BMJ 2014;349:g5133 doi: 10.1136/bmj.g5133 (Published 9 September 2014)

RESEARCH

Appraisal of evidence base for introduction of new implants in hip and knee replacement: a systematic review of five widely used device technologies

Marc J Nieuwenhuijse *research fellow ICOR and FDA*[1 2 3], R G H H Nelissen *professor*[2], J W Schoones *information specialist*[4], A Sedrakyan *associate professor*[1 3]

[1]Patient Centered Comparative Effectiveness Program and US Food and Drug Administration Medical Device Epidemiology Science and Infrastructure Center, Department of Public Health, Weill Cornell Medical College, New York, NY 10065, USA; [2]Department of Orthopaedics, Leiden University Medical Center, 2300 RC Leiden, Netherlands; [3]Office of Surveillance and Biometrics, Center for Devices and Radiological Health, FDA, Silver Spring, MD 20993, USA; [4]Walaeus Library, Leiden University Medical Center, 2300 RC Leiden, Netherlands

Abstract

Objective To determine the evidence of effectiveness and safety for introduction of five recent and ostensibly high value implantable devices in major joint replacement to illustrate the need for change and inform guidance on evidence based introduction of new implants into healthcare.

Design Systematic review of clinical trials, comparative observational studies, and registries for comparative effectiveness and safety of five implantable device innovations.

Data sources PubMed (Medline), Embase, Web of Science, Cochrane, CINAHL, reference lists of articles, annual reports of major registries, summaries of safety and effectiveness for pre-market application and mandated post-market studies at the US Food and Drug Administration.

Study selection The five selected innovations comprised three in total hip replacement (ceramic-on-ceramic bearings, modular femoral necks, and uncemented monoblock cups) and two in total knee replacement (high flexion knee replacement and gender specific knee replacement). All clinical studies of primary total hip or knee replacement for symptomatic osteoarthritis in adults that compared at least one of the clinical outcomes of interest (patient centred outcomes or complications, or both) in the new implant group and control implant group were considered. Data searching, abstraction, and analysis were independently performed and confirmed by at least two authors. Quantitative data syntheses were performed when feasible.

Results After assessment of 10 557 search hits, 118 studies (94 unique study cohorts) met the inclusion criteria and reported data related to 15 384 implants in 13 164 patients. Comparative evidence per device innovation varied from four low to moderate quality retrospective studies (modular femoral necks) to 56 studies of varying quality including seven high quality (randomised) studies (high flexion knee replacement). None of the five device innovations was found to improve functional or patient reported outcomes. National registries reported two to 12 year follow-up for revision occurrence related to more than 200 000 of these implants. Reported comparative data with well established alternative devices (over 1 200 000 implants) did not show improved device survival. Moreover, we found higher revision occurrence associated with modular femoral necks (hazard ratio 1.9) and ceramic-on-ceramic bearings (hazard ratio 1.0-1.6) in hip replacement and with high flexion knee implants (hazard ratio 1.0-1.8).

Conclusion We did not find convincing high quality evidence supporting the use of five substantial, well known, and already implemented device innovations in orthopaedics. Moreover, existing devices may be safer to use in total hip or knee replacement. Improved regulation and professional society oversight are necessary to prevent patients from being further exposed to these and future innovations introduced without proper evidence of improved clinical efficacy and safety.

Introduction

The introduction of new orthopaedic implants and related technologies has been the focus of major scientific and policy discussions since the failures of articular surface replacement and large head size metal-on-metal articulations in total hip

Correspondence to: A Sedrakyan ars2013@med.cornell.edu

Extra material supplied by the author (see http://www.bmj.com/content/349/bmj.g5133?tab=related#datasupp)

Appendix 1: Search terms used in study
Appendix 2: Ceramic-on-ceramic bearings in total hip replacement
Appendix 3: Total hip replacements with a proximal modular femoral neck
Appendix 4: Total hip replacements with an uncemented monoblock acetabular cup
Appendix 5: High flexion total knee replacements
Appendix 6: Gender specific total knee replacements

replacement were brought to light.[1-7] Scientists, clinicians, journalists, and policy professionals highlighted the need for a more thorough and evidence based introduction of devices and for development of an infrastructure for timely evaluation of these devices.[3-9] The scientific community recognised the differences between medical devices (orthopaedic implants) and pharmaceutical products,[5 9 10] and important guidance was issued by the US regulator.[11]

However, the consequences of uncontrolled device introduction worldwide may not be fully recognised and there is a high likelihood that current practice regarding evaluation of device innovations will not change. As such, there is a need to investigate whether the problems associated with the articular surface replacement and large head size metal-on-metal articulation are isolated events..

In this study, we systematically evaluate the evidence concerning the introduction of five substantial, innovative, relatively recent, and already widely implemented device technologies used in major total joint replacement. We evaluate comparative data from clinical trials, observational studies, and large national arthroplasty registries to study effectiveness and safety over existing, well proven, and comparable device solutions for the same condition. The five technologies are ceramic-on-ceramic bearings, modular femoral necks, and uncemented monoblock (not metal-on-metal) acetabular cups in total hip replacement, and high flexion implants and gender specific implants in total knee replacement.

Methods

Investigated technologies

The selected device innovations met the following criteria:

1. All technologies were innovative in nature and had an a priori rationale for expectation of superior clinical benefit over existing solutions (table 1⇓)
2. Survival outcome was reported by at least one of the national orthopaedic registries with ≥90% national completeness.[2 12] Reporting of outcome by a registry implicates that widespread commercial introduction has taken place and that the reported outcome is representative of actual regular clinical practice (that is, not only by skilled enthusiasts in selected settings)
3. All technologies had a representative safety benchmark or reference for comparison; that is, registry based survival could be compared with existing and representative ("traditional") alternatives for the same indication in comparable patients (total hip or knee replacement).

Identification of studies

In collaboration with an experienced information specialist (JWS), a thorough search strategy was constructed (see appendix 1) and the following bibliographies were searched up to April 2014: PubMed (Medline), Embase, Web of Science, Cochrane Library, CINAHL and Academic Search Premier. Reference lists of trials and reviews were assessed for additional studies. After study inclusion, a citation tracking search was performed and specific searches with the identified corresponding prosthesis brands were performed to identify potentially missed studies. References and references of relevant reviews were hand searched. We limited our selection to studies written in English (UK or US) and all European languages.

In addition, we worked with the Food and Drug Administration (FDA) to identify the summaries of safety and effectiveness for all pre-market application trials and relevant FDA mandated post-market studies reporting comparative information on the investigated technologies. Relevant publications related to pre-market application trials were identified to learn about follow-up results.

Annual reports and spin-off publications from validated registries with ≥90% national completeness were investigated for device survival data.[2 12] Registries report information on implant survival and do not report functional or patient reported outcomes per device. National registries are not intended to estimate device benefits but function as a post-market safety surveillance mechanism. Hence, they are considered only in the systematic appraisal of harms.

Inclusion criteria

Comparative clinical studies on primary total hip or knee replacement for symptomatic primary or secondary osteoarthritis in adults (≥21 years old) which reported any one of the clinical outcomes of interest (functional or patient reported outcomes, or complications, or both) in at least one patient group with the new device and one patient group with the reference device. Laboratory, biomechanical, or radiographic outcomes are not patient centred and therefore do not constitute an inclusion criterion.

Data abstraction

Search results were independently evaluated by at least two experienced abstractors (MJN, AS, AG, SJ, SLP, LR, see acknowledgements). In case of disagreement, consensus was reached with a referee (AS or MJN). After study inclusion, information was independently extracted by two abstractors and assessed for agreement. Information was extracted on study design, study quality aspects, setting, time period, number of implants and patients evaluated, sex, diagnosis (osteoarthritis or other), follow-up length and completion, and brands and manufacturers. The main outcome measures included functional and patient reported outcome measures as well as occurrence of dislocations and revisions.

Two experienced abstractors (MJN, AS) independently assessed data of the included national arthroplasty registries for survival outcomes on the selected technologies. Disagreements were discussed for consensus. The main outcome measure was implant survival.

Data analysis

Methodological quality

Methodological quality was evaluated based on adherence to recommendations by the CONSORT, STROBE, and Cochrane criteria.[13-15] Based on these aspects, overall quality was independently classified by two assessors (MJN, AS) as low (high risk of bias), low to moderate, moderate, moderate to high, and high (low risk of bias), and discussed for consensus.

Publication bias

This investigation covers new technologies and concerns comparative studies only. These studies are relatively uncommon in orthopaedic literature, and we believe that comparative studies on these topics are published irrespective of the findings. Hence we do not expect that publication bias is substantial. We evaluated funnel plots whenever quantitative data synthesis included more than five studies of similar design at one follow-up time.[16]

Qualitative and quantitative aspects

Study heterogeneity (clinical and methodological diversity) among the trials was evaluated by abstracting data on study design, included populations, and definitions of outcomes. Functional outcomes included range of motion, disease specific questionnaires, and patient reported outcome measures. Data on scores (continuous data), number of events (categorical data), and their 95% confidence intervals and P values were recorded. We report relevant outcome measures for the new implants compared with the closest, well established (traditional) implants.

Formal meta-analysis was performed in case of relatively complete reporting on functional outcome measures and in the absence of substantial study heterogeneity (as addressed by data abstraction and assessment of study quality). Meta-analysis was performed using a random effects model, and estimates are presented as means with 95% confidence intervals.[17 18] Statistical heterogeneity was estimated using the I^2 statistic; a value of >50% indicates substantial statistical heterogeneity, and additional estimates from a fixed effects model and subgroups of higher study quality will be reported if the magnitude of the overall estimate is influenced significantly.[19 20] STATA 12.0 (StataCorp) was used for statistical analyses.

Results

Table 2⇓ provides an overview of the results from our systematic review.

Ceramic-on-ceramic articulations in total hip replacement

Comparative effectiveness

We found and assessed 2911 abstracts, of which 42 studies based on 23 unique study cohorts could be included (appendix 2). In 10 cohorts a ceramic-on-ceramic (CoC) articulation was compared with a metal-on-polyethylene (MoP) articulation and in 13 cohorts CoC articulation was compared with a ceramic-on-polyethylene (CoP) articulation. These studies combined included 5442 total hip replacements in 4807 patients. Seventeen of these 23 cohorts were part of randomised controlled trials, and most reported mid-term follow-up. None of these studies found a relevant difference in functional and patient reported outcomes (appendix 2), nor did data synthesis (pooled estimates Harris Hip Score after short term, mid-term, and long term follow-up −1.7 to 0.2 points, fig 1⇓). Squeaking (an audible, component related noise), known to be bothersome to some patients, was reported in 0-8% of the CoC articulations and absent in MoP and CoP articulations. Implant fracture was more common in CoC articulations. Similar replacement dislocation rates were reported in short term and mid-term follow-up. Six studies, including three randomised controlled trials, had long term follow-up (≥10 years). In general, less wear was measured in CoC articulations, and two studies reported a higher late dislocation rate and subsequent revision rate in the MoP or CoP articulation compared with CoC articulations (one randomised controlled trial with 140 total hip replacements, one retrospective comparison in 126 patients with bilateral total hip replacement), although in both studies early generation polyethylene was used. The other four studies (2042 total hip replacements), including two comparing CoC with CoP articulations using the newer, highly cross-linked polyethylene, reported similar long term survivorship for bearing related revisions.

Study quality

High (1), moderate to high (4), moderate (6), low to moderate (7), low (5).

Safety and survival

The Australian registry reported a slightly higher 10 year revision risk of 1.09% (95% confidence interval 1.02 to 1.16) associated with CoC articulations when 50 533 CoC total hip replacements (cumulative revision percent 5.3% (5.0 to 5.6) were compared with 10 078 MoP replacements (cumulative revision percent 4.6% (4.4 to 4.8). The revision risk of CoC replacements was similar to that of 21 192 CoP replacements (5.3% (4.6 to 6.0).

The New Zealand Joint Registry reported a higher revision rate associated with CoC articulations over a 12 year period when 8177 CoC replacements were compared with 54 637 MoP replacements (revision rate per 100 component years 0.71 (0.62 to 0.81) versus 0.65 (0.63 to 0.68)). The revision risk of CoC replacements was similar to 14 093 CoP replacements (0.68 (0.62 to 0.74)).

The National Joint Registry of England and Wales reported a seven year revision rate of 4.3% (3.8 to 4.9) associated with 35 170 CoC total hip replacements, which was higher than the 3.4% (3.3 to 3.6) revision rate associated with 179 761 MoP replacements and 3.3% (2.9 to 3.7) associated with 30 704 CoP replacements.[6] This was also shown by analysis of all 35 386 most used, single brand, uncemented total hip replacements: CoC articulations were associated with a 1.55 (1.07 to 2.26) higher revision risk compared with MoP articulations.[21] In all three reports, however, it was noted that CoC replacements with small diameter heads (<28 mm) were particularly problematic and at least partially responsible for the increased revision rate when compared with MoP replacements. In addition, substantial interaction effects of fixation type and recipient sex have been found, suggesting survival of CoC replacements may be higher in specific subgroups who may benefit from CoC total hip replacement.[22]

A large US study which compared 5252 CoC total hip replacements with 93 929 MoP replacements in the Medicare database found no difference in incidence of thrombosis, dislocation, infection, loosening, periprosthetic fracture, or revision after a mean of two years of follow-up (relative risks 0.84 to 1.15).[23 24]

Total hip replacements using a stem with a proximal modular femoral neck

Comparative effectiveness

After review of 1966 identified abstracts, four retrospective comparative (non-randomised) studies were eligible for inclusion (appendix 3). Three studies had only short term follow-up (maximum two years) and compared dislocation rates as a secondary outcome: the dislocation rate in 1155 proximal modular total hip replacements (0.8% to 4.2%) was similar to the dislocation rate in 532 proximally non-modular replacements (1.8% to 4.2%). One of these three studies reported a two year postoperative Harris Hip Score, which was similar in both groups (n=466).

The fourth study reported the 14.5 year follow-up of a matched cohort of 163 modular or non-modular total hip replacements in 133 patients operated for symptomatic osteoarthritis secondary to developmental dysplasia of the hip (mean age 53.8 year). They found significantly, but not always clinically relevant, better outcomes in patients who received a modular

replacement (Harris Hip Score and abduction). No difference in complications or revisions was found.

Study quality

Moderate (1), moderate to low (2), and low (1).

Safety and survival

The Australian registry reported high failure rates in all registered 8971 primary proximal modular neck total hip replacements undertaken for symptomatic osteoarthritis (19 different brands): five and 10 year cumulative revisions were 7.4% (95% confidence interval 6.7 to 8.1) and 10.8% (9.6 to 12.1). Compared with 212 800 conventional replacements with a proximally non-modular neck (five and 10 year survival 3.7% (3.6 to 3.8) and 6.4% (6.3 to 6.6)), the revision risk was, irrespective of the type of articulation used, roughly doubled after correction for age and sex (10 year period hazard ratio 1.92 (1.76 to 2.10)). The Italian registry reported the implantation of 25 094 primary proximal modular neck replacements, but specific survival data for patients with primary osteoarthritis was not provided.

Total hip replacements with an uncemented monoblock acetabular cup

Comparative effectiveness

Of 1455 identified abstracts, five studies met the inclusion criteria and could be analysed (appendix 4). Three studies were randomised controlled trials, and two studies were retrospective comparisons of non-consecutive matched cohorts. Together they reported on 284 total hip replacements with an uncemented monoblock cup (all with polyethylene inner lining) in 281 patients after two to 5.6 years of follow-up. Only the randomised studies (178 monoblock cups in 178 patients after two to four years of follow-up) reported functional and patient reported outcomes, and these were found to be similar in total hip replacements with monoblock and modular cup designs. Also, bone mineral density around the monoblock cups was not higher. No difference in survival or complications was found.

Five additional studies met the inclusion criteria for analysis but reported on first generation uncemented monoblock acetabular components, which are no longer in use, and were excluded from analysis (appendix 4). All five studies were retrospective non-randomised studies of low quality, and none of these studies had the objective to compare the outcome of monoblock and modular cups. The only functional outcome measure reported by two studies was the Harris Hip Score, which was comparable between patients with monoblock and modular cups after five to 15 years of follow-up. Four studies investigated 5-15 year survival and one study investigated the 2-14 year wear rate; no difference was found between monoblock and modular cups. Thus, although the second generation uncemented monoblock cups are substantially different, based on these experiences a rationale for the renewed interest cannot objectified.

Study quality

Moderate to high (1), low to moderate (1), low (3).

Safety and survival

The New Zealand registry reported revision rates per 100 observed component years of in total 10 501 uncemented monoblock cups (MoP and CoP articulations), which varied between 0.23 (95% confidence interval 0.06 to 0.59) and 0.64 (0.45 to 0.88) and which were similar to or slightly better than revision rates with uncemented modular variants (0.64 (0.50 to 0.81) to 0.76 (0.65 to 0.88), n=35 650). The Swedish Hip Registry reported a five year survival of 95% (91 to 98) of all 210 uncemented monoblock cups in the registry, which was similar to the survival of the chosen reference, which consisted of all 1130 most commonly used uncemented metal-backed cup designs implanted during the same period (five year survival 97% (96 to 98)).[25] The Finnish registry reported 100% survival of 136 implanted Morcher uncemented monoblock cups. One large US single-institutional survival registry (n=9584 total hip replacements, Mayo clinics) reported similar survivorship of in total 634 monoblock cups of three different brands after 4.7 to 8.2 years of follow-up when compared with the reference cup (Harris Galante).[26]

High flexion total knee replacements

Comparative effectiveness

After assessment of 2410 search hits, 56 studies describing 52 unique cohorts could be included (appendix 5). Combined, 2851 high flexion total knee replacements of various types and brands were compared with 3872 conventional knee replacements. In 25 cohorts the high flexion variant of the widely used NexGen design was investigated (1210 high flexion replacements), and in 11 cohorts the PFC Sigma design was investigated (612 high flexion replacements). Of the included studies, 19 were randomised controlled trials, which together compared 1083 high flexion replacements with 1062 conventional replacements after one to 11 years of follow-up.

Pooled estimates after short term follow-up showed a statistically significant but clinically irrelevant increase in postoperative flexion with high flexion replacement (3.7° (95% confidence interval 2.0 to 5.4) in 26 non-randomised studies, and 1.6° (0.3 to 2.8) in 16 randomised studies) (fig 2⇓). Pooled estimates after mid-term and long term follow-up showed no statistically significant improvement in postoperative flexion with high flexion replacements (fig 2). No relevant increase in other postoperative outcome measures was found (Knee Society Score and Hospital for Special Surgery Score, appendix 5). No difference in survival or complications was reported.

Study quality

High (7), moderate to high (6), moderate (7), low to moderate (10), low (22).

Safety and survival

Only high flexion designs of the NexGen total knee replacement were represented in large numbers in registry reports. The Australian registry reported a 10 year cumulative percent revision for the high flexion version of cemented posterior-stabilised (PS) prostheses of 5.6% (4.9 to 6.3, n=19 326), which was higher than the percent revision of conventional cemented PS prostheses (4.6% (4.0 to 5.4), n=4948) but similar to the overall primary total knee replacement cumulative revision percent (5.5% (5.4 to 5.7), n=342 574). In cruciate ligament-retaining (CR) designs, five year survival of high flexion implants (n=21 088) was similar to that of conventional designs (n=10 316) irrespective of the type of fixation and was generally better than PS designs.

The New Zealand registry reported that, in cemented PS designs, the revision rate of the high flexion prostheses was 0.74 per 100 component years (0.58 to 0.92, n=3665), which was higher, but not statistically significantly higher, than the rate of conventional

prostheses (0.59 (0.46 to 0.74), n=2362), and the high flexion design was indicated to have a significantly higher revision rate than the overall primary total knee replacement revision rate of 0.52 per 100 component years (n=52 058). The revision rate of high flexion cemented CR prostheses of 0.56 per 100 component years (0.37 to 0.81, n=2,068) was also higher, but not statistically significantly higher, than that of conventional cemented CR designs (0.37 (0.29 to 0.47), n=2,723).

The Danish Knee Registry reported on 3998 total knee replacements with high flexion components. Ten year survival of the high flexion cemented PS prosthesis was 88.0% (79.9 to 97.0, n=234), similar to that of the conventional cemented PS prosthesis (92.4% (88.1 to 97.0), n=728) and the overall 10 year survival of primary total knee replacements (92.2% (91.9 to 92.4), n=78 911). Five year survival of the high flexion cemented CR prosthesis was 93.7% (91.6 to 95.8, n=1045), which was lower than that of the conventional cemented CR design (96.3% (95.7 to 96.8), n=5794) but similar to that of the overall primary total knee replacement (94.6% (94.4 to 94.8)). The Swedish Knee Registry noted that 19% of all NexGen primary total knee replacements (n=18 697), which was the most commonly used and best performing total knee replacement in Sweden in 2010, were high flexion designs. However, no specific survival of these high flexion designs was reported.

One large US registry based study of 64 017 total knee replacements with a median follow-up of 2.9 years reported an hazard ratio for revision of 1.76 (1.29 to 2.49) for high flexion designs (n=7810) compared with convention designs.[27]

Gender specific total knee replacements

Comparative effectiveness

Of the 1815 abstracts identified, 11 studies describing 10 unique patients cohorts met the inclusion criteria and could be included for analysis (appendix 6). All but one study investigated the same NexGen gender specific implant and together reported on 1879 total knee replacements in 1396 patients. Six studies were randomised studies that included only women, and five of them considered only patients who received bilateral total knee replacement in one surgical session, that is, one gender specific knee replacement and one conventional (unisex) replacement. The maximum follow-up was three years. Seven studies found no difference in functional and patient reported outcomes between gender specific total knee replacements and unisex total knee replacements, but two studies reported greater postoperative flexion-extension range of motion and one study reported a higher postoperative Knee Society Score. Meta-analysis showed a clinically insignificant increase in postoperative range of motion (pooled estimate 1.9° (0.8 to 3.1), P=0.001, fig 3⇓) and a comparable Knee Society Score and patient preference (appendix 6). No differences in complications or survival were reported by both groups.

Study quality

High (1), moderate to high (3), moderate (3), moderate to low (1), low (2).

Safety and survival

The Swedish Knee Registry noted that 4% of all NexGen primary total knee replacements (n=18 697) were gender specific designs. However, no specific survival of the gender specific designs was reported. The Danish Knee Registry reported the use of 541 NexGen gender specific designs; a five year survival of 92.8% (0.0 to 100%) was reported in 126 implants, which was comparable to that of 670 unisex counterparts (92.9% (90.5 to 95.4)). A large US registry based study of 64 000 total knee replacements found similar revision rates per 100 component years with gender specific implants (n=3376) and with their unisex counterparts (n=2908).[28]

Discussion

Principal findings

In this systematic appraisal of the evidence for the commercial introduction of five recent, well known and innovative implantable device technologies in orthopaedics, we found no clinically relevant improved benefits for these devices compared with older and established alternative implants. Furthermore, none of these five technologies was found to be safer or to have better survival than the established implants. In fact, some of the technologies (ceramic-on-ceramic articulations and modular femoral necks in total hip replacement and high flexion implants in total knee replacement) had inferior survival, which for modular femoral necks in hip replacement was substantial and could be on a scale similar to the failure of large head size metal-on-metal total hip replacements.[1-5] This result indicates that widespread and ongoing dissemination of these technologies cannot be justified from an evidence based perspective. We suggest that adoption and widespread use of these and other new implants or related technologies first requires thorough scrutiny by surgeons, regulators, and patients.

Context and rationale

This systematic review was part of a US Food and Drug Administration (FDA) initiated project on comparative effectiveness and safety of orthopaedic implants as a reaction to the recent disaster with metal-on-metal articulations.[1-9] Five innovative and relatively recent and ongoing technologies in major joint replacement were selected in order to investigate the past process of introduction of new orthopaedic technologies. All selected technologies had substantial potential for meaningful improvement, but also potential for an increased risk of adverse outcomes such as implant failure and inferior long term survival. Some technologies were novel, such as the gender specific total knee replacement, while others were recently re-introduced in their second iteration, such as the monoblock acetabular components, or are a topic of ongoing debate, such as ceramic-on-ceramic articulations in total hip replacement. Also, some technologies represent completely new designs (modular femoral neck implants), while others represent relatively minor implant modifications (high flexion total knee replacements). Consequently, this systematic appraisal covers a range of innovative technologies that is representative for the orthopaedic field of technology development.

Policy implications

The search for implants with superior functional outcome, increased longevity, and better survival is needed. Hence, this review is not intended to criticise the surgical community or orthopaedic industry. Our goals are to highlight that the status quo regarding the introduction of new device technologies is not acceptable, that substantial efforts are needed by all stakeholders to invest in systems of careful evaluation and to promote controlled and evidence based introduction of device innovations. Safety of new devices can only be established by the recording of adverse events and implant survival in large numbers of patients such as those included in registries. This is particularly important when the comparative benefits of new device related technologies have not been convincingly

demonstrated. Exposing large numbers of patients to new devices with uncertain safety could be considered unethical and should be avoided without controlled introduction and proper systems of evaluation. In this context, stepwise introduction of device related technologies has been proposed,[29 30] and new high accuracy measurement methods for implant performance have been standardised. For example, Roentgen stereophotogrammetry[30] is mandatory in some countries.[31] Registries and registry consortia, such as the International Consortium of Orthopaedic Registries,[32] are important and critical developments but they cannot influence the paths and loopholes associated with introduction of device related technologies on the commercial market. For now, the need for new device technologies and the associated expected benefits need to be carefully balanced against the possible associated risks. We suggest adherence to the IDEAL recommendations,[33-37] and when clinically relevant improvements have not been convincingly shown widespread commercial introduction should be deferred until appropriate long term safety comparable to existing technologies has been demonstrated.

The issues related to uncontrolled introduction of device technologies and the need for proper surveillance are not specific to orthopaedics. Recent issues with breast implants, cardiac leads, and urogynaecologic meshes suggest that this is a healthcare system issue in the Western world that requires major overhaul.[38 39] The FDA has recently recognised this and initiated a new vision to achieve this goal,[11] but appropriate actions may be necessary at a higher level in the US and worldwide.

Study limitations

Our study has several limitations. First, the evaluated device related technologies represent a selection of all introduced technologies. This was based on the authors' wish to cover a relevant, diverse, and representative selection, but the technologies were not chosen at random. They were, however, chosen without substantial knowledge of comparative effectiveness and registry reported survival and, to our best knowledge, are the five most widely marketed technologies adhering to the prespecified criteria. The unavailability of a suitable close reference device solution for comparison excluded several large innovations, such as resurfacing total hip replacements and unicompartimental knee arthroplasty. Highly cross-linked polyethylene in total hip replacement, for which results so far have been favourable, could not be evaluated in this study since this innovation can only be evaluated by long term wear and survival. Two other large innovative technologies, the mobile bearing total knee replacement and the metal-on-metal articulations in total hip replacement, were not selected since these have been extensively investigated elsewhere, but were not found to have any advantages either.[2 22 40]

Second, as with any systematic review or meta-analysis, the strength of our conclusions is limited by the quality of the individual studies. The number of included studies was (depending on the development) limited, the quality of the individual studies varied widely, and evident (study) heterogeneity was present.

Third, effects of a possible learning curve for new device technologies cannot be ruled out. However, all chosen developments represent important improvements on existing technologies but do not require a major overhaul of surgical technique. Hence, this effect is expected to be limited.

Finally, because of the observational nature of registry data, a comparison of survival of implants in registries is always subject to a substantial risk of introducing selection bias, since different implants may be chosen for different patients and case-mix differences may be present. Although most registry reports provide survival estimates based on large numbers of implants and adjusted for potential confounders, and despite our attempt to select appropriate reference groups for comparison, this can never be fully accounted for.

Conclusion

This systematic appraisal of the evidence based introduction of new orthopaedic device technologies indicates that new technologies are being introduced to the commercial market without sufficient high quality evidence for improved benefit over existing, well proven, and safe alternative implant solutions. Furthermore, the safety of several new technologies could be (substantially) compromised. Combined with recent disasters, we advise that actions should be undertaken by all stakeholders to prevent patients from being further exposed to new device related technologies without proper evidence of improved clinical benefit and safety.

We thank Anna Gambaryan (AG), MPH, Lucas Romero (LR), MPA; Samantha Jacobs (SJ), BS; Sofia Lang Perez (SLP), BS for their important contributions to the review process.

Contributors: MJN, RGHHN, and AR were responsible for study design and concept. MJN, AR, and JWS acquired the data, which were analysed by MJN and AR. MJN, RGHHN, and AR interpreted the results. MJN and AR drafted the manuscript and are the study guarantors. MJN and AR provided statistical expertise. RGHHN, AR, and D Marinac-Dabic DM-D provided administrative, technical, or material support. All authors critically revised the manuscript and revision for important intellectual content. The authors of this article are responsible for its contents. No statement in this article should be construed as an official position of the US Food and Drug Administration.

Funding: MJN received a fellowship from the US FDA Medical Device Epidemiology Network's (MDEpiNet) Science and Infrastructure Center (director AS) under the International Consortium of Orthopaedic Registries initiative (ICOR, www.icor-initiative.org). AS received funding from the US FDA for establishing the MDEpiNet Science and Infrastructure Center and is the principal investigator of ICOR contract.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The lead authors (the manuscript's guarantors, MJN and AS) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 MHRA. Medical device alert: all metal-on-metal (MoM) hip replacements. 2012. www.mhra.gov.uk/Publications/Safetywarnings/MedicalDeviceAlerts/CON155761.
2 Sedrakyan A, Normand SL, Dabic S, Jacobs S, Graves S, Marinac-Dabic D. Comparative assessment of implantable hip devices with different bearing surfaces: systematic appraisal of evidence. *BMJ* 2011;343:d7434.
3 Cohen D. Out of joint: the story of the ASR. *BMJ* 2011;342:d2905.
4 Meier B. With warning, a hip device is withdrawn. *New York Times* 2011 March 10.
5 Godlee F. The trouble with medical devices. *BMJ* 2011;342:d3123.
6 Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW, National Joint Registry of England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet* 2012;379:1199-204.
7 Graves SE, Rothwell A, Tucker K, Jacobs JJ, Sedrakyan A. A multinational assessment of metal-on-metal bearings in hip replacement. *J Bone Joint Surg Am* 2011;93(suppl 3):43-7.

What is already known on this topic

The introduction of new orthopaedic implants and related technologies has been the focus of scientific and policy discussions since the failures of articular surface replacement and large head size metal-on-metal articulations in total hip replacement

It is unclear whether these failures were isolated events or if there is a systemic problem affecting the introduction of a much wider range of implantable devices.

What this study adds

This systematic appraisal of the evidence based introduction of new orthopaedic device technologies indicates that new technologies are being introduced to the commercial market without sufficient high quality evidence for improved benefit over existing, well proven, and safe alternatives. Moreover, the existing devices may be safer to use in total hip and knee replacement

Improved regulation and professional society oversight are necessary to prevent patients from being further exposed to these or future innovations without proper evidence of improved clinical efficacy and safety

8 Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. *N Engl J Med* 2012;366:848-55.
9 Sedrakyan A. Metal-on-metal failures—in science, regulation, and policy. *Lancet* 2012;379:1174-6.
10 Skinner JA, Kay PR, Hart AJ. Introducing new joint replacements to clinical practice. *BMJ* 2011;343:d8188.
11 US Food and Drug Administration (FDA). National medical device postmarket surveillance plan. www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/ucm301912.htm.
12 Mission statement of the International Society of Arthroplasty Registries. www.isarhome.org/statements.
13 Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
14 Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453-7.
15 *Cochrane handbook for systematic reviews of interventions* . Version 5.1.0. 2011. http://handbook.cochrane.org.
16 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
17 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
18 Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13.
19 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
20 Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.
21 Jameson SS, Baker PN, Mason J, Rymaszewska M, Gregg PJ, Deehan DJ, et al. Independent predictors of failure up to 7.5 years after 35 386 single-brand cementless total hip replacements: a retrospective cohort study using National Joint Registry data. *Bone Joint J* 2013;95-B:747-57.
22 Smith AJ, Dieppe P, Howard PW, Blom AW; National Joint Registry for England and Wales. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. *Lancet* 2012;380:1759-66.
23 Bozic KJ, Ong K, Lau E, Kurtz SM, Vail TP, Rubash HE, et al. Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces. *Clin Orthop Relat Res* 2010;468:2357-62.
24 Bozic KJ, Lau EC, Ong KL, Vail TP, Rubash HE, Berry DJ. Comparative effectiveness of metal-on-metal and metal-on-polyethylene bearings in Medicare total hip arthroplasty patients. *J Arthroplasty* 2012;27(8 suppl):37-40.
25 Weiss RJ, Hailer NP, Stark A, Kärrholm J. Survival of uncemented acetabular monoblock cups: evaluation of 210 hips in the Swedish Hip Arthroplasty Register. *Acta Orthop* 2012;83:214-9.
26 Howard JL, Kremers HM, Loechler YA, Schleck CD, Harmsen WS, Berry DJ, et al. Comparative survival of uncemented acetabular components following primary total hip arthroplasty. *J Bone Joint Surg Am* 2011;93:1597-604.
27 Namba RS, Inacio MC, Cafri G. Increased risk of revision for high flexion total knee replacement with thicker tibial liners. *Bone Joint J* 2014;96-B:217-23.
28 Namba RS, Cafri G, Khatod M, Inacio MC, Brox TW, Paxton EW. Risk factors for total knee arthroplasty aseptic revision. *J Arthroplasty* 2013;28(8 suppl):122-7.
29 Malchau H. Introducing new technology: a stepwise algorithm. *Spine* 2000;25:285.
30 Nelissen RG, Pijls BG, Kärrholm J, Malchau H, Nieuwenhuijse MJ, Valstar ER. RSA and registries: the quest for phased introduction of new implants. *J Bone Joint Surg Am* 2011;93(suppl 3):62-5.
31 Valstar ER, Kaptein BL, Nelissen RG. Radiostereometry and new prostheses. *Acta Orthop* 2012;83:103-4.
32 ICOR Collaborators. The International Consortium of Orthopaedic Registries (ICOR) effort: a worldwide collaboration to advance the research methods and evidence in orthopaedics. *J Bone Joint Surg* 2011;93-A(suppl 3E):1-12, 43-47.
33 McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105-12.
34 McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK; IDEAL Group. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013;346:f3012.
35 Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG; IDEAL Group. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* 2013;346:f3011.
36 Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A; IDEAL Group. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* 2013;346:f2820.
37 McCulloch P, Barkun J, Sedrakyan A; IDEAL Collaboration. Implantable device regulation in Europe. *Lancet* 2012;380:729.
38 Horton R. Offline: the scandal of device regulation in the UK. *Lancet* 2012;379:204.
39 O'Dowd A. UK launches inquiry into safety of PIP breast implants. *BMJ* 2012;344:e11.
40 Namba RS, Inacio MC, Paxton EW, Robertsson O, Graves SE. The role of registry data in the evaluation of mobile-bearing total knee arthroplasty. *J Bone Joint Surg Am* 2011;93(suppl 3):48-50.

Accepted: 17 July 2014

Cite this as: *BMJ* 2014;349:g5133

Tables

Table 1| Details of selected device innovations

Development	Altered design features	Intended biomechanical benefits	Expected clinical benefit	Development-specific potential risks
Total hip replacement				
Ceramic-on-ceramic bearing	Material of articulating surfaces (cup liner and head) are alumina based ceramics, larger head sizes possible.	Increased longevity and less wear particles with subsequent less wear-induced osteolysis and lower loosening rates	Improved postoperative functional and patient-reported outcomes, lower dislocation and loosening rates	Material fracture due to vulnerability to point loading, squeaking (audible component related noise), liner chipping or canting
Modular femoral neck	Modularity between the stem and neck of the femoral implant, in addition to potential neck-head modularity; modular neck components come in various sizes, angulation, and lengths	Intraoperative adjustment of dimensions for optimal offset, leg length, and anteversion to reduce dislocation, impingement, and wear and allow better muscle balance	Improved postoperative functional and patient reported outcomes, reduced rates of dislocation and loosening, smaller incisions	Fracture, dissociation, corrosion (fretting, crevice, galvanic) with metal ion generation and potential pseudo-tumour formation, component mismatch
Uncemented monoblock acetebular component	Polyethylene liner is moulded into outer metal shell and the cup is a monoblock (non-modular) component which, additionally, allows thicker polyethylene, hemi-elliptical designs en tantalum trabecular metal designs	Prevention of potential (micro)motion induced backside wear and subsequent acetabular osteolysis and cup loosening, prevention of liner dislocation, longer time to revision for wear, less loosening due to more physiological acetabular bone loading due to hemi-elliptical shape and less rigid metal shells	Lower rates of liner dislocations and less concern about wear, resulting in improved postoperative functional and patient reported outcomes, less liner dislocations, and lower revision rates	More demanding implantation, higher risk of improper seating of cup in acetabular bone, intraoperative dimensional changes (that is, offset liners) not possible, isolated liner exchange not possible when worn out, with subsequent difficult revision procedure
Total knee replacement				
High-flexion components	Varies from a combination of extension of the posterior condyle of the femoral component with modifications to the cam and tibial spine and reduced anterior thickness of the polyethylene insert to isolated insert modifications	Increased articular contact during high flexion, improved stability and subluxation resistance, decreased stress on quadriceps mechanism, reduced risk of patellar impingement	Increased flexion and decreased incidence of anterior knee pain resulting in better functional and patient reported outcomes	Decreased stability due to higher stress at the cement-implant interface, increased edge loading and higher polyethylene wear, increased patellar impingement
Gender specific components	Enhanced aspect ratio: narrower medial-lateral dimension for a given anteroposterior dimension, reduced anterior flange, recessed and lateralized patellar sulcus	More close match to female anatomy, decreased rate of patellar (sub)luxations, less soft-tissue irritation or imbalance	Increased range of motion and decreased rates of knee pain and patella (sub)luxations resulting in better functional and patient reported outcomes	Unclear and not (yet) described; at least lack of benefit and cost effectiveness due to absence of evidence that women benefit less from unisex knee replacement than men, unknown risks

Table 2| Overview of results from systematic review of trials, comparative studies, and registries for comparative effectiveness and safety of five implantable device innovations

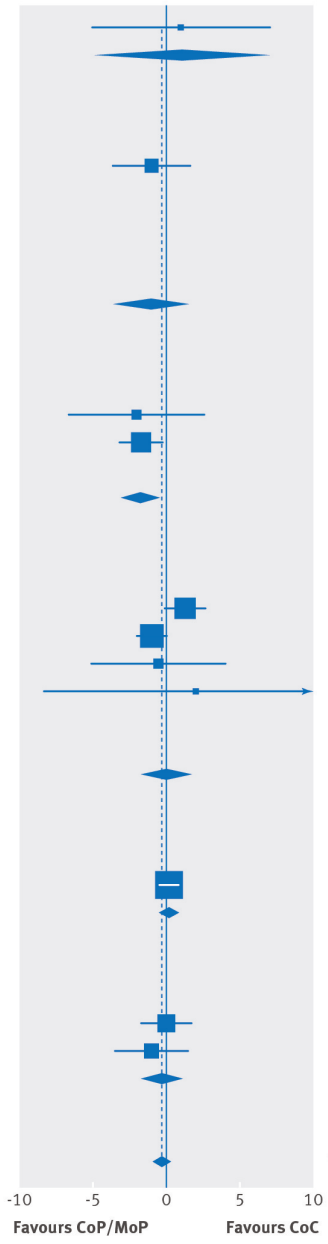
	Total hip replacement			Total knee replacement	
Device innovation	**Ceramic-on-ceramic articulation**	**Modular femoral neck**	**Uncemented monoblock acetebular component**	**High flexion components**	**Gender specific components**
Comparative effectiveness					
No of included cohorts (in studies)	23 (in 42)	4 (in 4)	5 (in 5)	52 (in 56)	10 (in 11)
No of implants (patients)	5442 (4807)	1730 (1700)	546 (540)	6835 (5769)	1879 (1396)
Follow-up term No of cohorts)	Short term (7) Mid-term (10) Long term (6)	Short term (2) Long term (1) Unknown (1)	Short term (2) Mid-term (3)	Short term (41) Mid-term (9) Long term (2)	Short term (10)
Study quality (No of cohorts)	High (1) Moderate to high (4) Moderate (6) Low to moderate (7) Low (5)	Moderate (1) Low to moderate (2) Low (1)	Moderate to high (1) Low to moderate (1) Low (3)	High (7) Moderate to high (6) Moderate (7) Low to moderate (10) Low (22)	High (1) Moderate to high (3) Moderate (3) Low to moderate (1) Low (2)
Main reported outcomes (No of cohorts)	Harris Hip Score (16) WOMAC (5) Squeaking (10)	Harris Hip Score (2) Hip flexion (1) Dislocation rate (4)	Harris Hip Score (2) Oxford Hip Score (2) Preference (2)	Knee flexion (52) Knee Society Score (32) Knee Society function (26)	Knee range of motion (10) Knee Society Score (6) WOMAC (4)
Results from reported differences and pooled estimates	No significant differences,* squeaking only in CoC group	Harris Hip Score significantly higher, dislocations comparable	No significant differences	Clinically irrelevant increased flexion,* no difference other outcomes*	Clinically irrelevant increased range of motion,* no difference other outcomes*
Conclusion	No evidence for clinically relevant improvement	Insufficient evidence for clinically relevant improvement	No evidence for clinically relevant improvement	No evidence for clinically relevant improvement	No evidence for clinically relevant improvement
Safety and survival					
Registries with relevant data	Australia, New Zealand, UK, Medicare	Australia, Italy	New Zealand, Sweden, US	Australia, New Zealand, Sweden, Denmark, US	Sweden, Denmark, US
No of implants	99 132 CoC 338 405 MoP 75 989 CoP	34 065 modular neck 212 800 conventional	11 345 monoblock cups 37 142 modular cups	57 955 high flexion 537 560 conventional	3917 gender specific 3578 conventional
Revision rate	Slightly higher compared with MoP (HR 1.0 to 1.55)	Nearly doubled (HR 1.92)	Comparable	Slightly higher for most common high flexion designs (HR 1.0 to 1.76)	Comparable
Complications	Comparable	Higher rate of dislocation and implant fracture	Not reported	Not reported	Not reported
Conclusion	Revision rate may be higher	Revision and complication rate higher	Comparable revision rate	Revision rate may be higher	Comparable revision rate

WOMAC=Western Ontario and McMaster Universities Arthritis Index, CoC=ceramic-on-ceramic, MoP=metal-on-polyethylene, CoP=ceramic-on-polyethylene, HR=hazard ratio

*Includes pooled estimates.

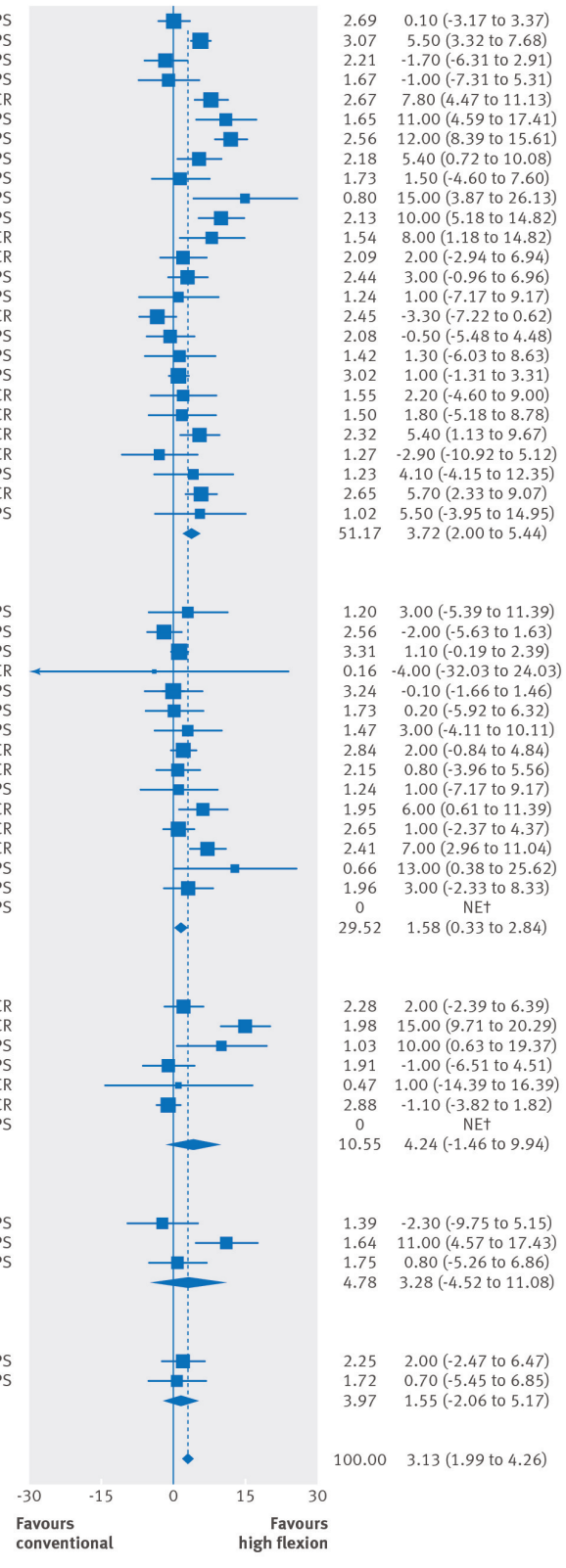
Figures

Trial	Reference device: No of implants	Reference device: Mean (SD) score	CoC device: No of implants	CoC device: Mean (SD) score	Follow-up time (years)	Reference bearing	Weight* (%)	Mean difference (95% CI)
Non-randomized, short term								
Plate et al 2013	30	90.0 (NA)	30	91.0 (NA)	3.3	MoP	1.09	1.00 (-5.08 to 7.08)
Subtotal							1.09	1.00 (-5.08 to 7.08)
Test for heterogeneity: I^2=not estimable								
Test for overall effect: z=0.32, P=0.747								
Randomized, short term								
Poggie et al 2007	137	93.0 (12)	315	92.0 (16)	2.0	CoP	5.28	-1.00 (-3.61 to 1.61)
Bal et al 2005	215	NA (NA)	229	NA (NA)	2.0	CoP	0	Not estimable†
D'Antonio et al 2002	151	96.0 (NA)	338	94.0 (NA)	3.0	MoP	0	Not estimable†
Hamilton et al 2010	76	93.8 (NA)	157	94.4 (NA)	2.6	CoP	0	Not estimable†
Schwämmlein et al 2002	25	93.7 (NA)	25	94.1 (NA)	2.0	CoP	0	Not estimable†
Subtotal							5.28	-1.00 (-3.61 to 1.61)
Test for heterogeneity: I^2=not estimable								
Test for overall effect: z=0.75, P=0.747								
Non-randomized, mid-term								
Lombardi et al 2010	44	92.0 (13)	64	90.0 (12)	6.1	CoP	1.83	-2.00 (-6.65 to 2.67)
Schmidt et al 2002	30	94.7 (3)	30	93.0 (3)	5.6	CoP	12.63	-1.70 (-3.19 to -0.21)
Capello et al 2005 (B)	95	96.4 (NA)	186	95.7 (NA)	6.5	MoP	0	Not estimable†
Subtotal							14.46	-1.73 (-3.15 to -0.31)
Test for heterogeneity: I^2=0.0%, P=0.904								
Test for overall effect: z=2.39, P=0.017								
Randomized, mid-term								
Bascarevic et al 2010	75	93.8 (5)	82	95.1 (4)	4.2	MoP	13.81	1.30 (-0.09 to 2.69)
Kim et al 2009	100	94.0 (4)	100	93.0 (4)	5.6	CoP	20.08	-1.00 (-2.00 to 0.00)
Nikolau et al 2012	32	91.5 (9)	34	91.0 (10)	5.0	MoP	1.88	-0.50 (-0.51 to 4.09)
Ochs et al 2007	21	89.0 (12)	22	91.0 (22)	8.1	CoP	0.38	2.00 (-8.41 to 12.41)
Amantullah et al 2011	95	NA (NA)	125	NA (NA)	5.0	CoP	0	Not estimable†
Capello et al 2005 (A)	101	97.0 (NA)	213	96.8 (NA)	5.0	MoP	0	Not estimable†
Subtotal							36.14	0.06 (-1.66 to 1.77)
Test for heterogeneity: I^2=57.8%, P=0.068								
Test for overall effect: z=0.07, P=0.948								
Non-randomized, long term								
Epinette et al 2014	120	98.3 (3.9)	163	98.5 (3.0)	10.9	CoP	26.97	0.20 (-0.49 to 0.89)
Subtotal							26.97	0.20 (-0.49 to 0.89)
Test for heterogeneity: I^2=not estimable								
Test for overall effect: z=0.57, P=0.568								
Randomized, long term								
D'Antonio et al 2012	72	96.5 (6)	144	96.5 (7)	10.3	MoP	10.26	0.00 (-1.72 to 1.72)
Kim et al 2013	100	95.0 (NA)	100	94.0 (NA)	12.4	CoP	5.79	-1.00 (-3.47 to 1.47)
Subtotal							16.06	-0.33 (-1.74 to 1.09)
Test for heterogeneity: I^2=0.0%, P=0.515								
Test for overall effect: z=0.45, P=0.650								
Overall							100.00	-0.32 (-0.96 to 0.32)
Test for heterogeneity: I^2=27.8%, P=0.180								
Test for overall effect: z=0.97, P=0.330								



CoC=ceramic-on-ceramic, MoP=metal-on-polyethylene, CoP=ceramic-on-polyethylene
*Weights from random effects analysis
†Not estimable because unavailability of mean or median or of SD, SE, range, and P value

Fig 1 Forest plot for comparison of Harris Hip Score in patients with ceramic-on-ceramic total hip replacement and metal-on-polyethylene or ceramic-on-polyethylene bearings.

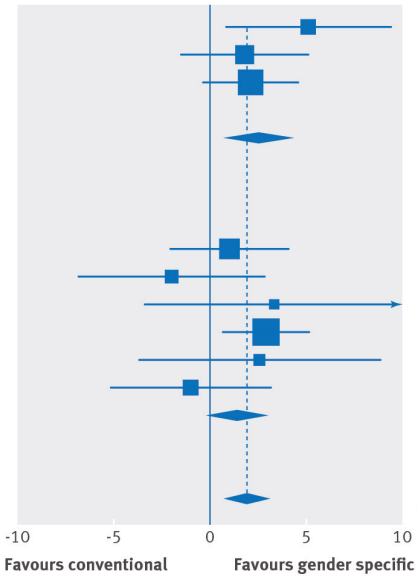


Trial	Reference device: No of implants	Reference device: Mean (SD) score	High flexion device: No of implants	High flexion device: Mean (SD) score	Follow-up time (years)	Design	Weight* (%)	Mean difference (95% CI)
Non-randomized, short term								
Bajammal et al 2006	87	113.5 (12)	87	113.6 (11)	1.0	PS/PS	2.69	0.10 (-3.17 to 3.37)
Bin et al 2007	90	124.3 (9)	90	129.8 (5)	1.0	PS/PS	3.07	5.50 (3.32 to 7.68)
Boese et al 2011	64	116.4 (NA)	64	114.7 (NA)	0.8	PS/PS	2.21	-1.70 (-6.31 to 2.91)
Coughlin et al 2007	11	129.0 (8)	8	128.0 (5)	3.0	PS/PS	1.67	-1.00 (-7.31 to 5.31)
Crow et al 2010	65	112.0 (11)	79	119.8 (9)	1.0	CR/CR	2.67	7.80 (4.47 to 11.13)
Gupta et al 2006	50	116.0 (10)	50	125.0 (11)	1.0	PS/PS	1.65	11.00 (4.59 to 17.41)
Huang et al 2005	25	126.0 (5)	25	138.0 (6)	2.0	PS/PS	2.56	12.00 (8.39 to 15.61)
Klein et al 2004	37	117.0 (6)	37	122.4 (6)	0.5	PS/PS	2.18	5.40 (0.72 to 10.08)
Kwak et al 2012	61	125.4 (16.2)	50	119.0 (19.1)	2.9	PS/PS	1.73	1.50 (-4.60 to 7.60)
Laskin et al 2007	39	118.0 (NA)	40	133.0 (NA)	2.0	PS/PS	0.80	15.00 (3.87 to 26.13)
Lee et al 2011	39	121.0 (12)	41	131.0 (10)	2.0	PS/PS	2.13	10.00 (5.18 to 14.82)
Mai et al 2013	69	109.0 (12.0)	75	117.0 (8.9)	2.7	CR/CR	1.54	8.00 (1.18 to 14.82)
Malik et al 2010	50	118.0 (14)	50	120.0 (12)	1.0	CR/CR	2.09	2.00 (-2.94 to 6.94)
Massin et al 2010	59	123.0 (10)	59	126.0 (12)	1.0	PS/PS	2.44	3.00 (-0.96 to 6.96)
McCalden et al 2010	1177	116.7 (12)	197	119.7 (13)	1.0	PS/PS	1.24	1.00 (-7.17 to 9.17)
Minoda et al 2009	89	115.3 (13)	87	112.0 (13)	1.0	CR/CR	2.45	-3.30 (-7.22 to 0.62)
Minoda et al 2010	28	125.5 (10)	28	125.0 (11)	2.0	PS/PS	2.08	-0.50 (-5.48 to 4.48)
Nam et al 2012	19	125.0 (12)	16	126.3 (10)	2.0	PS/PS	1.42	1.30 (-6.03 to 8.63)
Ng et al 2008	35	105.0 (13)	35	106.0 (14)	2.9	PS/PS	3.02	1.00 (-1.31 to 3.31)
Seon et al 2005	50	128.5 (16)	50	130.7 (15)	2.0	PS/CR	1.55	2.20 (-4.60 to 9.00)
Song et al 2012	40	134.4 (4)	40	136.2 (6)	2.0	CR/CR	1.50	1.80 (-5.18 to 8.78)
Stormont et al 2009	70	117.0 (11)	45	122.4 (12)	1.0	CR/CR	2.32	5.40 (1.13 to 9.67)
Suggs et al 2009	11	120.6 (11)	15	117.7 (10)	1.1	CR/CR	1.27	-2.90 (-10.92 to 5.12)
Tsai et al 2009	13	117.0 (13)	13	121.1 (8)	1.0	PS/PS	1.23	4.10 (-4.15 to 12.35)
Tsui et al 2009	122	111.3 (17)	112	117.0 (11)	1.0	CR/CR	2.65	5.70 (2.33 to 9.07)
Waisfisz et al 2010	24	119.7 (NA)	22	125.2 (NA)	0.0	PS/PS	1.02	5.50 (-3.95 to 14.95)
Subtotal							51.17	3.72 (2.00 to 5.44)
Test for heterogeneity: I^2=70.8%, P<0.001								
Test for overall effect: z=4.24, P<0.001								
Randomized, short term								
Ahmed et al 2009	28	107.0 (15)	28	110.0 (17)	1.0	PS/PS	1.20	3.00 (-5.39 to 11.39)
Choi et al 2010	74	130.0 (10)	76	128.0 (11)	2.3	PS/PS	2.56	-2.00 (-5.63 to 1.63)
Dennis et al 2013	81	126.8 (9.8)	81	127.8 (9.2)	1.0	PS/PS	3.31	1.10 (-0.19 to 2.39)
Fisher et al 2013	29	119.0 (12)	31	115.0 (11)	1.0	CR/CR	0.16	-4.00 (-32.03 to 24.03)
Guild et al 2014	123	121.0 (9.7)	119	120.9 (10.4)	2.0	PS/PS	3.24	-0.10 (-1.66 to 1.46)
Hamilton et al 2011	62	124.0 (NA)	66	124.2 (NA)	1.0	PS/PS	1.73	0.20 (-5.92 to 6.32)
Kim et al 2005	50	136.0 (11)	50	139.0 (11)	2.1	PS/PS	1.47	3.00 (-4.11 to 10.11)
Kim et al 2009	54	131.0 (14)	54	133.0 (10)	3.1	CR/CR	2.84	2.00 (-0.84 to 4.84)
Lutzer et al 2014	48	108.9 (12.5)	68	109.7 (12.9)	1.0	CR/CR	2.15	0.80 (-3.96 to 5.56)
McCalden et al 2009	49	123.0 (7)	49	124.0 (7)	2.7	PS/PS	1.24	1.00 (-7.17 to 9.17)
Nutton et al 2012	40	107.0 (NA)	36	113.0 (NA)	1.0	PS/CR	1.95	6.00 (0.61 to 11.39)
Seon et al 2009	50	134.3 (8)	50	135.3 (8)	2.0	CR/CR	2.65	1.00 (-2.37 to 4.37)
Thomsen et al 2013	33	120.0 (NA)	33	127.0 (NA)	1.0	PS/CR	2.41	7.00 (2.96 to 11.04)
Weeden et al 2007	25	120.0 (NA)	25	133.0 (NA)	1.0	PS/PS	0.66	13.00 (0.38 to 25.62)
Wohlrab et al 2005	28	109.1 (12)	29	112.1 (9)	2.9	PS/PS	1.96	3.00 (-2.33 to 8.33)
Biant et al 2009	50	106.0 (NA)	50	106.0 (NA)	0.5	PS/PS	0	NE†
Subtotal							29.52	1.58 (0.33 to 2.84)
Test for heterogeneity: I^2=34.7%, P=0.091								
Test for overall effect: z=2.47, P=0.014								
Non-randomized, mid-term								
Endres et al 2011	50	120.0 (13)	57	122.0 (11)	5.7	CR/CR	2.28	2.00 (-2.39 to 6.39)
Gobel et al 2008	22	100.0 (NA)	22	115.0 (NA)	5.0	PS/CR	1.98	15.00 (9.71 to 20.29)
Kim et al 2012 (A)	45	125.0 (7.0)	45	135.0 (6.0)	5.4	PS/PS	1.03	10.00 (0.63 to 19.37)
Lee et al 2012	40	121.0 (11.4)	94	120.0 (16.1)	7.4	PS/PS	1.91	-1.00 (-6.51 to 4.51)
Okamoto et al 2011	10	119.0 (13)	10	120.0 (9)	5.0	CR/CR	0.47	1.00 (-14.39 to 16.39)
Seon et al 2013	48	135.6 (7.0)	89	134.5 (8.1)	6.2	CR/CR	2.88	-1.10 (-3.82 to 1.82)
Li et al 2007	40	109.5 (NA)	15	104.3 (NA)	4.2	PS/PS	0	NE†
Subtotal							10.55	4.24 (-1.46 to 9.94)
Test for heterogeneity: I^2=84.4, P<0.001								
Test for overall effect: z=1.46, P=145								
Randomized, mid-term								
Nieuwenhuijse et al 2013	24	118.5 (11.6)	28	116.2 (15.0)	5.0	PS/PS	1.39	-2.30 (-9.75 to 5.15)
Seng et al 2011	27	117.0 (14)	36	128.0 (12)	5.0	PS/PS	1.64	11.00 (4.57 to 17.43)
Wohlrab et al 2009	19	117.5 (10)	22	116.7 (10)	5.0	PS/PS	1.75	0.80 (-5.26 to 6.86)
Subtotal							4.78	3.28 (-4.52 to 11.08)
Test for heterogeneity: I^2=76.0%, P=0.016								
Test for overall effect: z=0.82, P=0.410								
Randomized, long term								
Kim et al 2012 (B)	100	133.0 (NA)	100	135.0 (NA)	10.3	PS/PS	2.25	2.00 (-2.47 to 6.47)
Radetzki et al 2013	22	114.1 (9.4)	17	114.8 (10.1)	10.8	PS/PS	1.72	0.70 (-5.45 to 6.85)
Subtotal							3.97	1.55 (-2.06 to 5.17)
Test for heterogeneity: I^2=0.0%, P=0.738								
Test for overall effect: z=0.84, P=0.400								
Overall							100.00	3.13 (1.99 to 4.26)
Test for heterogeneity: I^2=69.9%, P<0.001								
Test for overall effect: z=5.39, P<0.001								

PS=posterior-stabilized prosthesis, CR=cruciate ligament-retaining prosthesis
*Weights from random effects analysis
†Not estimable because unavailability of mean or median or of SD, SE, range, and P value

Fig 2 Forest plot for comparison of flexion (in degrees) in high flexion total knee replacement and conventional total knee replacement.

Trial	Reference device: No of implants	Reference device: Mean (SD) score	Gender specific device: No of implants	Gender specific device: Mean (SD) score	Follow-up time (years)	Weight* (%)	Mean difference (95% CI)
Non-randomized, short term							
Eggers et al 2012	39	118.3 (9.7)	39	123.4 (9.8)	1.0	7.12	5.10 (0.79 to 9.41)
Emerson et al 2008	90	114.0 (NA)	106	115.8 (NA)	0.5	11.92	1.80 (-1.53 to 5.13)
Song et al 2012 (A)	40	134.3 (4.0)	40	136.4 (1.9)	0.0	21.70	2.10 (-0.37 to 4.57)
Tanavalee et al 2011	142	133.5 (NA)	157	134.1 (NA)	2.4	0	Not estimable†
Subtotal						40.73	2.54 (0.73 to 4.34)
Test for heterogeneity: I^2=0.0%, P=0.435							
Test for overall effect: z=2.76, P=0.006							
Randomized, short term							
Kim et al 2010 (A)	85	125.0 (NA)	85	126.0 (NA)	2.1	14.15	1.00 (-2.06 to 4.06)
Kim et al 2010 (B)	138	126.0 (NA)	138	124.0 (NA)	3.3	5.60	-2.00 (-6.86 to 2.86)
Klein et al 2011	40	109.6 (15.2)	40	112.9 (14.8)	0.1	2.91	3.30 (-3.44 to 10.04)
Singh et al 2012	100	119.9 (8.7)	100	122.8 (8.0)	2.1	25.60	2.90 (0.63 to 5.17)
Song et al 2012 (B)	46	121.4 (17.0)	46	124.0 (14.8)	2.6	3.34	2.60 (-3.69 to 8.89)
Thomsen et al 2012	24	132.0 (NA)	24	131.0 (NA)	1.0	7.66	-1.00 (-5.16 to 3.16)
Subtotal						59.27	1.41 (-0.17 to 2.99)
Test for heterogeneity: I^2=6.2%, P=0.377							
Test for overall effect: z=1.75, P=0.080							
Overall						100.00	1.91 (0.76 to 3.06)
Test for heterogeneity: I^2=0.0%, P=0.456							
Test for overall effect: z=3.26, P=0.001							



*Weights from random effects analysis
†Not estimable because unavailability of mean or median or of SD, SE, range, and P value

Fig 3 Forest plot for comparison of flexion-extension range of motion (in degrees) in gender specific total knee replacement and conventional total knee replacement