Metal-on-metal hip arthroplasty : local tissue reactions and clinical outcome

Weegen, W. van der

Chapter 12

Summary

Long term durability of hip replacement implants is mainly limited by wear of the bearing surfaces between the femoral and acetabular components. Different bearing materials have been used with the aim to reduce wear and prolong implant survival. Polyethylene (PE), commonly used as a bearing surface on the acetabular side, releases wear particles which induce osteolysis with subsequent component loosening and ultimately implant failure. In the constant strive to improve on implant design and materials, a second generation of Metal-on-Metal (MoM)surface bearings was introduced in the 1990s with the promise of reduced wear, thereby supposedly improving long term implant survival. The main support for this claim was achieved by in vitro testing, using wear simulators which were run under ideal conditions.

This re-introduction, after the use of the first generation MoM was discontinued due to unacceptable high failure rates, took place in the context of limited requirements on supportive clinical data to release new designs into the orthopedic market. After approximately a million MoM hip implants were inserted into patients (both total hip arthroplasty (THA) and resurfacing procedures), it became clear that unexpected complications occurred in soft tissue surrounding this MoM implant. These were due to metal debris released from the bearing surfaces. This eventually led in 2010 to a worldwide recall of one of the MoM hip implant designs, followed by a ban on the use of MoM large diameter hip implants in several countries. Since these reactions were unforeseen, no evidence based guidelines on how to diagnose and to treat these complications were available. Since there is no final data or consensus on the risk-benefit ratio for the use of MoM implants, the use of large diameter MoM hip implants is currently banned in some countries and still in use in many other countries.

Chapter 1

The first chapter introduces current issues raised with implant survival and bearing surfaces in THA and in Hip Resurfacing Arthroplasty (HRA). The four main aims of this thesis are presented, being (1) To report the implant survival of the current THA bearing options seen as gold standard for the young and more active patients. (2) To review all available literature on the different resurfacing systems (3) To investigate complications after MoM due to soft tissue reactions to metal

wear debris, and (4) To study the most used diagnostic tool, MRI, and the classification systems used to find and rate these complications.

Chapter 2

In chapter two we reviewed the long history of MoM bearing surfaces in hip arthroplasty. Since younger patients tend to be physically more active than elderly patients, their implants have to withstand higher biomechanical stress and these stresses also need to be endured for a more prolonged period of time, leading to accelerated wear of the bearing surfaces. To reduce bearing surface wear, surgeons, engineers and scientists have developed different bearing surfaces. For this purpose, Metal-on-Metal (MoM) surface bearings have a long tradition in THA. The re-introduction of the second generation MoM in the 1990s took place after the first generation of MoM was abandonded due to unacceptable high failure rates and as an answer to “polyethylene disease”, occuring with standard Ultra High Molecular Weight Polyethylene (UHMWPE) bearings. Wear simulation tests of second generation MoM bearings showed that wear rates were 20 to 100 times lower compared to metal-on-conventional polyethylene, and MoM bearing couples started to experience widespread clinical use in both hip resurfacing and total hip arthroplasty. The material properties allowed the use of large heads in thin acetabular shells, promising a reduced incidence of hip dislocation in younger and more active patients.

Despite the biomechanical advantages of MoM bearings, metal ion release over time and the potential detrimental effects of accumulated metal ions in the body remained a concern, and research started to identify implant failure modes in reaction to metal wear particles. The terms ALVAL [2005], pseudotumor [2008] and metallosis were used, together with a new umbrella term for these modes of failure: Adverse Reaction to Metal Debris (“ARMD”, 2010). These unforeseen complications revealed serious shortcomings in how orthopaedic innovations are introduced into clinical practice. The conflicting interests of making promising new hip implant materials and designs available so patients can benefit as soon as possible, and the fact that these same joint replacement devices have to perform well for over more than 10 years and preferably more than 20 years after implantation in the patient, make it difficult to design a model for market introduction that effectively and safely guards all these requirements. In comparison to pharmaceuticals which require multiple controlled clinical trials

prior to approval, which take a mean of nine years and cost an average of 800 million U.S. dollars, medical devices such as a new hip implant design can be released onto the market after in vitro testing and very limited clinical trials. As witnessed with the re-introduction of MoM bearings in THA, serious complications which were unforeseen at the time of introduction became only became known after a large number of patients (worldwide an estimated one million patients) had become at risk.

Chapter 3

Before MoM hip arthroplasty became available for clinical use in the Netherlands, uncemented THA with standard UHMWPE was the gold standard for younger patients. With this prosthesis design, PE wear remained an important clinical observation and to evaluate implant performance, we retrospectively measured radiographic wear and implant survival of the first 200 consecutative uncemented hip arthroplasties with standard UHMWPE used in our clinic. In this series we found a high proportion (53.4%) of implants with a wear rate of >0.2 mm per year, which is considered a threshold for accelerated wear. This was after a mean follow-up of 8.3 years. Somewhat in contrast, implant survival at a maximum of 12 years was acceptable (Kaplan-Meier survival probability 90.1%), and compliant to international guidelines such as the NICE criteria.

Chapter 4

With the re-introduction of MoM arthroplasty, all major orthopedic device manufacturers designed and introduced, sometimes slightly, different hip resurfacing implants. Individual studies using different resurfacing designs reported marked differences in short term implant performance, so we decided to sytematically review the peer-reviewed literature on implant survival of all contemporary MoM hybrid hip resurfacing designs. A total of 29 studies, compromising 10,621 patients, were included. All but one of the implants studied had insufficient follow up to be compliant with the NICE benchmark, of a revision rate of less than 10% at ten years, for choosing a prosthesis for primary THR. The study reporting a follow-up of longer than ten years had a revision rate of 16%, mainly for aseptic loosening of the implant. This high failure rate was attributed to the double-heat-treatment manufacturing process which is no longer in use. The prosthesis was superseded by the Cormet 2000 implant in 1996. Compared with

the three-year NICE entry-benchmark of implant survival ≥97%, 13 studies (44.8%) showed satisfactory survival. Eight used the BHR implant, two the Conserve plus, one the Durom, one the Cormet 2000 and one both the McMinn and BHR implants. Based on the results of this review we concluded that aseptic loosening was the most frequent failure mode and that none of the contemporaty hip resurfacing designs met the full 10 year NICE benchmark for survival. In this systematic review we were unable to include studies on the resurfacing implant used in our clinic.

Chapter 5

In this chapter we present the results of data prospectively collected in a series of 280 consecutative hip resurfacing procedures (ReCap, Biomet, Warsaw, USA) performed in our clinic. Mean follow up was 3.3 years (range: 1.0 to 6.3) and four patients were lost to follow-up. All patients were diagnosed with end-stage hip osteoarthritis, their mean age was 54 years and 76.4% of all patients were male. All were evaluated with standard radiographic imaging and clinical outcome scores before surgery and yearly after the index surgery. There were 16 revisions and four patients reported a Harris Hip Score <70 points at their latest follow up. Kaplan-Meier implant survival probability, with revision for any reason as endpoint, was 93.5% at six years follow-up (95%-CI: 88.8-95.3). There were no revisions for Adverse Reactions to Metal Debris (ARMD) and no indications of ARMD in symptomatic non-revised patients, although diagnostics were limited to ultrasound scans. We concluded that hip resurfacing is a demanding procedure, and that implant survival of the ReCap hip resurfacing system is on a critical level in our series. However, in non-revised patients, reported outcomes are generally excellent.

Chapter 6

In chapter six we presented the results of a pilot study in which we used an intensified screening protocol to detect pseudotumor formation after MoM hip resurfacing in three selected groups of patients: a group with a theoretically high risk for pseudotumor formation, a group with a very low risk for pseudotumor formation and a group scheduled for routine follow up with a mix of risk factors present. Risk factors were based on component size and orientation, gender, bilateral or unilateral MoM surgery and clinical symptoms. All selected patients

underwent blood metal ion level analysis and cross-sectional imaging using MARS-MRI. In this study we used a pseudotumor classification system devised by Anderson et al to grade pseudotumor severity. Pseudotumor formation was observed in all three groups, even in asymptomatic patients with normal blood metal ion levels. In 15 out of 44 MRI scans pseudotumors were observed (34.1%), of which six were graded with mild (13.6%), eight with moderate (18.2%) and one with severe MoM disease (2.3%). Twelve pseudotumors were present in asymptomatic patients (27.3%). Metal ion levels were normal in 80% of the MARS-MRI screened patients. As a consequence to our intensified screening protocol, one patient was revised for pseudotumor occurence and another patient scheduled for revision. Asymptomatic pseudotumors were observed in all three groups. We concluded that clinical outcomes and plain radiographs for screening MoM patients severely underestimated the presence of pseudotumors in MoM patients.

Chapter 7

Different pseudotumor grading systems had been described in the scientific literature, but no studies had compared these different systems for use in clinical practice and only limited data on the reliability of these grading systems was available. In chapter 7 we investigated the influence of using these different pseudotumor grading systems on how severe pseudotumors were classified. For this study we evaluated a cohort of 42 THA patients (49 MoM hips) using three different pseudotumor grading systems designed respectively by Anderson et al, by Matthies et al and by Hauptfleisch et al. Two experienced musculoskeletal radiologists evaluated all MARS-MRI scans with these systems, allowing us to calculate the interobserver reliability for each system. Our results showed that, regardless of the classification system used, grading pseudotumor severity on MARS-MRI had only a moderate interobserver reliability (ICC 0.65 to 0.68). The reliability of pseudotumor severity grading was high between the Matthies an Hauptfleisch system but low between the Anderson and the other two systems. We concluded that a more succinct pseudotumor severity grading system is needed for clinical use.

Chapter 8

Since we demonstrated in chapter 6 that standard radiographic follow up combined with clinical outcomes was not sensitive enough to detect pseudotumor formation after MoM hip arthroplasty, we extended our intensitied screening protocol to our complete cohort of MoM hip resurfacing patients. This study is presented in chapter 8. At the time this study was started, 248 MoM hip resurfacing procedures (214 patients, mean follow-up 4.6 years, range: 1 to 8.2) were available for follow up. Again the Anderson classification for pseudotumors was used. We found a pseudotumor prevalence of 36.3%: 61 pseudotumors were graded mild, 25 moderate and four were graded severe. Five revisions followed, all in symptomatic patients with elevated metal ion levels. Since the natural course of pseudotumors is largely unknown, and no validated treatment regime for pseudotumors after MoM hip arthroplasty exists, we suggested to repeat MARS-MRI in asymptomatic patients with mild to moderately severe pseudotumors combined with normal metal ion levels, rather than to immediately revise these cases. The use of this screening protocol and this pseudotumor grading system allowed us to be conservative with revision surgery for mild and moderate MoM disease. Of course patients with non-revised pseudotumors were kept under increased surveillance. These results could be used as a clinical guideline for management of observed pseudotumor after MoM hip resurfacing.

Chapter 9

As stated in the previous chapter, intensified follow up of cases with non-revised pseudotumors was needed to validate a more conservative approach in the management of observed pseudotumors. In chapter 9 we present the results of repeated MARS-MRI's which were used to follow up on identified pseudotumors from our previous studies. To monitor how pseudotumors developed in time, we repeated cross-sectional imaging 6 to 12 months after the initial MARS-MRI, together with repeated metal ion analysis and clinical examination. In this study, 14 unrevised cases with pseudotumour and a control group of 23 cases without pseudotumour on the first MARS-MRI were evaluated. The mean postoperative time to the first MARS-MRI was 4.3 years (range: 2.2 to 8.3) and mean time between first and second MARS-MRI was 8 months (range: 6 to 12). The majority of patients (35/37) showed no change in pseudotumor severity with the second MRI, one new pseudotumour was observed (Anderson C2 score, moderate) and

one pseudotumour was downgraded from C2 (moderate) to C1 (mild). We concluded that repeating of MARS-MRI within one year, in unrevised patients with asymptomatic pseudotumours after MoM hip resurfacing, was of limited use. But, since this was the first longitudinal study on pseudotumours using MARS-MRI, our findings need to be interpreted with caution.

Chapter 10

Since management of non-revised pseudotumors depends on both severity (based on location, type of content, growth rate) and on pseudotumor dimensions, it is relevant to have an accurate clinical measurement method of pseudotumor dimensions. In this chapter our objective was to validate clinical measurements of (change in) pseudotumor dimensions (maximum diameter and estimated volume) against three-dimensional region-of-interest (3-D ROI) volume measurements. Therefore, we had MARS-MRI scans available for 13 cases of non-revised pseudotumors after Metal-on-Metal hip resurfacing. Mean follow-up at the first MARS-MRI was 5.3 years (range: 2.4 to 7.5), a second MARS-MRI was acquired after a mean of 7.5 months (range: 6 to 12). On all scans pseudotumor dimensions were measured by (1) maximum diameter in one plane (MD) and (2) by estimating pseudotumor volume based on maximum diameter in three different planes (EV). (3) For validation, a 3-D ROI based volume (V) was calculated by the summation of all pseudotumor areas in each slice and multiplication by the slice profile. Correlations between MD, EV and V were calculated. Correlation was high between all three measurement methods, but the correlation was strongest between EV and V. EV overestimated V with a mean of 72.6%, and more so in non-ellipsoid pseudotumors than in ellipsoid pseudotumors. Median values for MD, EV or V were not significantly different between first and second MARS-MRI. Median change for MD was 0.0cm (range: -1.5 to 3.4), -0.5ml for EV (range: -16.4 to 45.5) and 0.5ml for V (range: -7.7 to 5.2). Percent change in pseudotumor dimensions was not significantly different between MD, EV and V.

We concluded that estimating pseudotumor volume in clinical practice using maximum pseudotumor diameter in three different planes has a strong correlation with a more elaborate 3D-ROI method. This method of estimating volume is easily attainable in clinical practice and can be used for monitoring change in pseudotumor volume over time.

Chapter 11

In this chapter, the findings of all studies conducted for this thesis are synthesized and discussed in their context, resulting in answers to the main study aims and propositions for future research. The first aim of this thesis was to report the implant survival of the current THA bearing options seen as gold standard for the young and more active patients. We concluded that for these patients, hip resurfacing with MoM bearing surfaces was not compliant with the international benchmarks for 10 year implant survival and uncemented, standard UHMWPE hip prostheses just barely reached this benchmark. For the UHMWPE prostheses, the high amount of wear was noticed as the biggest downside with a potential accelerated wear in the second decade after implantation.

The second aim of this thesis was to review all the different resurfacing systems on the market for implant survival results. After systematically reviewing the literature, we concluded that all reviewed hip resurfacing systems did not meet the international benchmark, and that there were hip resurfacing systems on the market of which no clinical studies were available for our review. It is noteworthy that the data presented in the studies we reviewed were collected before the unexpected adverse reactions to metal debris released by the MoM bearing surfaces were investigated. Therefore our conclusion that aseptic loosening was the main failure mode of MoM hip resurfacings needs to be seen in this context, and a future update of this review based on current knowledge might change the view on failure of reasons for current MoM systems.

The third aim of this thesis was to investigate complications after MoM due to soft tissue reactions to metal wear debris. We concluded that the incidence of these complications, diagnosed as pseudotumors, was higher than expected, that risk factors were difficult to interpret, and that cross-sectional imaging is necessary to find the true incidence, since many patients have these reactions without being symptomatic. We also found that the use of a pseudotumor classification system was helpful in managing treatment of these complications.

The fourth aim of this thesis was to study the most used diagnostic tool, MRI, and the classification systems used to find and rate these complications. Based on our validation studies we concluded that using these systems observers were able to idenfity pseudotumors but that the classification of pseudotumors severity needed more refinement. Therefore future studies need to validate the treatment which was chosen upon the pseudotumor severity grade that was seen with

MARS-MRI. For clinical practice, we found that a simple box model for estimating pseudotumor volume correlated well with a more elaborate three-dimensional region-of-interest system and was easily used in clinical practice, althoug clinicians using this method have to take some overestimation of pseudotumor volume into account.

CERAMIC-ON-POLYETHELENE DUAL MOBILITY HIP PROSTHESIS