but most of it was with similar products rather than ASR bearings [20]. Were the four steps described above followed? The present author would say that they were to some extent followed. So, at what stage did the simulator testing for the ASR development run into difficulty? The answers to these questions are very important and they are addressed in the subsequent sections of the present paper because it would appear that, once again, the reputation of "physical joint simulators" research had declined in the orthopaedic community.

Illustrative Example

So far in the present study, some general opinions have been given and it could be said that some rather obvious statements have been made. In this illustrative example, a much more specific treatment is given, particularly in describing what could have happened in the four step procedure regarding the identification of the risk of high clinical wear for the ASR implant systems prior to market release. The intent is not to examine why the ASR implants had high wear in some patients or if high wear is the only factor involved in the "higher than expected revision rate at five years" that is mentioned in the August 24, 2010 letter [21] to clinicians in which announced the ASR recall.

The year 2004 is taken as the approximate market release year for the ASR systems (actually it was July 2003 for Australia, March 2004 for the United Kingdom and December 2005 for the stemmed ASR XL in the United States). To address the requirements of steps 1 – 4, a fairly comprehensive "pre-market data set" for hip wear simulator studies of metal-on-metal implants has been obtained exclusively from the academic literature published up to 2004. Also, a second representative (and less comprehensive) "post-market data set" has been obtained exclusively from the academic literature published after 2004. The second post-market data set is only used

to consider whether the risks identified in the pre-market release testing actually manifested into a clinical wear problem. The pre-market data set is sufficient to develop the average wear rates under ID conditions and some considerations of CR+ conditions but unfortunately device-specific testing is not represented.

The pre-market data set is quite extensive but very difficult to formulate in a way that permitted following the above specified four steps. A number of assumptions have been made to get all the data into a single type of correlation that allows a comparison of simulator and retrieval wear rates.

Assumptions

- Let 2 million cycles (Mc) in a simulator be equivalent to 1 year in vivo.

- Estimate in vivo wear rate from simulator wear results by dividing the simulator wear (mm^3) by the number of cycles (Mc) at the end point of simulator testing and then divide this number by 2 to get the in vivo wear rate (mm^3/yr). This ignores the different run-in and steady state wear rates that many implant exhibit in simulator wear testing.

- If volumetric wear is available from retrieval studies, it can simply be divided by the implantation time (yr) to get a wear rate (mm^3/yr). Note that run-in and steady state wear rates are not readily available for implant retrievals.

- Linear wear (which is most often reported in retrieval studies) can be converted to volumetric wear using a correlation based on the academic literature published prior to 2004 (Appendix A).

- Plots of wear rate (V') versus the reciprocal of the effective radius (1/R) with effective radius (R) defined as

$$R = \frac{R_H \, R_C}{R_C - R_H} \tag{1}$$

where R_H = radius of head surface (m)
R_C = radius of cup surface (m)

can be used to obtain "positive" correlations that will identify average wear rates under ID conditions and higher wear rates under CR+ conditions.

The selection of 1/R for the above plots is motivated by its likely influence on lubrication and contact stress. Estimated fluid film thickness increases with R and the average contact stress (when the surfaces are not in motion) decreases with increasing R. This suggests that when 1/R is low, fluid films are thicker and average contact stresses are lower and because of both of these, the wear rate should be lower. Thus, as 1/R increases, the wear rate should increase thus giving a "positive" correlation. It is not essential that 1/R is the only variable influencing the wear rate; it must only capture enough of the wear rate variation to allow the identification of high wear caused by certain CR+ conditions. This will be demonstrated in the subsequent data exploration.

The correlation of wear rate with variables such as 1/R is not exactly new. A number of early studies correlated wear with film thickness estimates that were proportional to R [22, 23] or with R directly [24, 25]. However, these early efforts did not have, or did not identify, testing under CR+ conditions. They had more focus on the general understanding of metal-on-metal hip implant wear rather than attempting to identify the risk of higher clinical wear. As mentioned above, 1/R (rather than say R) is correlated in the present study to allow simple linear fits of the data and thus permits easier identification of groups with higher wear rates.

To provide an overview of the data used in the present study, a single plot with all the studies considered has been assembled (Figure 1) and supporting tables give some details of these studies (Appendix B and C). The data values were obtained directly from tabulated values in some of the studies but mostly using digitizer software (GetData Graph Digitizer 2.26, Digital River GmbH, Cologne, Germany) on plots presented in the various studies. All plots in the present study were developed using specialized software (Grapher 10.5, Golden Software Inc., Golden, CO, U.S.A.).

Both pre-market and post-market studies are included, although the pre-market studies are the only ones needed to follow the previously specified four steps to identify risks of high clinical wear. The data points that are red in Figure 1 are for simulator wear studies conducted with CR+ conditions [26-29, 10, 30-33]. The wear rates under CR+ conditions do not seem to be a strong function of 1/R, given that various high levels of wear rate occur over small ranges of 1/R. However, it may be that CR+ conditions cause cup edge contact wear or other types of wear which once established have wear rates that are not governed by the 1/R value. In addition, retrieval study data are included [34-38, 19] and tend to show higher wear rates than many of the

simulator studies, including those conducted under CR+ conditions. Digital data in tabular form was provided for the Rieker et al [37] study as a personal communication from its first author. While there is considerable overlapping data and trends are difficult to identify, the first step can be performed.

Step 1

Only the pre-market data set is considered and since most metal-on-metal implants had $1/R \leq 0.3$ m^{-1}, the $1/R$ values are clipped at this limit to focus and clarify the presentation. In all graphs of the present study, clipping is only related to $1/R$; the corresponding V' values are not clipped. Also, for some of the subsequent graphs, the $1/R$ clipping is removed because the data correlation is strengthened by considering the full range of $1/R$.

The resulting plot (Figure 2a) shows the data more clearly. Despite considerable overlap of simulator and retrieval data [34, 36, 37], it is apparent that some simulator studies [24, 39, 40-42] at low $1/R$ values have higher wear rates than is suggested by the retrieval data. Assuming that the retrieved implants had been subjected to various CR+ conditions, it is expected that their wear rates should be higher.

On this basis, a number of simulator studies, including two early studies involving the present author [39, 24], are identified as having unrealistic wear rates for the ID conditions apparently applied (Figure 2b). In the two studies involving the present author, implant specimens had some variation in sphericity and initial surface roughness that gave some quite high run-in wear and many of the early tests were of relatively short duration. Consequently, with the assumption that clinical wear rates could be estimated by simply dividing the endpoint wear volumes by cycle time in the conversion process to wear rates tended to produce unrealistically high wear rates. It is not known why the other studies [40-42], that have been judged as unrealistic, had high wear rates, although they (like our two studies) did show considerable scatter in wear rate when plotted against $1/R$.

To make the judgement that the results in these studies [24, 39, 40-42] are all unrealistic is perhaps too harsh a judgement. For example, the data presented in Dowson et al [42] included some ASR implants that, as a sub-group, did have very low wear rates. However, it has been decided that rest of the data presented had unrealistically high wear rates and for this reason the entire study is categorized as unrealistic. Fortunately, other data existed with ID conditions that

had lower more "realistic" wear rates [22, 43-46] which could be used to establish the average wear rate under ID conditions.

The "unrealistic" data is subsequently removed and the remaining "ID condition" data fell within and towards the lower side of the retrieval data. When this ID condition data is correlated over the entire 1/R range, a strong positive correlation results (Figure 3a) as shown below

$$V' = 4.331 \ (1/R) - 0.288 \qquad [R^2 = 0.73] \qquad\qquad (2)$$

where $V' =$ wear rate (mm^3/yr)
$\qquad\quad R =$ effective radius (m)
$\qquad\quad R^2 =$ coefficient of determination

When the retrieval data is added into the correlation, it weakens and the slope increases (Figure 3b) as shown below

$$V' = 8.057 \ (1/R) - 0.593 \qquad [R^2 = 0.09] \qquad\qquad (3)$$

It is noted that the retrieval results of Scott and Lemons [35] have not re-appeared in Figure 3b or in any figure in the rest of the present study because they were obtained for Sivash implants with a constrained head and thus are considered to have a very different tribology from all other metal-on-metal hip implants.

The increase in slope and weakening of the correlation (with R^2 dropping from 0.73 to 0.09) in Figure 3b suggests that many of the retrieved implants had sustained wear under some CR+ conditions. The retrieval implant wear rates, in general, fall above the simulator wear rates under ID conditions. Thus, if the simulator wear rates under CR+ conditions follow the retrieval wear rates, risk can be identified.

Step 2

Wear simulator results from tests with CR+ conditions (that are likely to occur in vivo to varying extents for all patients) are included in the pre-market data set. The CR+ conditions are stop-start motion [26] and microseparation [27-29]. Unfortunately, device-specific testing to

identify anticipated high wear scenarios is not available in the academic literature up to 2004. Thus, the present study cannot compare any wear rates that could have been obtained from such testing. For the ASR implant systems, simulator testing involving deformed cups, microseparation, high angles of cup inclination and various combinations could have addressed this aspect of the requirements of Step 2.

Step 3

This step requires the identification of wear rates under CR+ conditions that suggest the risk of high clinical wear. Thus, the available simulator wear results under CR+ conditions are added in for the range, $1/R \leq 0.3$ m^{-1}, and their resulting higher wear rates can be clearly identified (Figure 4). This identification would have been more distinct if wear rates for each implant had been published for the CR+ condition data under microseparation conditions [27-29]. Also, it is unfortunate that more wear simulator testing under CR+ conditions had not been available in the academic literature before ASR market release in about 2004.

Step 4

Although none of the simulator testing under CR+ conditions used ASR implants and much of the retrieval data was for smaller diameter components, there is a clear indication that the wear rates of metal-on-metal hip implants are quite sensitive to certain CR+ conditions (Figure 4). With the adopted data treatment, some level of risk of high clinical wear is shown by the wear rates found in the CR+ condition simulator studies. Thus, wear simulator testing could have identified risk prior to the 2004 market release.

Clinical Performance of ASR Implant Systems

The ASR implant systems were released to the market in about 2004. Hip wear simulator studies continued, some using ASR implants [31, 47] and both ID [31, 48] and CR+ [30-33] conditions were applied. Also, some new retrieval data was obtained [38] but this did not include

ASR implant systems. Eventually, a study with retrieval data for one of the ASR implant systems was conducted by Lord et al [18].

All of the above data is placed into a post-market data set and plotted. Essentially, a very similar pattern to Figure 4 is revealed except wear rates are all much higher (Figure 5). In particular, the retrieval wear rates for one of the ASR implant systems [18] are very high (up to 96 mm^3/yr).

The wear rates of the ASR retrievals follow the high wear rates of the hip simulator testing under CR+ conditions (involving high inclination angles and/or microseparation) thus confirming that the risk of high clinical wear has been realized for some ASR implants. The post-market data set is not comprehensive and, in particular, there may be more ASR retrieval data available from Campbell et al [49] but they have not yet revealed the component radii and implantation times corresponding to the measured volumetric wear of their retrieved implants. Such data values are needed for the present analysis.

Conclusions

There is a need for the steps described in the present study to be followed to allow wear simulator testing to anticipate and hopefully avoid clinical wear problems. It is important to develop and use correlations in data comparisons to guide the interpretation of simulator wear and implant retrieval results. In the particular case of the ASR implant systems, the hip wear simulator data did provide a warning of the risk of high clinical wear for the ASR implant systems using data from pre-market studies. This provides evidence to support the idea that, in general, physical joint simulator testing can anticipate clinical wear problems of new joint replacement implants prior to market release. Physical joint simulators should be used for this purpose.

Acknowledgement

Conflicts of interest

Over the last 5 years, the author has performed legal consulting for plaintiff law firms in the DePuy ASR litigation. These law firms are Weisman, Kennedy & Berris Co, LPA, Cleveland, OH, U.S.A.; Teplitsky, Colson LLP Barristers, Toronto, ON, Canada; Maurice Blackburn lawyers, Sydney, Australia; and Hendler Lyons Flores, Austin, TX, U.S.A.

References

1. Giles JW, Ferreira LM, Athwal GS, et al. Development and performance evaluation of a multi-PID muscle loading driven in vitro active-motion shoulder simulator and application to assessing reverse total shoulder arthroplasty. *Trans ASME J Biomech Eng* 2014; 136: 121007-1 -121007-10.

2. Liu A, Jennings LM, Ingham E, et al. Tribology studies of the natural knee using an animal model in a new whole joint natural knee simulator. *J Biomechanics* 2015; 48: 3004-3011.

3. Willert H-G, Buchhorn GH, Fayyazi A, et al. Metal-on-metal bearings and hypersensitivity in patients with artificial hip joints. A clinical and histomorphological study. *J Bone Jt Surg* 2005; 87-A: 28-36.

4. Brown C, Fisher J and Ingham E. Biological effects of clinically relevant wear particles from metal-on-metal hip prostheses. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 355-369.

5. Catelas I, Campbell P, Bobyn JD, et al. Wear particles from metal-on-metal total hip replacements: effects of implant design and implantation time. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 195-207.

6. Campbell P, Ebramzadeh E, Nelson S, et al. Histological feature of pseudotumor-like tissues from metal-on-metal hips. *Clin Orthop Rel Res* 2010; 468: 2321-2327.

7. Fice JB, Cronin DS. Investigation of whiplash injuries in the upper cervical spine using a detailed neck model. *J Biomechanics* 2012; 45: 1098-1102.

8. Mattucci SFE, Cronin DS. Strain rate dependent properties of younger human cervical spine ligaments. J *Mech Behav of Biomed Mat* 2012; 10: 216-226.

9. Fisher J. A stratified approach to pre-clinical tribological evaluation of joint replacements representing a wider range of clinical conditions advancing beyond the current standard. *Faraday Discuss* 2012; 156: 59-68.

10. Bowsher JG, Nevelos J, Williams PA, et al. 'Severe' wear change to 'as-cast' and 'double heat-treated' large-diameter metal-on-metal hip bearings. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 135-143.

11. Brandt J-M, Charron K, Zhao L, et al. Calf serum constituent fractions influence polyethylene wear and microbial growth in knee simulator testing. *Proc IMechE Part H: J Engineering in Medicine* 2012; 226: 427-440.

12. Jin ZM, Meakins S, Morlock MM, et al. Deformation of press-fitted metallic resurfacing cups. Part 1: experimental simulation. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 299-309.

13. Yew A, Jin ZM, Donn A, et al. Deformation of press-fitted metallic resurfacing cups. Part 2: finite element simulation. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 311-319.

14. Zywiel MG, Johnson AJ, Mont MA. Gradual introduction of orthopaedic implants: encouraging innovation and minimizing harm. *J Bone Jt Surg Am* 2012; 94: e158(1-5).

15. Muratoglu OK, Greenbaum E, Larson S, et al. Surface analysis of early retrieved acetabular polyethylene liners: a comparison of standard and highly crosslinked polyethylenes. *Ann Meeting of the ORS* 2002; Poster 1029.

16. Bragdon CR, Jasty M, Muratoglu OK, et al. Third-body wear of highly cross-linked polyethylene in a hip simulator. *J Arthroplasty* 2003; 18: 553-561.

17. Langton DJ, Joyce TJ, Jameson SS, et al. Adverse reaction to meal debris following hip resurfacing . The influence of component type, orientation and volumetric wear. *J Bone Joint Surg* 2011; 93-B: 164-167.

18. Lord JK, Langton DJ, Nargol AVF, et al. Volumetric wear assessment of failed metal-on-metal hip resurfacing prostheses. *Wear* 2011; 272: 79-87.

19. Underwood R, Matthies A, Cann P, et al. A comparison of explanted Articular Surface Replacement and Birmingham hip resurfacing components. *J Bone Joint Surg* 2011; 93-B: 1169-1177.

20. ASR XL Heads. *DePuy Brochure* 2005; Cat No: 9998-18-300 Version 1. www.rpa.spot.pt/getdoc/b5b63529-5902-4751-9145-8c3c29944ce7/ASR-xl-heads-(DePuy).aspx

21. Plouhar PL. Urgent information – recall notice. DePuy Orthopaedics. http://www.depuysynthes.com/sites/default/files/DPYUS1%20Recall%20Notice.pdf

22. Chan FW, Bobyn JD, Medley JB, et al. Wear and lubrication of metal-on-metal hip implants. *Clin Orthop Rel Res* 1999; 369: 10-24.

23. Dowson D. Tribological principles in metal-on-metal hip joint design. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 161-171.

24. Chan FW, Bobyn JD, Medley JB, et al. Engineering issues and wear performance of metal on metal hip implants. *Clin Orthop Rel Res* 1996; 333: 96-107.

25. Medley JB. Tribology of bearing materials. *Hip Resurfacing: Principles, Indications, Technique and Results* 2008; HC Amstutz: Saunders Elsevier, 33-44.

26. Chan FW, Bobyn JD, Medley JB et al. Simulator wear of metal-metal hip implants under adverse load conditions. *Ann Meeting of the ORS* 1999; 310.

27. Butterfield M, Stewart T, Williams S, et al. Wear of metal-metal and ceramic-ceramic hip prostheses with swing phase microseparation. *Ann Meeting of the ORS* 2002; 128.

28. Fisher J, Ingham E, Matthews B, et al. Functional biological activity and osteolytic potential of wear debris generated in artificial hip joints. *Engineers & Surgeons – Joined at the Hip. Refining Future Strategies in Total Hip Replacement*. IMechE 2002: C601/011/2002.

29. Williams S, Stewart TD, Ingham E, et al. Metal-on-metal bearings wear with different swing phase loads. *J Biomed Mater Res B Appl Biomat* 2004; 70: 233-239.

30. Williams S, Leslie I, Isaac G, et al. Tribology and wear of metal-on-metal hip prostheses: influence of cup angle and head position. *J Bone Joint Surg* 2008; 90-A: 111-117.

31. Leslie IJ, Williams S, Isaac G, et al. HIigh cup angle and microseparation increase the wear of hip surface replacements. *Clin Orthop Relat Res* 2009; 467: 2259-2265.

32. Al-Hajjar M. Wear of hard-on-hard hip prostheses: influence of head size, surgical position, material and function. *PhD Thesis* 2012; University of Leeds: Leeds, UK.

33. Al-Hajjar M, Fisher J, Williams S, et al. Effect of femoral head size on the wear of metal on metal bearings in total hip replacements under adverse edge-loading conditions. *J Biomed Mater Res Part B* 2013; 101B: 213-222.

34. McKellop H, Park S-H, Chiesa R, et al. In vivo wear of 3 types of metal on metal hip prostheses during 2 decades of use. *Clin Orthop Relat Res* 1996; 329S: S128-S140.

35. Scott ML, Lemons JE. The wear characteristics of Sivash/SRN Co-Cr-Mo THA articulating surfaces. *Alternative Bearings Surfaces in Total Joint Replacement* 1998; JJ Jacobs, TL Craig; STP 1346: 159-172.

36. Campbell P, McKellop H, Alim R, et al. Metal-on-metal hip replacements: wear performance and cellular response to wear particles. *Cobalt Base alloys for Biomedical Applications* 1999; JA Disegi, RL Kennedy, R Pilliar; STP 1365: 193-209.

37. Rieker CB, Schön R, Köttig P. Development and validation of a second-generation metal-on-metal bearing. Laboratory studies and analysis of retrievals. *J Arthroplasty* 2004; 19(8 Suppl 3): 5-11.

38. Medley JB, McGarry W, Campbell P, et al. Well-positioned large diameter surface replacements can have low wear in vivo. *Ann Meeting of the ORS* 2007; Poster 1696.

39. Medley JB, Chan FW, Krygier JJ, et al. Comparison of alloys and designs in a hip simulator study of metal on metal implants. *Clin Orthop Relat Res* 1996; 329S: S148-S159.

40. Firkins PJ, Tipper JL, Saadatzadeh MR, et al. Quantitative analysis of wear and wear debris from metal-on-metal hip prostheses tested in a physiological hip joint simulator. *Biomed Mat Eng* 2001; 11: 153-157.

41. Firkins PJ, Tipper JL, Ingham E, et al. A novel low wearing differential hardness, ceramic-on-metal hip joint prosthesis. *J Biomechanics* 2001; 34: 1291-1298.

42. Dowson D., Hardaker C, Flett M, et al. A hip joint simulator study of the performance of metal-on-metal joints. Part II: design. *J Arthroplasty* 2004; 19(8, Suppl 3): 124-130.

43. Farrar R, Schmidt MB. The effect of diametral clearance on wear between head and cup for metal on metal articulations. *Ann Meeting of the ORS* 1997; 71-12.

44. Medley JB, Dowling JM, Poggie RA, et al. Simulator wear of some commercially available metal-on-metal hip implants. *Alternative Bearings Surfaces in Total Joint Replacement* 1998; JJ Jacobs, TL Craig; STP 1346: 92-110.

45. Goldsmith AAJ, Dowson D, Isaac GH, et al. A comparative joint simulator study of the wear of metal-on-metal and alternative material combinations in hip replacements. *Proc IMechE Part H: J Engineering in Medicine* 2000; 214: 39-47.

46. Scholes SC, Green SM, Unsworth A. The wear of metal-on-metal total hip prostheses measured in a hip simulator. *Proc IMechE Part H: J Engineering in Medicine* 2001; 215: 523-530.

47. Leslie I, Williams S, Brown C, et al. Effect of bearing size on the long-term wear, wear debris, and ion levels of large diameter metal-on-metal hip replacements – an in vitro study. *J Biomed Mater Res B Appl Biomat* 2008; 87: 163-172.

48. Vassiliou K, Elfick APD, Scholes SC, et al. The effect of 'running-in' on the tribology and surface morphology of metal-on-metal Birmingham hip resurfacing device in simulator studies. *Proc IMechE Part H: J Engineering in Medicine* 2006; 220: 269-277.

49. Campbell P, Park S-H, Lu Z. Post-recall retrieval analysis of metal-on-metal total hips. *Ann Meeting of the ORS* 2015; Poster 1793.

50. Willert HG, Buchhorn GHH, Göbel D, et al. Wear behavior and histopathology of classic cemented metal on meal hip endo prostheses. *Clin Orthop Rel Res* 1996; 329S: S160-S186.

51. Glynn-Jones S, Roques A, Taylor A, et al. The in vivo linear and volumetric wear of hip resurfacing implants revised for pseudotumor. *J Bone Jt Surg* 2011; 93-A: 2180-2188.

52. Takamura KM, Amstutz HC, Lu Z, et al. Wear analysis of 39 conserve plus meal-on-metal hip resurfacing retrievals. *J Arthroplasty* 2013; 29(2): 410-415.

Appendix A

Development of a Formula for Converting Linear to Volumetric Wear

Data values are taken from both early analytical [39, 50, 44] and early experimental [34] studies and fit by a power relationship to produce the following conversion formula (Figure A1).

$$V = 0.0702\, L^{1.334} \quad [R^2 = 0.88] \tag{A1}$$

where L = linear wear which is the sum of maximum head and cup surface deviations from their original spherical shape that are caused by the wear (μm)

V = volumetric wear (mm^3)

R^2 = coefficient of determination

Some later studies [51, 52] produced experimental data that, although scattered, is quite well represented by the conversion formula based on the early data.

Appendix B

Details of the Simulator Studies Used in the Present Study

The simulator studies are listed in chronological order with some details of the test components and conditions (Table B1). The studies are identified by the first author and date of publication here and in the legends of the graphs in the figures of the main body. The reference numbers for these studies are included in Table B1.

Appendix C

Details of the Retrieval Studies Used in the Present Study

 The retrieval studies are listed in chronological order with some details of the components, implantation times and whether linear or volumetric wear was measured (Table C1). The studies are identified by the first author and date of publication here and in the legends of the graphs in the figures of the main body. The reference numbers for these studies are included in Table C1.

Captions

Table B1. Details of the wear simulator studies considered in the present study. See Legend below for definitions of the acronyms.

Table B1. Details of the wear simulator studies considered in the present study (continued). See Legend below for definitions of the acronyms.

Table C1. Details of the retrieval studies considered in the present study. See Legend below for definitions of the acronyms.

Figure 1. Correlation of all the data considered in the present study. The data points that are coloured red are for simulator wear under CR+ conditions (such as stop-start, microseparation, high inclination angle, "jogging" conditions). See Tables B1 and C1 for further details of the studies listed in the graph legends. The reference numbers for the various studies are found in Tables B1 and C1. The * superscript indicates a second data set from the same publication.

Figure 2. A focus in the region of interest ($1/R \leq 0.3$ m^{-1}) for the studies published up to 2004: (a) showing all of the data and (b) showing classified simulator data with the omission of data obtained for CR+ Conditions. Note that the $1/R$ axis limit is reduced to 0.3 m^{-1} and the V' axis limit can then be reduced to 20 mm^3/yr.

Figure 3. Linear fits for data from studies published up to 2004: (a) for ID conditions and (b) for ID Conditions and Retrievals together. Note that the $1/R$ axis limit is 1 m^{-1} (as in Figure 1) which allows the higher $1/R$ values to strengthen the correlations. The V' axis limit can remain at 20 mm^3/yr (as in Figure 2).

Figure 4. The identification of higher wear rates in simulator testing under CR+ conditions for studies published up to 2004. Note that the $1/R$ axis limit returns to 0.3 m^{-1} (as in Figure 2) and the V' axis limit can remain at 20 mm^3/yr (as in Figures 2 and 3).

Figure 5. The simulator wear rates under ID and CR+ conditions and wear rates from one retrieval study, all studies published after 2004. Note that the 1/R axis range remains at 0.3 m^{-1} (as in Figures 2 and 4) but the V' axis limit must increase to 100 mm^3/yr (as in Figure 1). The * superscript indicates a second data set from the same publication.

Figure A1. Formula for converting linear (L) to volumetric (V) wear. In the order listed in the legends, the early data was from [39, 50, 34 and 44] and the recent data was from [51, 52].

Table B1. Details of the wear simulator studies considered in the present study. See Legend below for definitions of the acronyms.

Study	Year	Ref	Sim/ Cycles (Mc)	Alloy ASTM	Dia (mm)	R_C-R_H (μm)	n	F_{max} (kN)	Notes
Medley	1996	[39]	MATCO 1.4-3.1	F1537(LC), F75	28 / 45	18-87 / 45-315	6 / 8	2.1	- custom implants - some high run-in wear
Chan	1996	[24]	MATCO 3.1	F1537(LC), F75	45	89-315	5	2.1	- custom implants - add cycles to some implants from [1]
Farrar	1997	[43]	MATCO 2.0	F1537	28	7-161	10	2.0	- 2 implants with R_C-R_H not included - wide range of R_C-R_H
Medley	1998	[44]	MATCO 1.5	F1537(HC) F1537(LC)	28	44-54	6	2.1	- Sulzer: HC - custom implants: LC
Chan	1999a	[22]	MATCO 3.0	F1537(HC) F1537(LC), F75	28	15-53	22	2.1	- custom implants - smooth, low sphericity
Chan	1999b	[26]	MATCO 0.7 – 1.0	F1537(HC) F1537(LC), F75	28	61	6	3.4	- custom implants - CR+ conditions: stop-start motion
Goldsmith	2000	[45]	ProSim 3.4 – 5.0	F1537(HC)	28 / 36	46-68 / 58-86	4 / 6	2.9	- Sulzer - DePuy: smooth,low sphericity
Firkins	2001a	[40]	Physiol 5.0	F1537(HC) F1537(LC)	28	22-33	3	-	- custom implants - mix HC and LC heads and cups
Firkins	2001b	[41]	Physiol 5.0	F1537(HC) F1537(LC)	28	29-31	3	-	- custom implants - LC heads against HC cups
Scholes	2001	[46]	Custom 5.0	F1537(LC)	28	22-40	4	2.0	- did not find a relationship between wear and clearance
Butterfield	2002	[27]	ProSim 5.0	-	28	30	3	-	- CR+ conditions: microseparation - wear not considered high
Fisher	2002	[28]	Physiol 5.0	-	28	30	-	-	- CR+ conditions: microseparation - caused increased wear
Williams	2004	[29]	ProSim 5.0	F1537, HC head, LC cup	28	30	5	2.0	- CR+ conditions: microseparation - wear "significantly increased "
Dowson	2004	[42]	ProSim 2.0-5.0	F1537 (HC) F75	16-54.5	30.5-143.5	25	2.5	- include 5 ASR's (54.5 mm Dia) - wide range of head sizes
Bowsher	2006	[10]	Custom	F75 as cast F75 HIP+SA	40	104-144	5	4.5	- CR+ testing: 1.75 Hz for 1 Mc - part of a duty cycle
Vassiliou	2006	[48]	Custom 5.0	F75	50	80-105	5	3.0	- low friction and mostly low wear - Stribeck analysis
Leslie	2008	[47]	ProSim 15.0	F75	38.5-54.5	53-63	10	3.0	- all ASR's, averaged values - more run-in wear for the smaller size

Table B1. Details of the wear simulator studies considered in the present study (continued). See Legend below for definitions of the acronyms.

Study	Year	Ref	Sim/ Cycles (Mc)	Alloy ASTM	Dia (mm)	R_C-R_H (μm)	n	F_{max} (kN)	Notes
Williams	2008	[30]	Custom 5.0	F1537(LC) F1537(HC)	28	30	4	-	- CR+ conditions: high angle - LC head against HC cups
Williams*	2008	[30]	Custom 2.0-5.0	as above	28	30	5	-	- CR+ conditions: high angle with microseparation
				F1537	39	63	6		
Leslie	2009	[31]	ProSim 2.0	F75	37.5	63	5	3.0	- CR+ conditions: high angle - ASR's
Leslie*	2009	[31]	Physiol 2.0	F75	37.5	63	5	3.0	- CR+ conditions: high angle with microseparation, ASR's
Al-Hajjar	2013	[32] [33]	Physiol 3.0	F75	28	40-43	3	3.0	- CR+ conditions: microseparation - Corin implants
					36	40-57	3		
Al-Hajjar*	2013	[32] [33]	Physiol 3.0	F75	28	40	3	3.0	- CR+ conditions: high angle with microseparation, Corin implants
					36	40-51	3		

Legend

Ref = reference number   Sim = simulator   Mc = millions of cycles

Dia = nominal head diameter R_H = radius of head surface R_C = radius of cup surface

n = number of implants tested F_{max} = peak load

F75, F1537 = ASTM alloy designations

HC = high carbon (> 0.20%)   LC = low carbon (<0.07%)

* indicates a second data set from the same publication (involving different CR+ conditions)

Table C1. Details of the retrieval studies considered in the present study. See Legend below for definitions of the acronyms.

Study	Year	Ref	Implant Name	n	Dia (mm)	R_C-R_H (μm)	Time in situ (yr)	Notes
McKellop	1996	[34]	McKee-Farrar Ring (1 only)	9	35-41	64-193	1-25	- convert linear to volumetric wear - measured head volumetric wear
Scott	1998	[35]	Sivash	13	28	60-115	1.8-22	- head held within the cup
Campbell	1999	[36]	Metasul	3	28	48-60	2.2-4	- convert linear to volumetric wear
Rieker	2004	[37]	Metasul	110	28	35-78	0.3-8.5	- convert linear to volumetric wear
Medley	2007	[38]	McMinn, BHR Wagner, Cormet	19	39.8-55.8	92-275	0.2-10	- convert linear to volumetric wear - surface replacement but not ASR's
Medley*	2007	[38]	Biomet, Accis	2	27.8-42.0	45-136	4.2-4.6	- convert linear to volumetric wear - stem-type implants
Lord	2011	[18]	ASR	22	40.5-50.5	50	0.7-4.8	- measured volumetric wear - surface replacement

Legend

Ref = reference number   n = number of implants   Dia = implant diameter

R_H = radius of head surface R_C = radius of cup surface

* indicates a second set of retrieval data from the same publication (for a different type of implant)
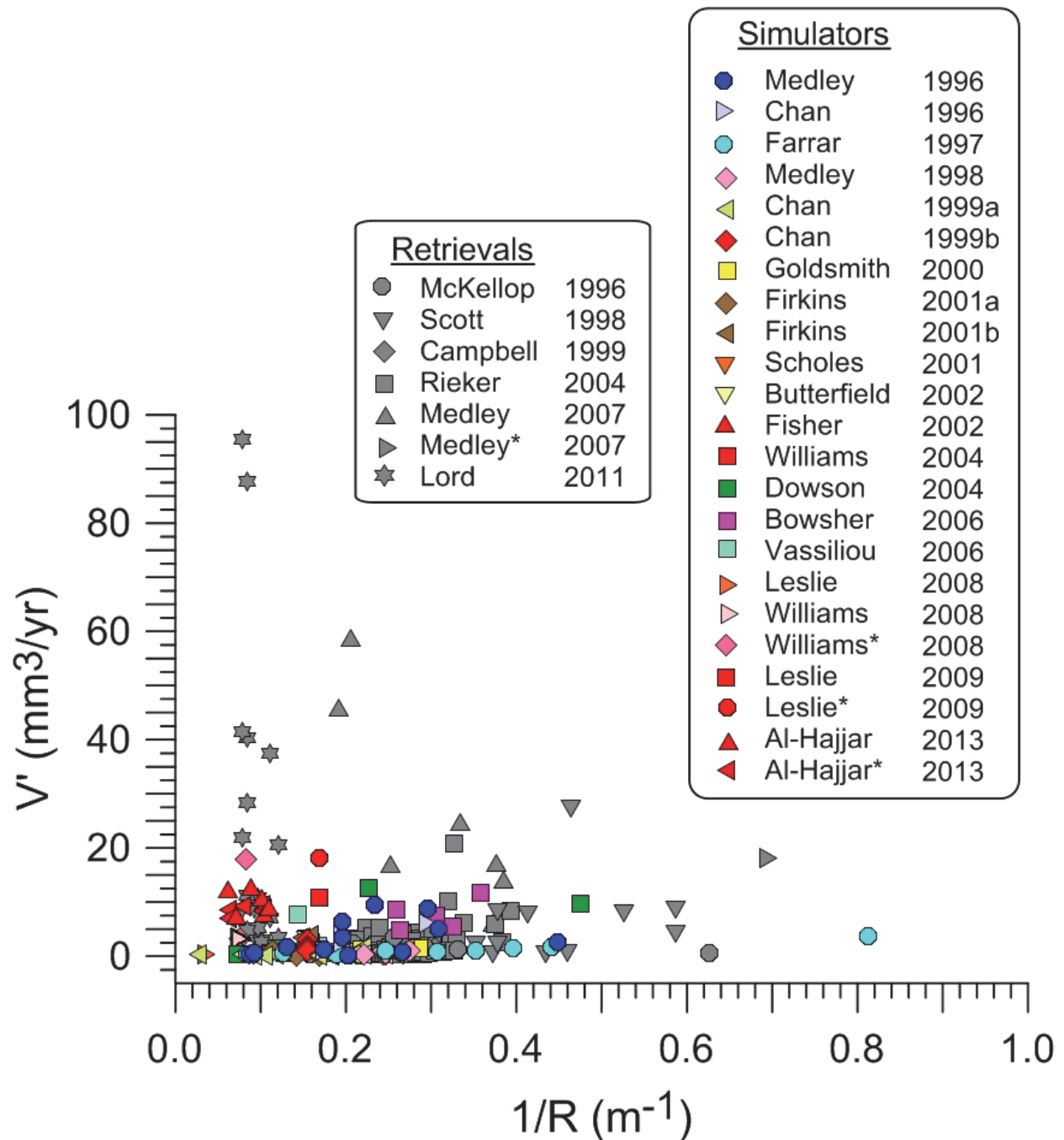
Figure 1. Correlation of all the data considered in the present study. The data points that are coloured red are for simulator wear under CR+ conditions (such as stop-start, microseparation, high inclination angle, "jogging" conditions). See Tables B1 and C1 for further details of the studies listed in the graph legends. The reference numbers for the various studies are found in Tables B1 and C1. The * superscript indicates a second data set from the same publication.
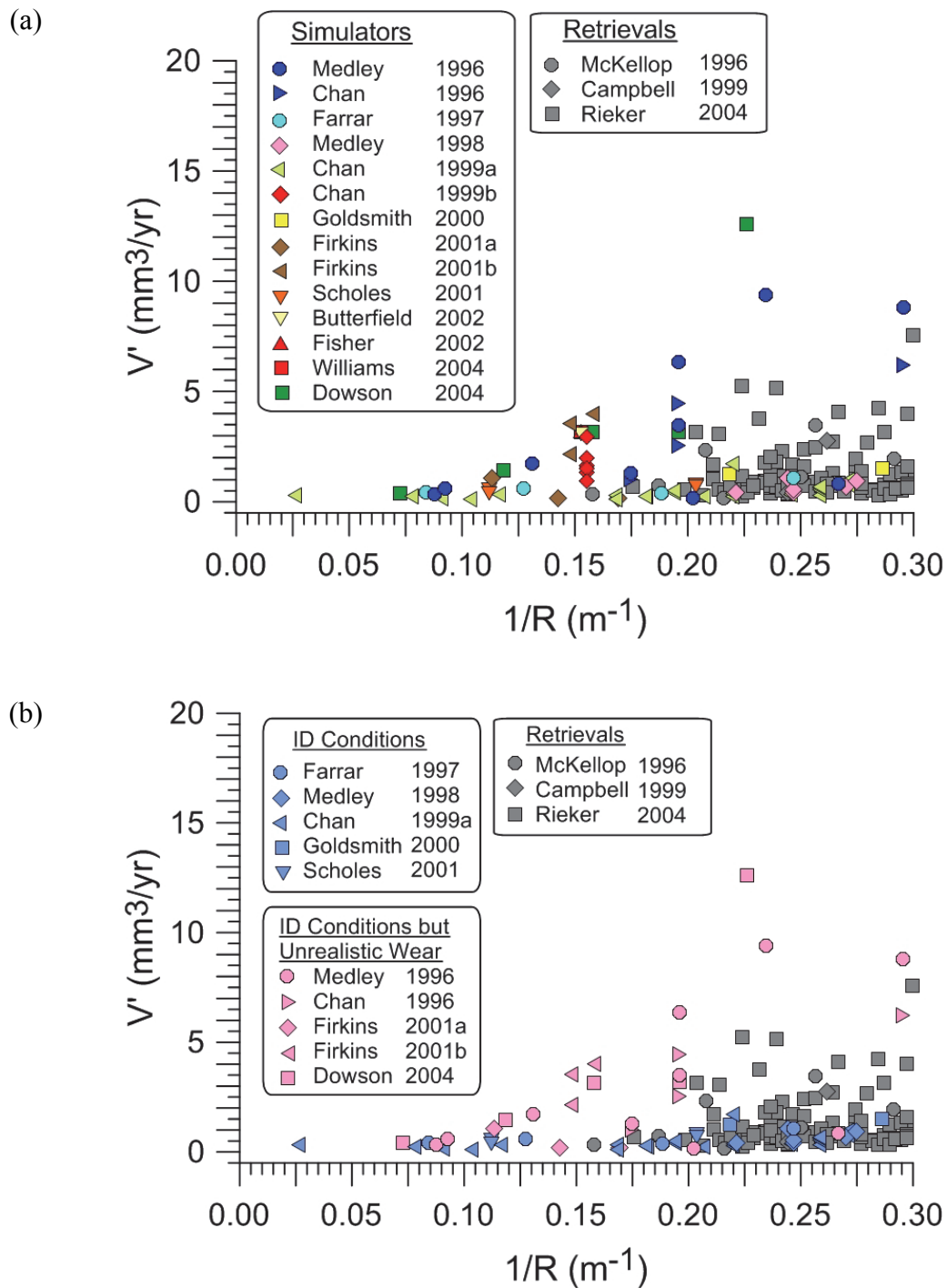
(a)

(b)

Figure 2. A focus in the region of interest ($1/R \leq 0.3$ m^{-1}) for the studies published up to 2004: (a) showing all of the data and (b) showing classified simulator data with the omission of data obtained for CR+ conditions. Note that the $1/R$ axis limit is reduced to 0.3 m^{-1} and the V' axis limit can then be reduced to 20 mm^3/yr.
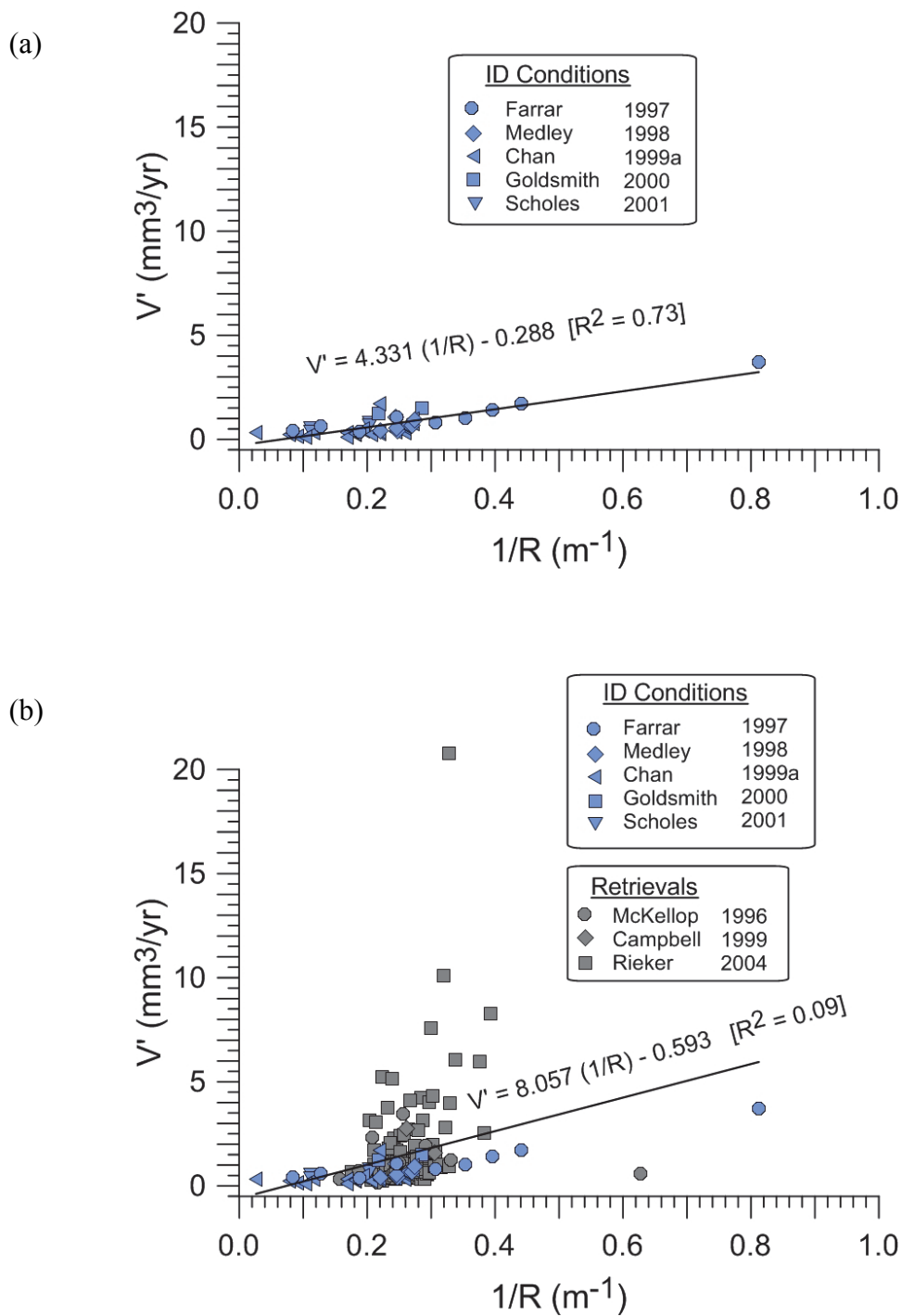
(a)

(b)

Figure 3. Linear fits for data from studies published up to 2004: (a) for ID conditions and (b) for ID Conditions and Retrievals together. Note that the 1/R axis limit is 1 m[-1] (as in Figure 1) which allows the higher 1/R values to strengthen the correlations. The V' axis limit can remain at 20 mm^3/yr (as in Figure 2).
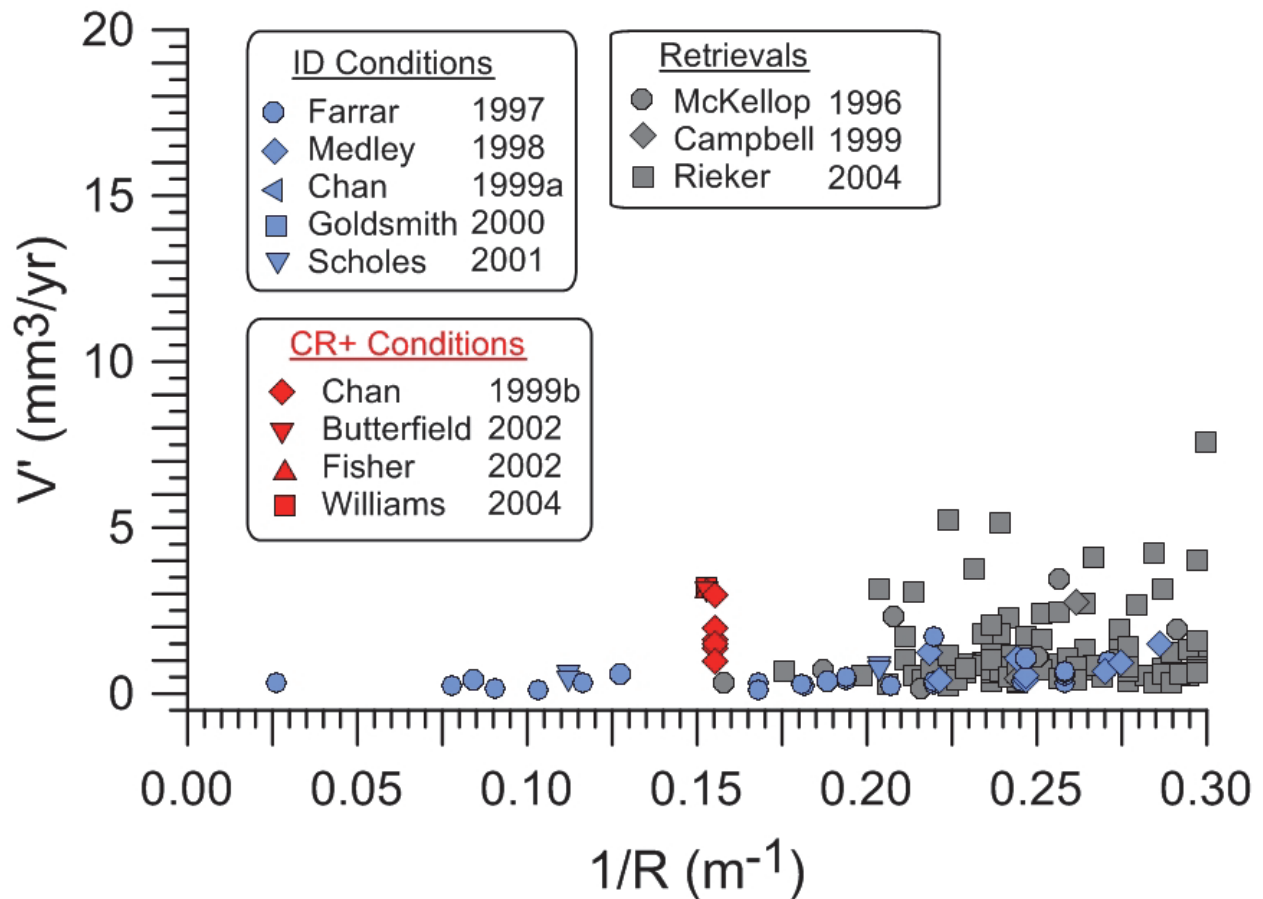
Figure 4. The identification of higher wear rates in simulator testing under CR+ conditions for studies published up to 2004. Note that the 1/R axis limit returns to 0.3 m^{-1} (as in Figure 2) and the V' axis limit can remain at 20 mm^3/yr (as in Figures 2 and 3).
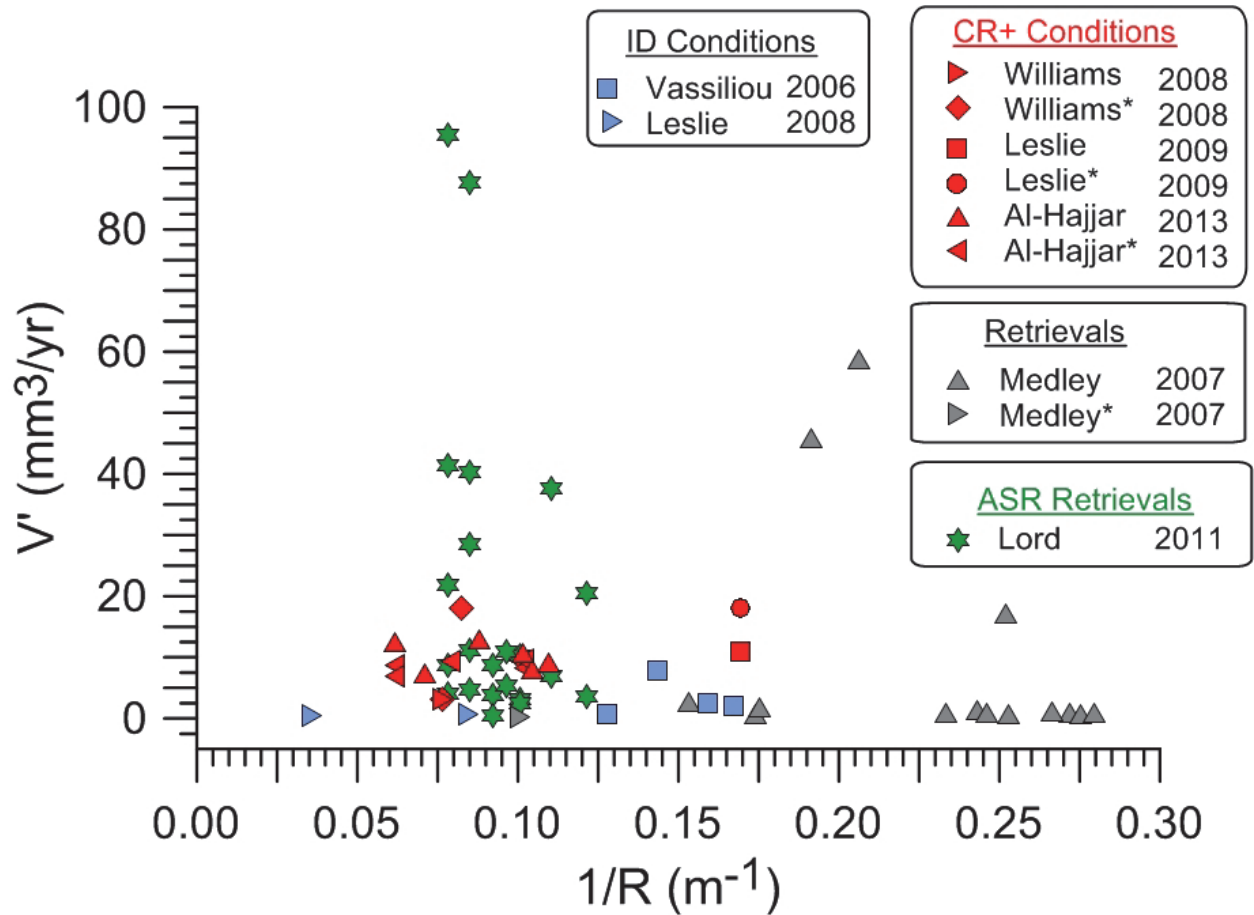
Figure 5. The simulator wear rates under ID and CR+ conditions and wear rates from two retrieval studies, all studies published after 2004. Note that the 1/R axis range remains at 0.3 m^{-1} (as in Figures 2 and 4) but the V' axis limit must increase to 100 mm^3/yr (as in Figure 1). The * superscript indicates a second data set from the same publication.
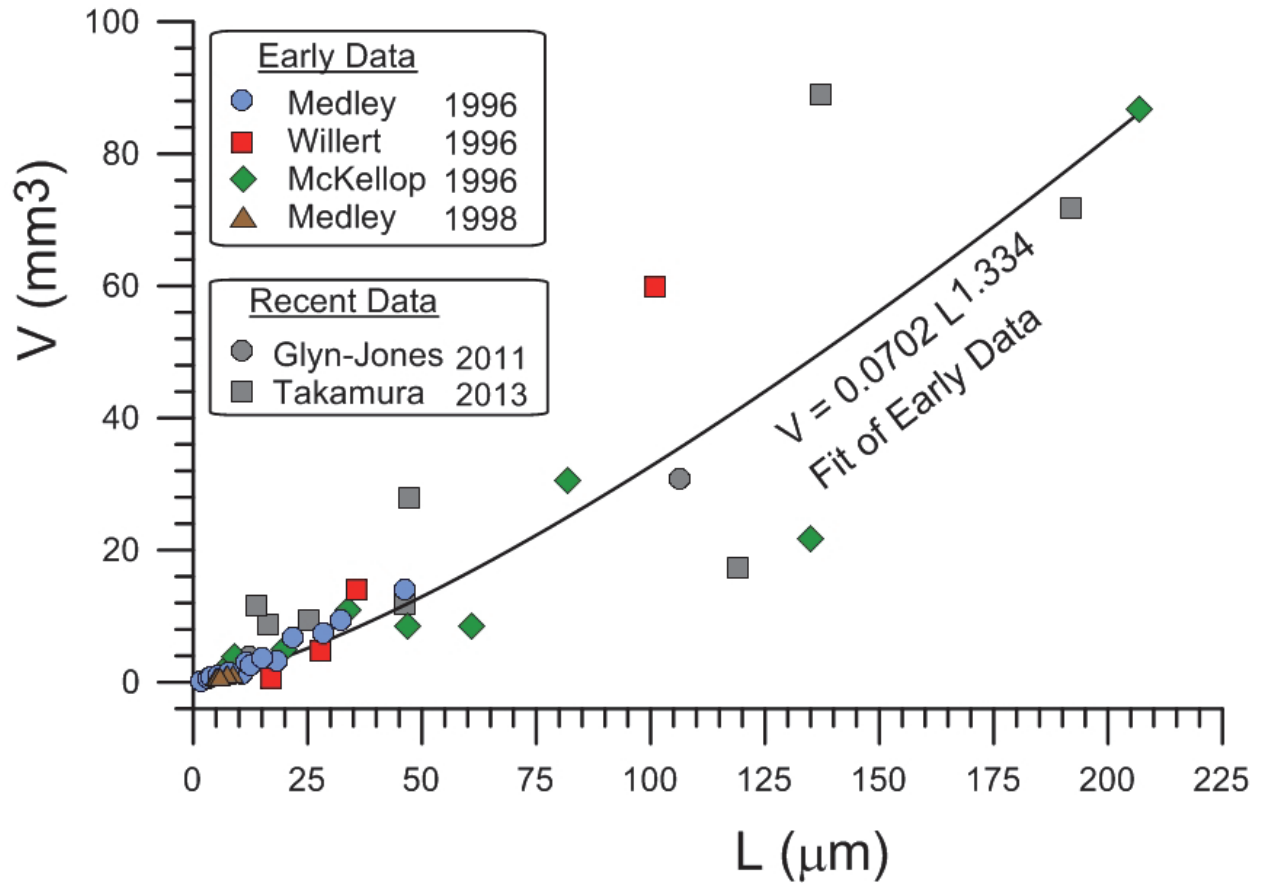
Figure A1. Formula for converting linear (L) to volumetric (V) wear. In the order listed in the legends, the early data was from [39, 50, 34 and 44] and the recent data was from [51, 52].