Washington State Health Care Authority

Health Technology Assessment

Artificial Disc Replacement

Peer Review and Public Comments and Responses

September 18, 2008

CONTENTS	Page
Spectrum Research Response to Peer Review Comments	3
Michael J. Lee, MD	3
Sean D. Sullivan, RPh, PhD	3
Ann M. Derleth, PhD	4
Jens R. Chapman, MD	5
Brian M. Drew, MD	5
Spectrum Research Response to Public Comments	5
Industry (DePuy)	5
Industry (Medtronic)	8
Clinician/Professional Organization	10
Spectrum Research Response to Washington State Agency Comments	12
Peer Review Comments	
Michael J. Lee, M.D., Assistant Professor, University of Washington, Spine Service	15
Sean D. Sullivan, RPh, PhD, Professor of Pharmacy and Health Services, Director of the Pharmaceutical Outcomes Research and Policy Program, University of Washington	17
Ann M. Derleth, PhD, Health Services Research Postdoctoral Fellow, VA HSR&D, Seattle, Wa	19
Jens R. Chapman, M.D., Professor; University of Washington, Director, Spine Service	20
Brian M. Drew, M.D., Assistant Clinical Professor, McMaster University, Medical Director, Hamilton General Hospitals Spine Unit	24
Public Comments	29
Industry (DePuy)	29
Industry (Medtronic)	32
Clinician/Professional Organization	39
Patients	46
Washington State Agency Comments	55

SPECTRUM RESEARCH RESPONSE TO PEER REVIEW COMMENTS

Note: Spectrum is an independent vendor contracted to produce evidence assessment reports for WA HTA program. For transparency, all comments received during the comments process are included. However, comments related to program decisions, process, or other matters not pertaining to the report are acknowledged through inclusion, but are not within the scope of response for report accuracy and completeness.

1. Michael J. Lee, M.D., Assistant Professor, University of Washington, Spine Service

Dr. Lee's comment 1 response: Methods section, cervical – Peng Fei et al and Nabhan et al were removed from the level of evidence summary table. See L&I Comment 23 response on page 13 below.

Dr. Lee's comment 2 response: Results section, Key question #3 – Added a comment that no studies were found evaluation L-ADR in workers compensation populations.

2. Sean D. Sullivan, RPh, PhD, Professor of Pharmacy and Health Services, Director of the Pharmaceutical Outcomes Research and Policy Program, University of Washington

Dr. Sullivan's comment 1 response: Superiority can be concluded from an inferiority study, but not the other way around. I added the following text on page 49 of the report. "A non-inferiority clinical trial design is often used in FDA trials to show that a new treatment is no worse than a reference treatment. In order to accomplish this, a pre-stated margin of non-inferiority is defined for the treatment effect of a primary outcome. The new treatment will be recommended if it is similar to or better than the existing one, but not if it is worse by more than the pre-stated margin. It is acceptable to assess whether the new treatment is superior to the reference treatment using the appropriate statistical test.[124,152,168]"

Dr. Sullivan's comment 2 response: The following text was added to the report: There were no reports of death relating to the device or surgical procedure with either ADR or fusion in either study.

Dr. Sullivan's comment 3 response: The short term complications (up to 2 years) are similar with L-ADR and fusion, and fewer with C-ADR compared with ACDF. The real effect on cost will be determined when longer term data are available.

Dr. Sullivan's comment 4 response: The rate of device failure up to 2 years is low with most of the failures reported at one and two years. We plotted the information and concluded that a figure would not be particularly helpful. However, we agree that time to

device failure and reoperation are valuable pieces of information, especially as longer term data become available.

Dr. Sullivan's comment 5 response: Appraisal was changed to Assessment.

Dr. Sullivan's comment 6 response: The following statement was added to the summary:

- One study suggests that surgeons and institutions with a high volume of L-ADR cases have shorter operating time and hospital stay, and lower complication rates which may have an economic effect. No effect on clinical outcomes was reported between high and low volume surgeons or institutions.

Dr. Sullivan's comment 7 response: We added information from three local payers.

Dr. Sullivan's comment 8 response: We accessed the literature references from the manufacturers of the devices undergoing FDA IDE clinical trials.

Dr. Sullivan's comment 10 response: Registry studies are considered observational studies and depending on the quality of the registry and the design of the study, would be evaluated as a cohort or case series. No large registry study was found for this report.

Dr. Sullivan's comment 11 response: Meta analysis is performed to estimate the size of the pooled association between treatment and outcome, to seek evidence that the association varies according to the level of some other factor, and to estimate a variance so that the precision of the pooled estimate may be determined using a confidence interval.[1] This can be done with two or more studies that are similar or homogeneous both clinically and statistically. See comment to Washington State L&I on page 12 below for our rationale for conducting a meta analysis.

[1]Cummings P. Meta-analysis based on standardized effects is unreliable. Arch Pediatr Adolesc Med; 158, 2004, 595-6

3. Ann M. Derleth, PhD, Health Services Research Postdoctoral Fellow, VA HSR&D, Seattle, Wa

Dr. Derleth's comment 1 response: Background – we made corrections throughout the report to reflect that one indication for surgery was failed conservative care for six months for the lumbar spine and six weeks for the cervical spine.

Dr. Derleth's comment 2 response: Methods – we used the FDA data when there was an unresolved conflict between the FDA reports and the published articles because the FDA data were often more completely reported. We added a text in the report to state this reasoning.

Dr. Derleth's comment 3 response: Results – The summary scores from the SF-36 physical and mental (PCS, MCS) were used and reported as such within the results section.

Dr. Derleth's comment 4 response: Results - Text added to clarify that ASD rates were among patients receiving L-ADR.

Dr. Derleth's comment 5 response: The footnote was corrected to point out that the risk difference in the Prodisc trial favored ADR with respect to major complications.

Dr. Derleth's comment 6 response: We added the MAUDE database into the methods section.

4. Jens R. Chapman, M.D., Professor; University of Washington, Director, Spine Service

Dr. Chapman's comment 1 response: The phrasing of the key questions comes from the Washington State HCA to Spectrum Research, the independent vendor.

Dr. Chapman's comment 2 response: We interpreted ADR as mechanical total disc arthroplasty and added a sentence to reflect this under the key questions listed in the executive summary.

Dr. Chapman's comment 3 response: We defined safety profile as complications, adverse events, device failure and reoperation. This is included in Key question 2.

Dr. Chapman's comment 4 response: Background – we added a phrase to emphasize that this frequently stated comment was anecdotal.

Dr. Chapman's comment 5 response: See response 2 above.

Dr. Chapman's comment 6 response: We added a short paragraph on the success of peripheral total joints as a motivation for spinal ADR.

Dr. Chapman's comment 7 response: We added some information about the Bristol disc, a precursor to the Prestige.

Dr. Chapman's comment 8 response: See response to Clinician's/Professional Organization Comment 4 Response on page 10 below.

Dr. Chapman's comment 9 response: The definition of the composite score is listed under section 2.5, Description of study outcomes, just preceding the results section.

Dr. Chapman's comment 10 response: We used the FDA recommended 15 point cut off, and we added that to the figures for ODI to help clarify this point.

5. Brian M. Drew, M.D., Assistant Clinical Professor, McMaster University, Medical Director, Hamilton General Hospitals Spine Unit

Dr. Drew's comment 1 response: We added a bullet in the summary to emphasize the issue of high volumes and its possible effect on outcomes/safety.

SPECTRUM RESEARCH RESPONSE TO PUBLIC COMMENTS

Responses to Industry Association Comments

DePuy Spine Comment 1 Response:
There are many systems available to evaluate the Level-of-Evidence in an Evidence Based Medicine environment. Spectrum Research has chosen a system adapted from the orthopedic surgery field and used by the *Journal of Bone and Joint Surgery*.[1] Its current system for articles pertaining to therapeutic intervention is reproduced below and can be accessed from the Journal's website, http://www2.ejbjs.org/misc/instrux.dtl.

Levels of Evidence for Primary Research Question[1]	
	Therapeutic Studies—Investigating the Results of Treatment
Level I	• High-quality randomized controlled trial with statistically significant difference or no statistically significant difference but narrow confidence intervals • Systematic review[2] of Level-I randomized controlled trials (and study results were homogeneous[3])
Level II	• Lesser-quality randomized controlled trial (e.g., <80% follow-up, no blinding, or improper randomization) • Prospective[4] comparative study[5] • Systematic review[2] of Level-II studies or Level-I studies with inconsistent results
Level III	• Case-control study[7] • Retrospective[6] comparative study[5] • Systematic review[2] of Level-III studies
Level IV	Case series[8]
Level V	Expert opinion

1. A complete assessment of the quality of individual studies requires critical appraisal of all aspects of the study design.
2. A combination of results from two or more prior studies.
3. Studies provided consistent results.
4. Study was started before the first patient enrolled.
5. Patients treated one way (e.g., with cemented hip arthroplasty) compared with patients treated another way (e.g., with cementless hip arthroplasty) at the same institution.
6. Study was started after the first patient enrolled.
7. Patients identified for the study on the basis of their outcome (e.g., failed total hip arthroplasty), called "cases," are compared with those who did not have the outcome (e.g., had a successful total hip arthroplasty), called "controls."
8. Patients treated one way with no comparison group of patients treated another way.

This chart was adapted from material published by the Centre for Evidence-Based Medicine, Oxford, UK. For more information, please see www.cebm.net.

This system is designed to distinguish between high- and lesser-quality randomized controlled trials. Of course, the hallmark feature of a properly conducted randomized controlled trial is that the random assignment of trial participants tends to minimize differences between study populations in factors that may influence outcome. In other words, it minimizes the effect of selection bias. As much as a randomized controlled trial is desired, it must be remembered that there are other places within a clinical trial where other forms of bias may enter. A potentially significant bias can result when the patient or the evaluator is not blinded to treatment. Blinding the patient is difficult for many surgical procedures, especially when compared with non-surgical care. Nevertheless, whether a study did not or could not blind the patient, the result is that bias is possible. In the current study, it is likely that many patients sought to be enrolled hoping they would receive ADR (a "newer" treatment). To the extent that was the case, those who were randomized to the ADR group would likely be more satisfied and report better outcomes than the fusion group. With respect to evaluator blinding, we expect any evaluation reported for the clinical study to be done with knowledge of the intervention when possible. When not possible, we expect the evaluator to be independent of the investigating team.

DePuy Spine Comment 2 Response:
Surgical intervention for lumbar DDD in these trials is offered to patients who continue to have symptoms after receiving at least six months of nonoperative care. We acknowledge that within this population, operative and nonoperative options may not be "competitive/interchangeable" in the sense that these patients are more likely to seek the surgical option and have greater expectation for improvement compared with continued nonoperative care. The SPORT study cited is a good example to illustrate this point (albeit in a different patient population) in that half of those randomized to nonoperative care for degenerative spondylolisthesis after at least 12 weeks of failed conservative care opted for surgery. However, after 2 years of follow-up, only 64% of those randomized to surgery underwent surgery. What happened to the 36% who didn't undergo surgery is not completely known; nonetheless, it is reasonable to assume that some improved without surgery. The optimum nonoperative care for lumbar DDD continues to be debated. What is needed is a better mechanism to identify which subgroups of patients that will positively respond to different treatment strategies.

Reference:
1. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am.* 2003;85-A(1):1-3.

\

Medtronic Comment 1 Response: Background - We added a section that better discusses the historical perspective of ACDF.

Medtronic Comment 2 Response: See DePuy Spine Comment 1 Response on page 6 above.

Medtronic Comment 3 Response: Summary Table 26 was edited to state that motion at the index segment for L-ADR is maintained or improved compared with preoperative levels.

Medtronic Comment 4 Response: Omission of studies – Anderson et al (2008), Sasso et al (August 2008), Riina et al (2008), and Yang et al (2008) were added to Pubmed after our search date but will be evaluated in future updates. The purposes of Sasso et al (Feb 2008) and Kim et al (2008) were to evaluate motion or sagittal balance primarily, not complications.

Medtronic Comment 5 Response: In addition to the national coverage plans, we added some state coverage policies from Washington State to include Premara Blue Cross, Regence, and Group Health Cooperative.

Medtronic Comment 6 Response: This report did not include data from presentations or abstracts. The two presentations listed will be reviewed when they are published in the peer-reviewed literature.

Medtronic Comment 7 Response: The references are organized in alphabetical order to facilitate citation identification.

Medtronic Comment 8 Response: We made corrections throughout the report to reflect that one indication for surgery was failed conservative care for six months for the lumbar spine and six weeks for the cervical spine.

Medtronic Comment 8 Response: Wear debris citation corrected.

Medtronic Comment 9 Response: The longitudinal study citation was separated from the case series so that it reads clearer.

Medtronic Comment 10 Response: Changed the sentence to reflect that many technology assessments were performed prior to any RCTs.

Medtronic Comment 11 Response: Reference to Tables 4 and 6 were added which contain the nine HTAs.

Medtronic Comment 12 Response: We did not detail the identification of the case series based on the report's inclusion/exclusion criteria found on Table 7 of the report.

Medtronic Comment 13 Response: The citation was corrected.

Medtronic Comment 14 Response: Non-inferiority studies can be evaluated for superiority. See L&I Comment 32 Response on page 14 below for more discussion on superiority in non-inferiority studies.

Medtronic Comment 15 Response: Statistical significance is not important in comparing baseline differences since P-values depend on sample size, variance and effect size. The effect size (the magnitude of the difference between the two groups) is what is important. Remember, the P-value is the probability that the difference is due to chance. In a randomized controlled trial, this makes no sense since the probability that any difference is from chance is 100% (given that random allocation was done correctly).

Medtronic Comment 16 Response: See Clinician's/Professional Organization Comment 4 Response on page 10 below.

Medtronic Comment 17 Response: Though VAS pain was not the primary outcome for this study, Nabhan et al reported that there was no statistical difference between groups for this outcome. Our comment is meant to reflect that for this outcome, this no statistical difference may be a result of a small sample.

Medtronic Comment 18 Response: We made the following corrections: % was changed to points, and the changes in the SF-36 scores were rectified.

Medtronic Comment 19 Response: Goffin et al (reference #124) reports one evacuation of a paravertebral hematoma.

Responses to Clinician's/Professional Comments

Clinician's/Professional Organization Comment 1 Response:
With respect to the level-of-evidence rating for surgical trials, please see the discussion on page 5 above, *Depuy Comment 1 response.*

Clinician's/Professional Organization Comment 2 Response:
We acknowledge that lumbar and cervical ADR are indicated for different spinal conditions; one essentially treats the symptoms of pain thought to arise from the degenerative disc (lumbar) while the other treats signs associated with neurological compromise (cervical). The report attempted to make this clear in all major sections to include the results and summary sections. Spectrum is an independent vendor, and the decision to include lumbar and cervical in one report belongs to the Washington State Health Care Authority (HCA).

Clinician's/Professional Organization Comment 3 Response:
The Washington State HCA asked us to see if there was evidence available to compare ADR with nonoperative therapy. We concluded that there was not.

Clinician's/Professional Organization Comment 4 Response:
The comment was meant to point out that the two populations being compared had some potentially important differences at baseline. With a big enough sample, we expect that these will even themselves out from random assignment. However, they don't always do so and therefore, studies should list robust baseline characteristics so that the reader can see if potentially important differences occur. When they do, we believe they should be adjusted for in the analysis or at least evaluated to see if they are potential confounders. One mistake that authors often make is to compare the differences in characteristics between groups using a statistical test, and concluding if the P-value is not statistically significant, then the difference is not big enough to be important (see CONSORT statement)[1]. An example is the smoking proportions in the Prodisc study. The effect size (11% difference in the proportion of patients who smoked between the two treatment groups) is relatively large even though the P-value is "non significant". As to whether the smoking could have confounded the results due to the unequal distribution of smokers, the point made by the clinician's comment is well taken with respect to its effect on fusion. Since it had no apparent effect on the fusion rates, we removed this discussion point.

With respect to the number of patients who were "enrolled" versus "treated", the issue for analyses is how many patients received random assignment. From the Prodisc study using the published journal article or the SSED, we were unable to determine for certain if those who were enrolled were randomized even if they did not receive treatment. The fact that there were 21 of these in the ADR group and 13 in the fusion group implies that they were randomly assigned and therefore should count in the follow-up even though they did not receive treatment. This type of analysis process helps to ensure the integrity of the random allocation process.[1]

[1]Altman DG, Schulz KF, Moher D, et al. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Ann Intern Med.* 2001;134:663-694. See especially page 677.

Clinician's/Professional Organization Comment 5 Response:
The comparison we were looking for was between ADR and ***continued*** nonoperative care in cohort or RCT study design. Therefore, the sentence was changed to state that there were no studies found comparing lumbar ADR with continued nonoperative care.

Clinician's/Professional Organization Comment 6 Response:
Spectrum Research is an independent vendor, and as such received no such mandate from the Washington State HCA. It is curious that the readers perceived that the analyses were structured in a way to emphasize the negative aspects and to downplay the positive aspects when the bullet points on efficacy/effectiveness state:

- that there is moderate evidence that the efficacy/effectiveness of L-ADR as measured by the composite measure of overall clinical success, Oswestry Disability Index (ODI) improvement, pain improvement, neurological success, SF-36 improvement, and patient satisfaction is comparable with anterior lumbar interbody fusion or circumferential fusion up to two years following surgery
- There is moderate evidence for the cervical spine that C-ADR is superior to ACDF with respect to overall clinical success (77% versus 68%) and neurological success (92% versus 86%), and is comparable with ACDF with respect to Neck Disability Index, and pain up to two years following surgery.

Spectrum Research Response To Washington State L&I Comments

L&I Comment 1 response. This statement is not meant to convey that the fusion is the standard of care; rather, that fusion is the SURGICAL standard. That is, when surgery has been decided, fusion is the current surgery of choice. We ensured that this statement and all others similar have the words "surgical standard" included.

L&I Comment 2 response. This statement is not meant to imply that this has been proven. It simply meant to state that this is one of the aims of ADR.

L&I Comments 3&4 response. We acknowledge that there is debate on which studies should be pooled for a meta analysis. In general, variation across studies (heterogeneity) should be considered in two main areas: clinical and statistical heterogeneity. Clinical heterogeneity has a subjective component that should take into account the similarity of the patient populations, the treatments and the outcomes among studies. Though we recognize that there are some differences between the studies for this technology assessment (the case for all meta analyses), we judged that there was sufficient clinical homogeneity to pool. Consider the following: With respect to the patients in the two lumbar FDA trials, their demographics (age, gender, race, BMI, prior spine surgeries) were similar. With respect to ADR treatment, one ADR was semiconstrained, one unconstrained. At this point, there are no data to suggest outcomes from one are different than another. One control group received ALIF, one circumferential fusion. 91% of the ALIF patients fused compared with 97% of the circumferential fusion patients. A 2005 Systematic Review[1] was unable to draw conclusions about the relative effectiveness of anterior, posterior, or circumferential fusion due to lack of evidence. With respect to outcomes, both lumbar studies relied on the ODI by itself as a functional outcome and as the core to a composite score of "overall clinical success". And though the sponsors had different cutoffs for minimal clinically important differences (MCID) in ODI improvement, each provided data for the FDA for the 15 point MCID cut point which we were able to use. In that regard, the outcomes were homogeneous. It is noted that the Prodisc study had the addition of any improvement in the SF-36 score and radiological success in their composite score for clinical success compared with the Charite study. However, we agree with The Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care, who state, "It was thought that the addition of these 2 variables to the composite definition of clinical success would make it harder to achieve clinical success and therefore not bias the result in favour of clinical success. Because of this, synthesizing the data from these slightly different definitions was thought to be acceptable." (MSAC HTA, page 40). With respect to the statistical heterogeneity, we did not pool when heterogeneity was present.

[1]Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis. Cochrane Database Syst Rev. 2005 Oct 19;(4):CD001352. Review.

L&I Comment 5&6 response. We agree with these comments. We also believe that our statement that the connection between motion and ASD is unclear, and the connection between ASD and patient symptoms is not established articulates this point.

L&I Comment 7 response. See response 3 & 4 above.

L&I Comment 8 response. We removed the section in question.

L&I Comment 9, 10 response. See response 1 above

L&I Comment 11-14 response. The information in this section is for context purposes primarily and includes a wide range of estimates based on marketing data. We added some text to emphasize that the potential impact of these devices on cost of medical care is dependent on the extent that certain predictions are correct. We omitted a press release and added a statement on post approval studies.

L&I Comment 15 response. Additional text was added in the background.

L&I Comment 16 response. We added a comment about the technical demands of L-ADR vs. fusion. It should be noted that some surgeons believe that the learning curve for ADR may be the same as fusion, and is probably a function of surgeons being more comfortable with fusion surgery due to the long history of the procedure.

L&I Comment 19 response. Both are expected to last for 40-50 years.

L&I Comment 20 response. We included a comment to reflect this.

L&I Comment 21 response. The FDA statistical review noted that there was no *a priori* statistical plan initially submitted. The peer reviewed Blumenthal article states, "The sample size was computed using the Blackwelder methodology, 13 assuming that 70% of the patients in both the investigational and control groups would have a successful result and that a clinically insignificant difference in success rates between groups (delta) was 15%. Choosing a type I error of 5% (one-sided) and 80% power, the sample size in the investigational group was 174 patients, and the sample size in the control group was 87 patients, for a total of 261. Allowing for a potential dropout rate of 10% resulted in approximately 194 patients in the treatment group and 97 patients in the control group, for a total of 291 patients."

L&I Comment 22 response. The data presented were interim data on the Bryan FDA panel summary.

L&I Comment 23 response. We added a sentence that there was not enough information in the methods section of the Peng-Fei or the Nabhan articles to warrant a level of evidence rating.

L&I Comment 24 response. The current trend when fusing for back pain associated with lumbar DDD is to do a 360 degree or circumferential procedure which was done by Zigler et al. Blumenthal et al appeared to do a stand alone ALIF. See response 3 & 4 for the rationale for pooling the data from these two studies.

L&I Comment 25 response. We included an intent to treat and sensitivity analysis on the pooled data providing information on the various scenarios for imputing missing or discontinued data.

L&I Comment 26 response. See response 3 & 4.

L&I Comment 27 response. Yes, and text is added to reflect that.

L&I Comment 28 response. Putzier et al values were included and text added to highlight the heterotopic ossification/spontaneous fusion rates in two studies with 10 plus years of follow-up, and a possible explanation as to the difference in rates.

L&I Comment 29 response. Noted. We attempted to summarize adverse events and listed each from the FDA trials in the appendix.

L&I Comment 30 response. Mehren et al was not initially included in the table because their rates were based on the number of spinal segments, not the number of people in the denominator. We have subsequently added this study to the table and text

L&I Comment 31 response. The figure may be helpful for some.

L&I Comment 32 response. Superiority can be concluded from an inferiority study, but not the other way around. I added the following text on page 49 of the report. "A non-inferiority clinical trial design is often used in FDA trials to show that a new treatment is no worse than a reference treatment. In order to accomplish this, a pre-stated margin of non-inferiority is defined for the treatment effect of a primary outcome. The new treatment will be recommended if it is similar to or better than the existing one, but not if it is worse by more than the pre-stated margin. It is acceptable to assess whether the new treatment is superior to the reference treatment using the appropriate statistical test.[124,152,168]"

L&I Comment 33 response. Since we don't have all the data available, we omitted this sentence.

L&I Comment 34 response. Longer term results from case series still report a wide variance in the number of failures and spontaneous fusions. In the two lumbar studies with over 10 years of follow-up (Putzier and David), the results were very different (60% HO/fusion vs. 3%). One difference between the two studies is the postoperative motion. They both had their patients initially immobile for several weeks following surgery. David changed his postoperative follow-up after he noted some HO forming, and allowed the rest of his patients early mobilization. He noted a significant reduction in HO in those later patients. We need longer term data from the RCTs and we need prospective safety data from case series.

L&I Comment 35 response. See response 3 & 4.

PEER REVIEW COMMENTS

1. Michael J. Lee, M.D., Assistant Professor, University of Washington, Spine Service

INTRODUCTION Comments

Lumbar: The overview is well defined. Specific questions and goals of the paper are well defined and exhaustively researched. The topic is important to address and the public policy and clinical relevance are well delineated in the introduction. The introduction/executive summary provides a concise overview of the paper.

Cervical: Overview of topic is adequate. The introduction adequately describes the clinical scenario relevant to cervical artificial disc replacement. Because the technology is newer than lumbar disc replacement, the report also adequately contrasts the indications for lumbar and cervical disc replacement. In addition, a nice historical background is provided leading up to the advent of ACDF. The incidence of adjacent segment disease is well described on page 25 of 224. It should be noted that while adjacent segment degeneration is widely discussed and supported by biomechanical and clinical studies, some surgeons feel that "adjacent segment degeneration" may only be a progression of the "natural history of cervical spondylosis" and "probably unaffected by the operative management" (Hilibrand et al JBJS 1999). These opinions originate from the original study that is oft quoted for its 2.9% rate of adjacent segment degeneration.

BACKGROUND Comments

Lumbar: The literature review and background are sufficient. On page 16, line 11, prior to stating "in 2001 122,469 lumbar fusion surgeries were performed…", I believe there should be statement as to what the cause of pain is. There seems to be a disconnect between lower back pain, and then fusion. I would recommend a statement to the effect that the degenerated disc is believed to be the pain generator, and traditionally, fusion has been used to eliminate motion at the pain generator site and subsequently the patient's pain. Then I would continue with the fusion statistics. I believe this allows the reader to make the connection of why fusion is being used to treat DDD.

Cervical: Literature review and background are sufficient. Biomechanical studies (Eck et al Spine 2002) would further support increase stresses at segments adjacent to a fusion. Adjacent segment degeneration is well described, and the authors do a good job of differentiating radiographic adjacent segment degeneration, clinically symptomatic ASD, and clinically symptomatic ASD requiring surgery.

REPORT OBJECTIVES & KEY QUESTIONS Comments

Lumbar: If the formal report begins on page 16, then I do not see the report objectives in this report. They are clearly defined in the Appraisal section of the report (pg 10-15). It seems it may allow for more linear thought process to restate the objectives and questions on page 36 prior to introducing Section 2 the evidence. Otherwise, the objectives and key questions are clearly defined in the Appraisal and Executive Summary.

Cervical: The objectives and key questions are well delineated in the Summary, however it may be nice to revisit them prior to addressing the methods section, so the reader may follow what questions are being answered while assessing the literature.

In regard to key Question #1, I believe it is inappropriate to compare C-ADR to non-operative treatment. The most valid comparison would be to operative treatment, the ACDF, which is the gold standard with a well-documented history of success.

METHODS Comments
Lumbar: The method for identifying relevant studies is well defined on page 39.
The method for selecting appropriate studies is adequate. Exclusion and Inclusion criteria are well defined. Level of Evidence rating is appropriate. It should be noted that no study evaluated in the report held a Level I rating. It should be further noted that blinded assessment is not possible (from a reviewer) and not ethical (from a patient). Therefore, no study examining lumbar ADR can qualify as a Level 1 study using these criteria.

Cervical: This was a little confusing. The 3 FDA studies are well described. Initially it was not clear to me which studies were the FDA studies. After reviewing the Pang Fei and Nabhan summary, I initially was confused why certain studies were excluded (Sasso et al Dec 2007 Spine). For me, the summary of the Pang Fei and Nabhan studies added confusion. While certainly important to note, the meat of the analysis really lies in the FDA study comparisons (which is well done). It may be less confusing to address the other studies afterwards the FDA comparison.

RESULTS Comments
Lumbar: The detail in the results section is exhaustive and appropriate.
The key questions are answered appropriately. Key question #1 is well answered in detail. Regarding "patient satisfaction", I would note that pre-operative impressions of L-ADR vs fusion are important data not reported. Anecdotally, many patients sought to be enrolled in these studies because they wished for a L-ADR. Everything else being relatively equal, the ones randomized to L-ADR would likely be "more satisfied" than those with fusion because these patients (anecdotally) were seeking ADR.

Key question #2 is well answered.

Key question #3 reports the available data examining the question. There are limited reports at this time looking at "special populations." This will likely be investigated in future studies. Of note, the Key Question #3 mentions workers compensation populations, but does not address them in the text.
Key question #4 is addressed as best can be by the available literature. The authors provide a detailed evaluation of available reports examining cost effectiveness.

Cervical: The results were easy to follow. The charts were easy to follow. The analysis was appropriately done with and without the Bryan data for completeness. The Key questions are answered as best can be in this early stage of analysis. As stated in the text,

C-ADR is newer than L-ADR and studies with longer follow-up are required to fully investigate the safety issues and the incidence of adjacent segment disease.

CONCLUSIONS Comments
Lumbar: The conclusions essentially state that at this time, the current literature suggests that L-ADR appears to be comparable to lumbar fusion in regards to clinical improvement and safety and efficacy. As the authors clearly state, long term data are still required to better assess the incidence of adjacent segment disease. In addition, the authors appropriately point out that different lumbar disc replacement designs and fusion strategies may affect future comparisons.

Cervical: The conclusions are valid. As stated in the conclusion, studies with longer follow-up are required to further investigate safety and sequelae of these procedures. At this time, the clinical improvement appears to be comparable between ACDF and C-ADR, however longer follow-up is required.

OVERALL PRESENTATION and RELEVANCY Comments
This review is very well structured and organized. The main points are very clearly presented. The executive summary does an outstanding job of summing up the major points. The depth of reporting data in the text can be challenging to follow, however does accurately report the current literature. As stated in the background, DDD is a major source of disability and is quite relevant to clinical medicine and public policy and health.

2. Sean D. Sullivan, RPh, PhD, Professor of Pharmacy and Health Services, Director of the Pharmaceutical Outcomes Research and Policy Program, University of Washington

(1) Executive Summary. You indicate there is moderate evidence that C-ADR is superior to ACDF? You then cite 2 non-inferiority studies. Were the trials powered for superiority, even though non-inferiority was the main design feature? Can you really say the evidence suggests superiority?

(2) Are surgical mortality rates in patients undergoing the ADR procedures versus non-ADR surgical procedures comparable? It seems that these data would be available somewhere, even if they did not come from a clinical trial.

(3) I note that the economic data and previously conducted HTA report suggest that complications rates (and therefore costs) may be higher in ADR? Should this be reflected more prominently in the risk section of your report?

(4) One of the main economic drivers for the cost-effectiveness of ADR is device failure and re-operation. It would be interesting to see a chart with time on the x-axis and failure rate on the y-axis so that decision-makers can visualize the failure rates for the ADR technology and the non-ADR comparators – even if you only have 24 months of data

from the published reports.

(5) On page 10, you use the term Appraisal. Are you required to use this term by Washington state? If not, you might consider changing this to Assessment. You will note that the UK NHS process defines Assessment as the systematic evaluation of the evidence (what you are doing) and Appraisal as the process that the decision makers use to review the assessment and make a recommendation to the NHS. The HTAi and EUnetHTA organizations make the same distinction.

(6) It would seem to me that one of the findings to highlight in the executive summary is that higher surgical volume is associated with better outcome in lumbar procedures. This is important, because if Washington state issue a positive coverage determination, they may decide to make coverage conditional upon use of a high-volume surgeon.

(7) In table 5 and 6 you describe payer policies for the ADRs. The policies would be more useful to the Washington state HTA program if you included local payers like GHC, Regence and Premera.

(8) Did you query the manufacturers for studies? Did you submit the list of studies to the manufacturers and ask them if you may have missed any recent reports?

(9) I like that you used the QHES.

(10) From Table 9, it is not clear how you would rate a high quality registry study? Would you place registries alongside a cohort study? In any event, were there large registries available for ADRs?

(11) You know I am not an expert on meta-analysis. However, I question the need to perform a meta-analysis on 2 or 3 studies, unless the study designs, research questions and treatments were exactly the same. Can you assure a reader that this is the case?

(12) You did a nice job with the economic section.

Ann M. Derleth, PhD, Health Services Research Postdoctoral Fellow, VA HSR&D, Seattle, Wa

Introduction:
Overview of topic is adequate and important to address. With new technology it is important to assess whether it provides an improvement on the current standard of care.

Page 10 - lines 7-13 Important to be clear that increased incidence of procedures performed is not necessarily increase in underlying condition it is seeking to correct. It can be either increase in incidence of the condition or change in surgical practice where the procedure is used more frequently or a combination of the two.

Background:
The literature review is thorough and sufficient.
Page 23 Line 30: clarify whether this is six weeks or six months.

Report objectives and key questions:
The aims clearly address relevant policy and clinical issues. The key questions are clearly defined and adequate to achieve the aims.

Methods:
The search methods for identifying the relevant studies is thorough and well presented, including specification of inclusion and exclusion criteria. The Level of Evidence (LoE) clearly explained and provides an excellent way to characterize the rigorous standards used to evaluate the reports reviewed. Methods of data abstraction and analysis are very good.
Page 41 Line 5: Explain why the FDA data were used vs those in peer reviewed reports when there was a conflict - are the FDA standards more restrictive?

Page 43 First paragraph: QHES is well presented and evaluated.

Results: the results section is very well presented in terms of organization, level of detail and clarity of tables and figures. Limitations are well stated.
Page 54 - State which score of the SF-36 is used - whether it is one or more scale scores (there are 8) or a summary score (PCS or MCS). This matters because it is reported as >15 point difference and that is easier to achieve on a scale score than on a summary score.

Page 65: paragraph on ASD, line 2 - it is not stated whether the patients are fusion, ADR or both.

Page 79: Table 20 footnote: risk difference in Bryan trial is reported to be in favor of ADR but the CI is (-.06-0.01)

Page 80, bottom of page: I was pleased to see mention that the MAUDE database was searched. But I didn't see this mentioned in the methods section - suggest it should be.

Conclusions: The conclusions reached are valid and well stated.

Overall Presentation and Relevancy
This is a well organized and thorough review of the literature and available information on the current state of knowledge for lumbar and cervical artificial disk replacement technology. It is very relevant to clinical medicine and important for public policy and public health. It is appealing to use new technology when it appears it might lead to improved patient outcomes, but often can be implemented before long terms results are known. This kind of assessment provides an objective guide to policy makers and clinicians for their decision making.

4. Jens R. Chapman, M.D., Professor; University of Washington, Director, Spine Service

INTRODUCTION Comments

Lumbar: This report fundamentally suffers from lack of scientifically sound " four key questions" as basis for its analysis, which clearly hampered the attempts of this research group to try to answer the questions posed. Each of these "key questions" by themselves reflects lack of familiarity with the subject matter or poor scientific background by those who were asking them and introduces a potential for considerable bias introduced by their phrasing. If it is the goal of the HCA to obtain a fair and unbiased review of a current or emerging health technology these 4 key questions do not provide the basis for such an analysis. I am afraid that a great opportunity has been squandered and Washington state tax-payer dollars have been wasted by not asking questions which may actually benefit interested or affected citizens of the State of Washington to gain insight into emerging health care technologies such as disc replacement technology. Pertinent and highly interesting questions were either not posed or addressed in a roundabout fashion.

Aspects of the limitations of this analysis lie in the phrasings of its key questions 1-4:
Key question 1
What is the evidence of efficacy and effectiveness of ADR compared with comparative therapies (including nonoperative therapy, spinal fusion, other surgery)?

The term lumbar disc arthroplasty is not defined anywhere in this text. There are many variants of mechanical total disc arthroplasties, which are listed in some detail (page 18, paragraph 2). What about nucleus replacements, anular reconstruction techniques and other forms of intradiscal spacers, which are all variants of lumbar disc arthroplasty? What about disc transplants, disc regeneration techniques? The wide area of disc replacement surgery has not been defined from the onset. It seems that this analysis is concentrated on the assumption that lumbar artificial disc replacement is synonymous with mechanical total disc arthroplasties, a hypothetical premise which should have been defined from the onset.

The premise of comparison of disc arthroplasty to nonoperative care is flawed from the start as all the US trials mandate a failure of all supervised nonoperative management as premise for exclusion. To assume parity of fusion results to nonoperative care based on some European PRCT study populations is a highly problematic assumption based on highly divergent study populations between the US and the European cohorts at hand.

Key question 2
What is the evidence related to the ADR safety profile (including device failure, reoperation)?

What is a ' safety profile', how is it defined? Again, the use of a non-defined term does not provide a basis for a scientific analysis.

Key Question 3

What is the evidence of differential efficacy or safety issues amongst special populations (including but not limited to the elderly and workers compensation populations)?

Convoluted and conditional question writing does not provide a sound basis for any exploration, especially not for a complex subject matter, such as the one at hand. What, please are 'special populations', do they bear any similarities with contestants at 'Special Olympics'? What is an elderly population? I am still waiting for general comparative workers compensation population studies comparing 'differential efficacy' of workers compensation systems of various states in the U.S. and other countries compared to those in the State of Washington.

Key Question 3

What are the cost implications and cost effectiveness for L-ADR?

Yet again, definitions, please. What are cost implications? I have never heard this term used in any analysis before. Did the question writers wish to compare disc arthroplasty patients to any specific other cohort?

Cervical: The study at hand suffers from poorly phrased key questions, which serve as determinants for the project at hand. Each of the key questions fails in providing defined terms for its questions and posing answerable questions. Since the key questions posed by the HCA of Washington were left unchanged the same criticisms applied to the lumbar disc assessment apply here as well and will not be reiterated. The fundamental differences in human anatomy, biomechanics, clinical indications and expected long term outcomes that present as differences between lumbar and cervical disc pathology were not at all reflected in these questions. I am not sure how the citizens of the state of Washington were helped with this project in the context of these undifferentiated questions.

BACKGROUND Comments
Lumbar: Page 17, paragraph 2, line 4. No reference for Fernstrom failure rates given, stated argument of 'After a short period of symptom relief, the prosthesis ultimately failed secondary to subsidence of the implant within the spine vertebra leading

to abandonment of the technique' is hearsay and should be eliminated unless details quoted. Long-term data on Fernstrom cages suggests differently.

Page 117-18. All explanations for disc arthroplasties pertain to mechanical total disc replacements and do not address 'disc arthroplasty' as term posed in question.

Page 26. The references to other health technology reviews having taken place was illustrative and helpful.

I would expect a reference to other forms of arthroplasties in this part, such as hip and knee replacements, which are considered some of the most successful heath related quality of life procedures known in medicine. We derive much of our knowledge and concerns for spinal mechanical arthroplasties from the 50 year history of utilizing these devices in major extremity joints.

Cervical: Page 13, paragraph 1, line 5. No mention of electrodiagnostics as supplement to physical examination is given. Page 20, C-ADR and onward. No mention of the Bristol disc and its lengthy track record is made in this paragraph. The Bristol disc is relevant as the minimally modified precursor to the Prestige disc. There is some 20 year data available for this model of disc replacements.

REPORT OBJECTIVES & KEY QUESTIONS Comments

Lumbar: See above comments regarding key questions.
I remain confused as to the objectives. Were these objectives formulated by the HCA or by the research organization? How are the objectives and key questions supposed to interface? Is the objective of this undertaking to formulate health care policy, advance medical knowledge, improve informed decision making of affected Washington State patients or produce a summary statement on the state of research of clinical studies pertaining to lumbar disc arthroplasties? The purposes of this undertaking are not articulated and spelled out, which adversely affects its relevance.

Cervical: As with the L-ADR I am missing a clear objectives statement in conjunction with the key questions. The flaws of the key questions have been outlined above and in my comments on L-ADR and are not repeated herein, but remain in full effect.

METHODS Comments

Lumbar: A thorough and comprehensive attempt at compiling articles pertaining to total mechanical disc arthroplasties was made. In- and exclusion criteria for review of studies were reasonable and meet scientific and fairness and relevancy standards.
Level of evidence determination deserves further commentary. The Oxford Centre for Evidence-based Medicine, precepts outlined by the Grades of Recommendation Assessment, Development and Evaluation (GRADE) Working Group[4] and recommendations made by the Agency for Healthcare Research and Quality (AHRQ) are widely accepted, no doubt. These standards were developed with pharmacologic trials in mind and are incompatible with current surgical practices and the realities of performing research in most countries of the Western civilized world. The downgrading of the 2

main PRCT's (Blumenthal and Zigler) to moderate or poor quality RCT's is harsh and as stand-alone statement can induce a bias against the undeniable and unprecedented quality of either of these trials. It would improve the fairness of this presentation to provide a more balanced and detailed discussion of why the downgrading to IIb) occurred. In this context the realities of FDA study regulations and time periods to insurance approval for some elective surgeries may explain differences in enrollment numbers and treated numbers. These are factors clearly are outside of the control of any US investigator and usually will not introduce a methodological bias, thus should not be used as a reason for downgrading. The same goes for completers versus ITT population. The data acceptable to the FDA has to be obtained in specified time sections, data outside of that specified window is not of interest to the FDA, and will thus usually not be published.
A downgrading of LoE for differences in smoking status appears not appropriate since the data analyses in these 2 study populations don't support different complication rates for fusion patients. Extrapolation from previous publications in this regard doesn't meet scientific standards if the data presented specifically contradicts this variable having a confounding influence.

Cervical: Assignment of Level of Evidence status to category II: The reviewers have pointed out the incompatibility of the GRADE system with surgical trials under FDA premises (which, for instance, explicitly governs timing of device disclosures to patients). Page 67. Please define 'success' as a composite 'all-or-nothing' term introduced by the FDA and list its subcomponents. This definition of success has never been validated statistically and should be used with some caution.

RESULTS Comments
Lumbar: Page 56. The issue of "success" of disc arthroplasty is discussed prominently in the opening paragraphs of the result section in reference to the 2 'index studies' by Blumenthal and Zigler. Nowhere is it made clear that the definitions of success are "all-or nothing " composite scores created by the FDA without scientific validation based on a number of outcomes scores, radiographic observations and absence of complications, which were different for the 2 trials. It would seem reasonable to point these circumstances out and cut and paste these FDA derived definitions into the text for a clearer understanding of the definition of 'Success' for the benefit of the readership at large.

Page 59: does the ODI data presented, for instance in Figure 8, assume a > 15 point difference throughout, or are there different standards applied by either study?

Cervical: The available results are presented in a fair and clear fashion.

CONCLUSIONS Comments
Lumbar: The evidence tables provided reach valid conclusions within the diffuse parameters set by the study sponsor.

Cervical: A fair representation of the data available to date has been made

OVERALL PRESENTATION and RELEVANCY Comments
Spectrum has done a fair and reasonable job with the foundations it was provided trying to answer the questions it was given. The main issues concerning disc arthroplasty unfortunately have not been adequately addressed. These include identifying patients with especially good results and differentiating these from those with poor results. What are failure mechanisms of disc arthroplasty – are there any predictors for poor outcomes based on indications, surgical technique and postoperative care? What are the biomechanical and clinical foundations for development of - and prevention of adjacent disc degeneration? Are there differences in efficacy and effectiveness, and if so what factors play a role? How did complications affect HrQoL outcomes? Did maintenance of motion lead to better or worse outcomes? Does quality of surgery with disc replacements influence outcomes and complication rates?

5. Brian M. Drew, M.D., Assistant Clinical Professor, McMaster University, Medical Director, Hamilton General Hospitals Spine Unit

INTRODUCTION Comments
I thought the introduction was accurate in terms of the scope and overview of DDD and the general surgical indications for DDD, specifically when ADR is an option.

In the summary and implications section (page 7, last paragraph) I thought the 6 week period of conservative treatment is too short and should be 3 months. I did not feel 6 weeks is wrong but just a bit on the aggressive side unless there was progression of neurological signs. Otherwise this section was clinically relevant, particularly with respect to a) special subpopulations and b) the last paragraph on page 9 regarding the different biomechanical designs

Page 10, 3rd paragraph—very important to highlight the fact that adjacent segment disease is a controversial issue and not well understood. This is addressed here well and at several other points in this document.

Page 12-15 - The primary and secondary outcomes and complications are clinically relevant measures and issues.

The comment at page 13 regarding new indications and off-label use is clinically very relevant. There is a long history of new products coming to market with strict indications that then widen overtime creating new complications and issues not previously known.

On page 15, the last paragraph is important as it highlights the issue of high-volumes. Most of the studies referenced would be performed by surgeons with significant experience in this procedure. The complications would be expected to increase with lower volume surgeons performing the procedure. The extent of the increase is difficult to quantify.

BACKGROUND Comments
Page 16. 1.1-- The epidemiology is accurate. The anatomy and pathophysiology is basic but gives an accurate and sufficient basis for the understanding of the rationale for the use of an ADR in DDD.

Page 17-25. 1.2-- A good description of the history of L-ADR is given as well as the various types of products are well described. The biomechanical principles, classifications and material components are accurate. The clinical symptoms and unique design differences in C-ADR are highlighted and are accurate. The surgical description and potential complications of C-ADR are accurate.

The indications and contraindications are sufficient.

Page 19 paragraph 3 highlights the importance of the lack of long term data to help clinicians understand what possible adverse outcomes may lay ahead. Surgeons currently lack a good understanding of how they may need to deal with implant failures over the long term.

The summary of the indications and contraindications of L-ADR are reasonable

Page 23-25. The alternative non-operative and operative treatment options to L&C-ADR are described sufficiently and accurately.

Section 1.4 summarizes all relevant technology assessments in a comprehensive and well organized fashion. This includes both lumbar and cervical ADRs. The tables are particularly helpful (tables 3 & 4)

Finally in section 1.6 it reveals a comprehensive and very current review of the latest evidence despite it being incomplete. Caution is required in interpreting this data until full enrollment and follow-up is achieved. But it is important to note this clinically important work is currently being undertaken.

REPORT OBJECTIVES & KEY QUESTIONS Comments
I believe this was done well. Certainly the aims and objectives were clearly addressed. They represent relevant policy and the important current clinical issues.

The key question were clearly described and thoroughly addressed and discussed throughout the document. Limitations in what the literature had to offer were well described as it pertained to the key questions.

METHODS Comments
Page 36-38. Table 7 summarizes the inclusion and exclusion criteria well. The population, intervention, study design and outcomes are appropriate.

Page 38-50. The search strategy & the algorithm for article selection are described well and are appropriate. Figures 2 & 3 help explain the rationale for selection and the relationship to the key questions in an organized fashion.

In reviewing the document I believe the relevant studies were reviewed or at least considered based on the studies methodological merits and its relationship to answer the key questions. The studies selected from the current literature were of high quality. Tables 10 through 13 help to summarize this well.

The definitions used to differentiate the levels of evidence for articles on therapy fit standard definitions and are quite appropriate. This use of these definitions was instrumental in helping to rate the LoE. The use of meta-analysis to interpret primary outcomes is well described and then a good flow sheet is available in Appendix D.

The Quality of the studies used to evaluate both the lumbar and cervical ADR are well described on pages 45 to 51. I felt this was quite comprehensive.

Page 51-55. The description of the study populations and study outcomes for both the lumbar and cervical are summarized well and are more than adequate.

RESULTS Comments

My comments in this section do not include page numbers as my comments are more general in nature and are aimed helping to answer the questions above.

The detail in the section is very comprehensive. It includes all clinically relevant outcome measures needed to determine clinical success and evaluate the safety of the various ADRs. The figures and tables were quite helpful in summarizing the large quantity of date and they highlighted the major findings well.

The key questions were answered thoroughly. The strongest available data was sought out and I believe interpreted appropriately to assist in answering these four questions well. The data regarding possible adverse outcomes was summarized very well and this is critical when evaluating new technologies.

The limitation of the literature was highlighted well with respect to adverse events and complications. The document recognizes and highlights the gap present here. The document acknowledges these devices are designed for long term implantation yet the available studies lack the long term follow-up necessary to objectively analyze this.

The implications of the major findings were described well. Perhaps the clinical significance or nature of some of the complications (for example DVT or vessel laceration) could have been described in a little more detail to give a better general understanding of them and their significance. This may help in the understanding that these complication could in some circumstances be life threatening. A lay person reading

this document may not understand the significance of some of the complications. However the breadths of complications were well covered.

Key Question 3 is important but details in the literature are sparse. This explains the lack of clinical data to review on this area. This is a significant gap in the literature that is highlighted well in this document and is clinically important. The populations listed here (smokers, athletes and the elderly) are not insignificant in society and they will experience DDD of the lumbar and cervical spine. There will be no doubt that they will request these forms of treatment from surgeons as these devices come to market. There will be pressure on surgeons to use these treatment modalities but the lack of data on this population exists and hence the importance of this question. Despite the sparse data here it is the best information available to date.

I do not have any experience or much knowledge regarding the cost implications and economic analyses and therefore can not comment on the details in Key Question 4. However it is clearly an important issue and it is extensively addressed in this document.

CONCLUSIONS Comments

The Summary and Implications section summarizes the clinically important issues succinctly. The conclusions reached are an accurate interpretation of the current literature. They reflect the important clinical issues and address the lack of data on other clinically important issues.

I believe the conclusions were reached in a valid manner. The methods section was well described and appropriate search strategies were used to identify and eliminate appropriate evidence.

OVERALL PRESENTATION and RELEVANCY Comments

The review was certainly well organized and easy to follow. The order in which the information was described helped achieve this goal. The use of tables, charts and figures was helpful to clarify and summarize the results.

The main points were clearly presented. The written descriptions were easy to follow and the again the use or tables, charts and figures helped with this. The main points were often repeated to various degrees in multiple sections including the conclusion to help with clarity and importance.

The details of this document cover all important clinical material with respect to ADR. Especially the complications and adverse event components, indications and contraindications and the clinical difficulties surgeons encounter to reach certain treatment decisions. This was particularly well described when noting the difficulties in selecting patients for lumbar fusion or L-ADR. These issues are controversial. The issue of ASD is well addressed and is controversial as well. ASD is a key argument in the use of L & C-ADR but it is not well understood. This important point was addressed sufficiently.

It is important to note the significant differences in clinical indications and patient selection between the 2 devices(L & C-ADR). This was addressed but I think could have been stressed a little more.

Although I stated my knowledge on the specific economics of these devices is limited it is clearly an important public policy issue given the expense of these devices and relative lack of long term data when compared to current treatment.

DePuy Spine, Inc.
325 Paramount Drive
Raynham, MA 02767-0350 USA

Toll Free Customer Service: +1(800) 227-6633
Toll Free Receptionist: +1(800) 365-6633
Direct Receptionist: +1(508) 880-8100
Fax: +1(508) 828-8122

September 9, 2008

Ms. Leah Hole-Curry
Program Coordinator
Washington State Health Care Authority
Health Technology Assessment
Health Care Authority 676 Woodland Square Loop SE P.O. Box 42712 Olympia, WA 98504-2712

Re: Health Technology Assessment Draft Report – Artificial Discs Replacement (ADR)

Dear Ms. Hole-Curry:

Thank you for the opportunity to comment on the Artificial Discs Replacement Draft report from Aug 26, 2008, for the Washington State Health Care Authority.

DePuy Spine, Inc. is an operating company of DePuy, Inc. one of the world's leading designers, manufacturers and suppliers of orthopedic devices and supplies. We are known throughout the medical world for the development of innovative solutions for a wide range of spinal pathologies.

The two issues discussed in this letter relate to: 1) the methodology utilized by Spectrum Research Inc. (SRI) to assess the quality of the evidence; and 2) clinical comparisons of ADR to continued conservative nonoperative care.

1. Methodology to Assess the Quality of the Evidence.

SRI's methodology to assess the quality of the evidence uses a 4-Level grading system, defined on page 44 of the draft HTA. Level I evidence is defined as a "Good Quality RCT" and requires all of the following criteria: concealment; blind or independent assessment for important outcomes; co-interventions applied equally; follow-up rates of 85%+, adequate sample size and intent-to-treat. Evidence from studies that violate any of these methodological criteria is graded Level II ("moderate or poor quality RCT"), Level III (moderate or poor quality cohort or case-control) or Level IV (case-series).

SRI's evidence assessment system may be appropriate for pharmaceutical studies. However, there are unique considerations related to surgical device trials. For example, "blind or independent assessment for important outcomes" may not be feasible in surgical device trials[1], and as such, no trial can possibly be designed to qualify for Level I. As listed in Appendix G, none of the studies reviewed in this report was graded Level 1. This grading system may therefore not be appropriate for reviewing surgical evidence for spinal devices.

2. Clinical comparisons of ADR with continued conservative nonoperative care.

The Summary and Implications section page 92-94 reports the lack of studies comparing ADR to continued conservative nonoperative care. The use of nonoperative control arms was previously discussed in great details at the Washington State Health Technology Assessment on Fusion, on November 16, 2007. Whether a nonoperative control arm is compared to a fusion or an ADR group, similar considerations apply. Specifically:

1) The assumption that surgery (whether fusion or ADR) and nonoperative care are competitive/interchangeable treatments utilized under similar circumstances is incorrect, as was stated by Dr. McCormick during the Washington State Health Technology Assessment on Fusion on November 16, 2007. In fact, patients are only considered surgical candidates after failing nonoperative care. The same applies to arthroplasty: in the CHARITÉ IDE study, patients in the ADR group were on non-operative care for an average of 32.4 months (median: 23.0 months) while patients in the fusion group were on nonoperative care for an average of 26.7 months (median 19.0 months) [non-published data – on file at DePuy Spine].
2) The fact that surgery is a treatment offered when nonoperative care fails has been made apparent in the SPORT trial. In this study, out of 145 patients assigned to nonoperative treatment for lumbar degenerative spondylolisthesis, 49% (71) underwent surgery. The magnitude of this cross-over rate illustrates the fact that as patients worsen, surgery becomes the main treatment option[2]. It also points out to the inherent difficulty in generating statistically meaningful data that can conclusively address the issue of nonoperative care vs. surgery.
3) No standardized nonoperative treatment exists for patients with degenerative disc disease. While pilot studies have discussed the potential effectiveness of specific rehabilitation programs, these programs have not been validated. The clinical effectiveness of nonoperative treatments still need to be established, prior to being used as controls to ADR in randomized controlled trials[3,4].

As the comments included herein may potentially impact the overall interpretation of available evidence for ADR, we would like to respectfully suggest that these points be considered in the final version of the ADR HTA.

Sincerely,

Chantal E. Holy, PhD
Director of Scientific Affairs
DePuy Spine

Reference List

1. Lilford,R., Braunholtz,D., Harris,J., & Gill,T. Trials in surgery. [Review] [66 refs]. *British Journal of Surgery* **91**, 6-16 (2004).

2. Weinstein,J.N. *et al.* Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N. Engl. J Med.* **356**, 2257-2270 (2007).

3. Fairbank,J. *et al.* Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ* **330**, 1233 (2005).

4. Brox,J.I. *et al.* Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* **28**, 1913-1921 (2003).

Medtronic, Inc.
State Government Affairs
2600 Sofamor Danek Drive
Memphis, Tennessee 38132

www.medtronic.com

September 9, 2008

Ms. Leah Hole-Curry
Program Director
Health Technology Assessment Program
676 Woodland Square Loop SE
Olympia, Washington 98504-2712

RE: HTA Artificial Disc Replacement Draft Report Comments

Dear Ms. Hole-Curry:

We appreciate the opportunity to comment on the Artificial Disc Replacement Draft Report. As you are probably aware, Medtronic Spinal and Biologics Division manufactures products that treat a variety of disorders of the spine. These products are utilized by spinal and orthopedic surgeons to treat patients and restore their quality of life. As the manufacturer of the first cervical disc to market, we are very interested in this review and want to ensure that patients in Washington retain access to the latest and most effective technologies.

We have reviewed the Draft Report prepared by Spectrum Research, Inc. (Spectrum) and found it to be thorough. However, we do have several comments pertaining to the findings and analysis regarding cervical disc arthroplasty.

Summary of Findings Does Not Reflect the Review/Analysis

As stated in the report, there is moderate evidence in support of the safety and effectiveness of C-ADR compared to ACDF (see questions 1 and 2). However, in the Summary of Findings on page 9, the report diminishes the strength of the data with the statement that there is "insufficient evidence to draw extensive efficacy/effective conclusions comparing ADR with a broad range of treatment options." In this regard, statements are made that there is no direct comparison to conservative operative care or other forms of surgical intervention. We believe these statements are misleading for two reasons. First, patients who are candidates for disc arthroplasty would have already exhausted an appropriate period of conservative care. And second, ACDF is historically the standard of care. See further discussion below.

Additionally, the methods, grading, rating, and application of the evidence are unclear, particularly in regard to reference numbers 76, 77, and 78; further explanation would be useful and helpful to ensure consistent evidence evaluation.

Non-operative Care and Other Surgical Procedures

Patients who are indicated for C-ADR have undergone at least six weeks (note the report has a couple of erroneous references to six *months*) of non-operative care without success. As noted on page 24, non-operative care does not reverse or stop the disease progression or resolve pain in 50-70% of myelopathy patients and 25% of radiculopathy patients. Further weakening and worsening pain often occurs in patients with cervical disc herniation or spondylosis causing radiculopathy or myelopathy. To relieve these symptoms, decompression is required, and as noted on page 26, the current definitive standard of care for these patients is ACDF.

The background of surgical options for these patients is a necessary component of the report in laying out the history of treatment options for these patients (see pages 11,17, 18,25 and 26) and ultimately reaching the conclusion that ACDF is the relevant comparison to C-ADR. However, we find that further explanation of the historical perspective and reference to broader literature (i.e. beyond two articles on myelopathy, and one on Blue Cross Blue Shield's disc arthroplasty technology assessment in references 5, 6 and 7) would improve the quality of the report. In general, beginning in the early 1900s and for many years, posterior decompressions were the standard of care. However, with limited access and exposure to midline disc fragments and calcified spurs, anterior approaches were introduced in the 1950s. Instrument reconstruction and fusion was necessary to promote fusion, allow for earlier return to activities of daily living, and avoid kyphosis. Posterior decompression continues to be a treatment option for soft accessible disc fragments and foraminal osteophytes in radiculopathy. However, anterior decompression and fusion have become the standard of care for central and paracentral disc herniation, radial osteophytes and uncovertebral joint spurs in radiculopathy and myelopathy.

It is important to clarify that not all patients who currently undergo cervical spine surgery would be candidates for C-ADR. Auerbach (2008) conducted a retrospective study of 167 patients who underwent cervical spine surgery. Based on assessment of the patients' history in terms of the indications and contraindications for C-ADR, 43% would have been candidates.

Level of Evidence: Questions 1 and 2

The Spectrum report cites several sources on rating the evidence; it is not clear from the report how the rating methods were selected or utilized to rate the studies. The AHRQ report (#78) notes that there are numerous methods for rating clinical evidence, and that "[u]sers wishing to adopt a system for rating the quality of RCTs will need to consider the topic under study, whether they prefer a scale or checklist, and ease of use of the system." Two other reports were cited by Spectrum to support their rating of the

evidence, the Phillips/Oxford Centre's guidance for rating evidence (#76) and the Atkins' criteria (#77). Because the AHRQ report conclusions on rating evidence suggest a system related to the topic studied, the Phillips/Oxford Centre's rating system is fairly complex to interpret and Atkins' is relatively general, we suggest that more details are necessary and should have been included in order to accurately support the bases of the ratings.

In other methodologies, it is more typical to rate a randomized control study, such as the studies cited in the report, as Level I studies with several subgroups (SIGN 2008 and van Tulder 2003). Due to the inherent difficulties of conducting a randomized, blinded medical device surgical trial, compared to a drug trial, it is not logistically or ethically feasible to meet 100% of the criteria for the highest rated RCT. Nevertheless, the quality of the referenced studies warrants a higher rating than moderate, level II.

Ms. Leah Hole-Curry, Program Director
September 9, 2008
Page Three

Based on the analysis of three level II RCTs (see pages 66-81), the evidence meets the criteria for quality, quantity and consistency evidence showing that C-ADR is superior to ACDF for overall study success and neurologic success, and comparable for NDI, pain, and safety. The report acknowledges this with its statement on page 97, third column, that "this result is based on FDA criteria for overall success and pooled estimates from two completed trials and interim FDA analysis of a 3rd trial." To reflect this, "quantity" in the evidence strength table (see pages 97 and 98) should be changed from "-" to "+". In addition, for these listed outcomes, the comment regarding further research in column two should be removed.

With regard to motion, as noted on pages 9 and 93, there is evidence that motion is maintained or improved up to four years. For text on pages 74 and 97, "improved" should be included with "maintained."

For adjacent segment disease, Mummaneni (2007, #98) and Robertson (2005, #124) report lower risk of ASD requiring surgery for C-ADR vs. ACDF (see page 76). While longer term data are necessary, these two year results are worthy of acknowledgement on pages 9, 93 and 97.

Omission of Studies: Question 2

There are several studies addressing key question 2 that we recommend Spectrum include in its analysis. In addition to the 22 studies cited in the technology assessment contributing to the evidence on the safety of cervical disc arthroplasty, the following six studies also provide information on safety outcomes.

- Anderson et al. (2008) compared the adverse events associated with the Bryan artificial disc to anterior cervical athrodesis in a randomized controlled trial (n=463). This study found that both procedures had a low incidence of significant adverse events related to the procedure. Statistically, more serious adverse events and reoperations occurred in the fusion group while a significantly greater number of less serious surgically related events occurred in the investigational group.

- Sasso et al. (August 2008) found no evidence of migration, no subsidence at 24 months, and no evidence of bridging bone across the implant disc spaces in cases implanted with the Bryan disc in the same randomized controlled trial reported on by Anderson et al (2008). The radiologists did find a 2.5 % incidence of anterior osteophytes in the investigational patients.

- Rates of adverse events between fusion and artificial cervical disc (Prestige ST) arms of a single center randomized controlled trial (n=19) were similar in Riina et al. (2008) after 24 months follow-up.

- In Sasso et al. (February 2008), flexion/extension range of motion was not determined to be significantly different between populations (randomized clinical trial comparing fusion to Bryan cervical disc replacement, n=22) at adjacent segments. There was a significant difference in translation at the level above the fusion after the surgery. To accomplish similar flexion/extension range of motion at the level above the fusion, increased translation was found in the fusion group. This increased translation at the adjacent level may place excessive loads on the annulus and the facet joints above a cervical fusion.

Ms. Leah Hole-Curry, Program Director
September 9, 2008
Page Four

- Yang et al. (2008) identified no cases of prosthesis subsidence or excursion in a case series of 19 patients implanted with the Bryan artificial disc after an average of 24 months follow-up.

- In a 47 patient case series of patients who received the Bryan artificial disc, Kim et al. (2008) reported that the overall sagittal balance of the cervical spine was usually preserved. The study also reported that no definite clinical deterioration due to kyphogenesis of the functional spine unit or overall cervical alignment was observed.

We believe these studies provide additional evidence to support the safety profile question reviewed by Spectrum Research; and furthermore, that the report would not be complete without them.

Treatment Guideline/Coverage Policy Omissions

The Spectrum report provides an overview of payer assessments and policies for cervical disc arthroplasty. However, many key payer policies and state workers' compensation treatment guidelines were not included. In order to provide a complete and comprehensive analysis, these policies should be included.

Currently, several state workers' compensation policies and/or treatment guidelines allow coverage of the cervical artificial disc and others allow coverage and payment on a case-by-case basis. Colorado's workers' compensation guideline was created by a physician advisory panel that reviewed the clinical evidence and determined that the cervical disc should be covered. Montana established a coverage policy in the workers' compensation program for FDA approved devices used in a single level after a period of conservative care. Wyoming also has a positive workers' compensation guideline for cervical discs which was established by a physician committee. Finally, many states have proposed guidelines that establish coverage for cervical discs, including New York and Oregon. These guidelines/policies are currently in the regulatory process and have not yet been finalized.

In addition, there are various positive commercial payer policies including Aetna's national coverage decision and certain Blue Cross Blue Shield plans. The Federal Employee Health Benefit Plan also provides positive coverage that allows reimbursement for any FDA approved device. We would encourage review of those plans and inclusion in Table 6 on page 34 of the report.

Economic Data Omitted

There are two recent economic presentations that we would encourage Spectrum to include in the disc arthroplasty review:

- Anderson, Paul, Traynelis, Vincent. Economic Analysis of Artificial Cervical Disc Replacement versus Anterior Cervical Fusion Surgery in the Non-Elderly: Impact on Hospital and Societal Costs. Presented at the North American Spine Society Meeting, Seattle, Washington, September 27-30, 2006.

Ms. Leah Hole-Curry, Program Director
September 9, 2008
Page Five

- Menzin, Joseph, Zhang, Bin. The Economic Impact of The Prestige Cervical Disc System: Results From A Randomized Clinical Trial. Presented at American Association of Neurological Surgeons Meeting. Chicago, Illinois, April 27-May 1, 2007.

The Anderson study is an economic analysis of three prospective, multi-center, randomized clinical trials and 2 single arm trials assessing arthroplasty and anterior cervical fusion. The study included 649 disc and 580 fusion patients with single level radiculopathy or myelopathy with a mean age of 44 years. The results show that disc surgery saves $200 per patient, on average, relative to fusion. From a societal perspective, the savings were $5273 per patient favoring disc and the finding was based on a 35-day faster return to work.

The Menzin study is a randomized clinical trial of 541 patients with single-level disease; 276 of the patients received cervical disc arthroplasty and 265 received fusion surgery. Clinical data were collected preoperatively and postoperatively for a maximum time period of two years and the study measured direct medical costs and work productivity. The results showed that compared to fusion, disc arthroplasty resulted in higher neurological success rate and better functional outcomes, fewer secondary procedures and an earlier return-to-work. The net economic benefit, defined as the difference between value of work productivity and direct medical costs, was $5988 for the cervical disc arthroplasty patient.

Although these economic studies have not yet been published, they have been presented at two leading physician specialty society meetings and represent valid information that should be considered in a review of cervical disc arthroplasty. We would encourage Spectrum to include this information in the final report.

Errors to Be Corrected

Prior to release of this report in a final form, we recommend a final quality check of the document, including consistency of the narrative and accuracy of the references. Examples follow.

Page	Description
General	In various tables, author/year only citations are provided. As the bibliography is not ordered alphabetically, it is not easily possible to find these citations. If an author/year reference is included, the citation number should also be provided.
21	6 weeks vs. 6 months of conservative care
23	Discussion of wear debris evaluation; reference to Singh (#9), which is a paper on C-ADR impact on physician practices Reference to the Prestige and ProDisc indications is incorrectly listed as #40, which is for the Bryan disc.
24	6 weeks vs. 6 months of conservative care
26	Sentence including “”the most methodologically rigorous longitudinal study” includes references 62-67; #62 only is correct
29	Statement made that all assessments performed prior to any RCTs. This is not accurate as the Hayes and NICE study reference RCTs.

40	"Nine" HTAs are referenced. It's not clear which studies comprise the nine.
42	The flow chart focuses solely on the RCTs. It's not clear how the search/exclusion/selection of the additional 22 articles was completed.

Ms. Leah Hole-Curry, Program Director
September 9, 2008
Page Six

47	In reference to the 2007 Bryan panel overview, two references (101 and 102) are provided relative to the Bryan Disc. One is incorrect (Nabhan 102), as it includes the ProDisc.
50	The Mummaneni study included a secondary hypothesis for superiority, which is not cited.
50, 54	The baseline differences in patient characteristics were not statistically significant; this should be stated.
51	References to the "high percent of lost to follow-up" for the Mummaneni study are inaccurate. A priori, the analysis intended to look at the first 250 completers in each group and the lost to follow-up are reported for these patients. The other cases were not yet due for follow-up.
52	Nabhan's study was not designed to assess VAS pain; therefore the statement regarding the small sample size is inappropriate.
73	SF-36: all references to "%" should be points. The Mummaneni changes in SF-36 scores should be verified; rather than 11 and 9, and 7 and 8, it appears that 13.1 and 11.8 and 7.4 and 7.5 are correct.
81	In addition to the total number of patients with complications, addition of a column with the total number of patients in the "x" specified articles would be helpful to put in perspective the range of complications. Double check references. As an example, #124 may not report on hematomas.
91	With exception of the 2007 cases, all prior years would represent patients in IDE trials.

Thank you again for the opportunity to comment on the Artificial Disc Replacement Draft Report and to participate in the HTA process. We stand ready to answer any questions on these comments and will gladly respond to non-proprietary information requests from Spectrum. Such requests for information, however, should be directed to my attention rather than our customer service or sales staff.

Sincerely,

Dena Scearce, JD

Director, State Government Relations
Medtronic, Inc.
Spinal and Biologics Division
1800 Pyramid Place
Memphis, TN 38132
Office: 901.344.1573
Cell: 901.428.3516
dena.l.scearce@medtronic.com

Additional References

Auerbach J, Jones K, Fras, C, et al. The prevalence of indications and contraindications to cervical disc replacement, Spine Journal. 8(5): 711-6, 2008.

SIGN, Scottish Intercollegiate Guidelines Network, SIGN 50: A guideline developer's handbook, Revised edition, January 2008, NHS, Quality Improvement Scotland.

Van Tulder M, Furlan A, Bobbardier C, et al. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group, Spine. 28 (12): 1290-99, 2003.

Anderson PA. Sasso RC. Riew KD. Comparison of adverse events between the Bryan artificial cervical disc and anterior cervical arthrodesis. Spine. 33(12):1305-12, 2008 May 20.

Kim SW. Shin JH. Arbatin JJ. Park MS. Chung YK. McAfee PC. Effects of a cervical disc prosthesis on maintaining sagittal alignment of the functional spinal unit and overall sagittal balance of the cervical spine. European Spine Journal. 17(1):20-9, 2008 Jan.

Riina J. Patel A. Dietz JW. Hoskins JS. Trammell TR. Schwartz DD. Comparison of single-level cervical fusion and a metal-on-metal cervical disc replacement device. American Journal of Orthopedics. 37(4):E71-7, 2008 Apr.

Sasso RC. Best NM. Cervical kinematics after fusion and Bryan disc arthroplasty. Journal of Spinal Disorders & Techniques. 21(1):19-22, 2008 Feb.

Sasso RC. Best NM. Metcalf NH. Anderson PA. Motion analysis of Bryan cervical disc arthroplasty versus anterior discectomy and fusion: results from a prospective, randomized, multicenter, clinical trial. Journal of Spinal Disorders & Techniques. 21(6):393-399, 2008 Aug.

Yang S. Wu X. Hu Y. Li J. Liu G. Xu W. Yang C. Ye S. Early and intermediate follow-up results after treatment of degenerative disc disease with the Bryan cervical disc prosthesis: single- and multiple-level. Spine. 33(12):E371-7, 2008 May 20.

September 9, 2008

Leah Hole-Curry, JD
Program Director
Washington State Health Care Authority
Health Technology Assessment Program
P.O. Box 42712
Olympia, WA 98504-2712

VIA E-MAIL

RE: HTA Draft Evidence Report on Artificial Disc Replacement (ADR)

Dear Ms. Hole-Curry:

We would like to thank the Washington State Health Care Authority Health Technology Assessment Program (HTA) for the opportunity to provide comment on the draft health technology assessment to systematically review the evidence available on the safety, efficacy and cost-effectiveness of artificial disc replacement (ADR). We fully endorse and applaud the HTA's ultimate goal of improving patient care through application of scientifically grounded therapies, including newer health technologies. As medical specialty societies representing the primary providers of ADR, we have some concern about the content of the evidence report, but more about the process by which it was achieved. The comments provided herein are submitted with the intent of assisting in providing the residents of Washington State with the best, most cost-efficient healthcare possible.

HTA Draft Report: Artificial Disc Replacement (ADR) 8.26.08
General Comments on the Lumbar Arthroplasty Section of the Assessment. This draft evidence report summarizes the preclinical and clinical literature available on lumbar arthroplasty, and defines the levels of evidence presented in the articles based on a 4-point scale (page 44). Level-1 data requires studies with blinding of treatment and analyses, follow-up rates of 85%, adequate sample size and intent-to-treat analyses. Violation of any of these conditions down classifies trial results to lower levels of evidence.

This methodology is particularly challenging in the realm of spinal device trials. Surgeons are obviously not blinded to treatment arms, and patients are aware of the nature of their implants immediately post-surgery. Blinding of imaging results for analyses purposes is also not achievable, as various devices are clearly identifiable on x-rays.

As a result, and not surprisingly, all RCTs reviewed in this report are described as Level-II studies or "Moderate or Poor Quality RCT," despite the fact that these studies were mandated, reviewed and accepted by FDA using strict clinical and statistical methodologies. In fact, it is unclear whether any RCT conducted to date for spinal surgery could possibly qualify as a Level I study. It is therefore questionable whether this 4-point scale is adequate to qualify RCTs for spinal surgery and lumbar arthroplasty. This specific issue was raised and discussed

recently by Lilford *et al.*, who similarly confronted the issue of blinding and overall quality of resulting evidence, from surgical trials.[1]

In November 2004, the National Institute for Clinical Excellence (NICE – UK) issued a Guidance on Prosthetic Intervertebral Disc Replacement, indicating that "current evidence on the safety and efficacy of prosthetic intervertebral disc replacement appears adequate to support the use of this procedure." This report was based on data available before January 2004. Since that time, both the Blumenthal *et al.* and Zigler *et al.* studies were published, further describing the safety and efficacy of lumbar arthroplasty.

A common consideration among technology assessments is the lack of data to determine the longer term safety and efficacy of lumbar arthroplasty compared to fusion (e.g., page 93 of the WA HTA draft report). The five-year CHARITE Artificial Disc IDE study, recently completed and presented at CNS/AANS Joint Section and EuroSpine 2008, addresses this shortfall (see attached abstract). This data was accepted for publication by *The Spine Journal* on August 5, 2008, and is currently in press.[2] This study represents the largest and longest RCT performed on arthroplasty to date, and addresses the need for long-term safety and efficacy data, as indicated in the WA HTA draft report.

Combined Review of Lumbar and Cervical ADR. One overall concern is that, despite disclaimers, the results from lumbar and cervical ADR appear to have been blended. These two treatments are very different—lumbar ADR is an alternative to fusion for the primary treatment of mechanical disabling low back pain, while cervical ADR is a motion alternative to the segmental reconstruction that is required after decompression for a primary extrinsic neurologic problem. Blending the two types of ADR is like comparing a car to a building because they are both made of steel. Their functions are very different. Assessment of these entities needs to be made separately.

Executive Summary. Efficacy/Effectiveness of Artificial Disc Replacement (ADR) (p. 8). The report indicates that "neither the type of conservative treatment nor the level of patient compliance with pre-study conservative treatment was detailed in the published studies used in this technology assessment and therefore, unknown." We would refer you to the comments below regarding the section *Results 3.1.* However, it is also arguable that if the type and compliance with conservative treatment are unknown, the comparison between ADR and nonoperative treatment cannot be effectively made in this technology assessment.

Critical Appraisal of Study Methods, ProDisc-L (p.49). The report refers to "a number of methodologic flaws..." that dropped the study to a Level of Evidence II. However, only two "flaws" are mentioned:

1. The report indicates that there were 32% smokers in the fusion group and only 21% smokers in the ADR group, and states "smoking has been shown to increase the risk of nonunion in patients undergoing lumbar fusion." However, the fusion rate in this study, verified by independent third party radiologists on digital radiographs, was 97%. The independent radiologists felt that only 1 of the 75 fusion patients did not meet strict radiographic criteria for fusion (and that patient

was clinically asymptomatic). What is the methodologic "flaw," when smoking did not have any significant deleterious effect on fusion?

2. The report points out that although 183 ADR patients and 93 fusion patients were enrolled, only 162 ADR and 80 control patients were treated. This occurred because once the threshold for treated patients was reached, the study stopped. There were 21 + 13 patients in the "pipeline" awaiting insurance authorization, medical clearance, surgical scheduling, etc. who were enrolled, but not treated. Once the study numbers had been reached and the study closed, these patients were not subsequently treated within the study. They had to choose between more conservative care, either accepting conventional surgical treatment (fusion) or wait for another FDA clinical study. They were no longer considered part of the ProDisc-L study population. Continuing to include these patients in the overall follow-up rates, as the report suggests, is not logical. The FDA had no interest in including these non-treated patients, since they had no treatment data points.

Results 3.1 (p. 57). The report states that, "There were no studies found comparing lumbar ADR with nonoperative care." This is untrue. Minimum requirements for patient enrollment in the ProDisc-L IDE study were six months of failed conservative nonoperative treatment. In fact, the average patient in the ProDisc-L IDE study had nine months of conservative nonoperative treatment.

The baseline Pain Visual Analog Scale (VAS) and Oswestry Disability Index (ODI) scores for patients in this study represent the best each patient could achieve with nine months of conservative care. Within the first six weeks after surgery, this patient population demonstrated an immediate and significant improvement in both pain VAS and ODI, which was maintained to the two year study window (and has now been shown to be maintained out to five years on subsequent reporting). The only variable introduced between the preoperative baseline score and the six week postoperative score was the surgical intervention. Nine months of static, failed nonoperative therapy with an immediate and significant change postoperatively is a fair comparator.

In response to the criticism that the nonoperative care was not standardized, we would point out that the nonoperative care used in the study was the conservative care patients receive in communities across the US. The value of a multicenter, multisurgeon study is exactly that: it normalizes the variations one might see in a single facility or single surgeon's practice. Since there is so little agreement on what constitutes adequate conservative care, this actually represents a better nonoperative control than one designed as part of a study, since consensus would never allow all readers to agree that this structured treatment was adequate. This was a real-life, same-patient conservative care control model that could easily be considered a third study arm.

Summary and Implications (p. 92-93). Remarks on all five points and subpoints are negatively biased to the degree that it gives the perception that this study group was given a mandate to show negative results. The analysis appears structured to emphasize the negative aspects of this new technology, and to downplay positive aspects.

Disclaimer (p. 2). The disclaimer on the report is appropriately included and should be considered. "…Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider

this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability."

The HTA Process

The work group would like to provide comments based upon its experience with the process in an effort to continue to improve upon it.

Dedicated Review Time for Draft Evidence Report. One of the primary goals of the health technology assessment program is … to make the "coverage decision process more open and inclusive by sharing information, holding public meetings, and publishing decision criteria and outcomes." (www.hta.hca.wa.gov). At least for this topic, inadequate review time was allowed for the public comment period on the draft evidence review. The 200+ page draft evidence report took months to write. A two week review period (including a holiday weekend) was not enough time to generate substantive public comments. At least one month needs to be made available to potential reviewers to allow truly inclusive and substantive comment.

Technology Selection. Given that three of the first ten topics selected for assessment by HTA are directly related to spine (lumbar fusion, discography, ADR), the work group is concerned that there is an inordinate focus on spine. This raises concern about bias in the selection process.

Although topics under consideration for selection are eventually ranked according to a specified process, the initial selection of topics for briefing and ranking is done in such a manner that there is a concern about bias. The initial topic suggestions are made by agency medical directors alone (at least until a public process is implemented) which allows political bias and budget conflicts to potentially enter the process and bias which topics are put in the pipeline for consideration before briefing and ranking in a more transparent manner occurs. The fact that technologies not selected still remain on the list for future consideration is also concerning. Each technology should be individually vetted at the time of consideration, not wait-listed if initially rejected.

Clinical Committee and Panel Hearing. We would also encourage the participation of experts in the process for each topic area considered. In addition, scheduling of the panel meeting in conflict with a professional medical meeting of major stakeholders discourages input from key stakeholders.

The HTA should also consider the concept that there is variability of opinion in the selection of any treatment. A mature process brings in individuals who represent the spectrum of variation. This inclusion of diversity of opinion at the start of the process allows the best critical analysis, weighing the advantages and disadvantages of new or existing interventions. It also has to weigh the evidence for benefit of the alternative treatment. In this process of technology assessment, cost is not supposed to be a consideration. It is recognized that the follow-on step is allocation of scarce resources. In order to apply that step appropriately, cost-effectiveness analysis is then required. Unfortunately, in most surgical interventions, robust cost-effectiveness data is limited and cost minimization is substituted for cost-effectiveness analysis which does not optimize patient care.

Lumbar disc arthroplasty is a potentially valuable technology that may ultimately play a significant role in the treatment of patients with axial back pain. Currently, there are significant knowledge gaps regarding the true benefit of lumbar disc arthroplasty in patients previously considered candidates for fusion. It is apparent that the indications for arthroplasty may not be the same as the indications for fusion and that patients who are candidates for one procedure may not always be candidates for the other. Prospective series and randomized trials have demonstrated that these devices do provide substantial pain relief and functional benefits for some patients. We encourage the Washington State HTA to consider the potential benefits of both lumbar and cervical devices on a case-by-case basis and not categorically restrict covered patients access to evolving technologies.

Once again, we would like to congratulate the State on its initial steps towards using a logical, evidence-based process to evaluate technologies for coverage. Thank you for this opportunity to comment and we look forward to participating in the October panel meeting.

James R. Bean, MD
American Association of Neurological Surgeons

Thomas A. Zdeblick, MD
Cervical Spine Research Society

Anthony L. Asher, MD
Congress of Neurological Surgeons

Tom Faciszewski, MD
North American Spine Society

Karin Buettner- Janz, MD, PhD
Spine Arthroplasty Society

References

1. Lilford R, Braunholtz D, Harris J, Gill T. Trials in surgery. [Review] [66 refs]. *British Journal of Surgery.* 2004; 91:6-16.

2. Guyer RD, et al. Prospective, randomized multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the Charité™ artificial disc versus lumbar fusion—5 year follow-up. *The Spine Journal.* In press. 2008.

Attachment: 5-Year Charité Abstract—EuroSpine 2008

MONDAY, MAY 26

EUROSPINE - MOTION PRESERVATION

SP1

PROSPECTIVE, RANDOMIZED, MULTICENTER FOOD AND DRUG ADMINISTRATION INVESTIGATIONAL DEVICE EXEMPTION STUDY OF LUMBAR TOTAL DISC REPLACEMENT WITH THE CHARITÉ ARTIFICIAL DISC VERSUS LUMBAR FUSION,
5 YEAR FOLLOW-UP.

RD Guyer, PC McAfee, RJ Banco, F Bitan, A Cappuccino, FH Geisler, H Hochschuler, LG Jenis, JJ Regan, SL Blumenthal
Texas Back Institute, Texas, USA

BACKGROUND CONTEXT AND PURPOSE: The purpose of this study was to compare the safety and effectiveness at the 5-year follow-up time point of the CHARITÉ Artificial Disc with that of anterior lumbar interbody fusion (ALIF) with BAK cages and iliac crest autograft.
STUDY DESIGN/SETTING: Randomized controlled trial, 5-year follow-up, 90 arthroplasty patients and 43 fusion patients.
METHOD: Of the 277 CHARITÉ IDE eligible for the 5-year study, 160 patients completed the 5-year follow-up (133 randomized and 27 non-randomized - only randomized cases presented). Clinical evaluations included the following validated outcome scales: ODI, VAS, and SF-36, as well as patient satisfaction and work status. In addition, individual patient success/failure status as per the non-validated FDA binary criteria was defined as improvement greater than or equal to 15 pts in ODI vs. baseline, no device failure, absence of major complications and maintenance or improvement of neurological status.
RESULTS: Mean changes for ODI (CHARITÉ: -24.0 points vs. BAK: -27.5 points), VAS pain scores (CHARITÉ: -38.7 vs. BAK: -40.0) and SF-36 questionnaires (SF-36 PCS: CHARITÉ: 12.6 points vs. BAK: 12.3 points) were similar across groups. In patient satisfaction surveys, 78% CHARITÉ patients were satisfied vs. 72% BAK patients. Sixty-six percent patients in the CHARITÉ vs. 46.5% patients in the BAK group were employed full-time. This difference was statistically significant (p=0.0403). Long-term disability was recorded for 8.0% CHARITÉ patients and 20.9% BAK patients, a difference that was also statistically significant (p=0.0441). Additional index-level surgery was performed in 7.7% in the CHARITÉ group and 13.9% in the BAK group. Overall success was 57.8% in the CHARITÉ group vs. 51.2% in the BAK group (Blackwelders test: p=0.0359, D=0.10).
CONCLUSIONS: The results of this 5-year, prospective, randomized multi-center study are consistent with the 2-year reports of non-inferiority of CHARITÉ Artificial Disc vs. ALIF with BAK and iliac crest autograft. No statistical differences were found in clinical outcomes between groups.

Note: Personally identifiable health information redacted.

To whom it concerns:

This letter is to express my frustration and disgust with the insurance industry and their stranglehold on good, practical medical treatments that are available. The particular treatment to which I am referring is Artificial Disc Replacement (ADR) for pain and disability due to injury and/or degenerative disc disease.

I am a 54-year-old firefighter. I have been a career firefighter [redacted] Washington, since March 20, 1978 (30+ years now). The job is demanding, both physically and mentally, and I love it. I have had back pain for a fair portion of my life. Off and on at first, and constantly for the last several years, with some days being severe. I saw Dr. [redacted] a neurosurgeon [redacted] several years ago after he had performed a single level ADR surgery on one of my coworkers. The insurance carrier for the [redacted] paid for his surgery, which was part of a clinical trial, after they had initially denied coverage for it. That surgery was performed in [redacted], Washington.

I had a knee injury at work in 2007. While recovering from that, I was seen by my orthopedic doctor because I was having pain and numbness in my right hand and arm. He offered me a choice of injections in my neck or physical therapy. For anything other than that he said he would have to send me elsewhere as he would not do surgery on my neck. The physical therapy did little or nothing for the problems in my neck, nor for the pain and numbness in my arm and hand.

In October of 2007, after finally being pronounced fit for duty following the knee injury and its surgical repair, I woke up in the morning with worse than my normal back pain. My back got progressively worse over the next couple of days and left me hardly able to walk. My wife finally got me in to see an orthopedic surgeon who is a back specialist. He took x-rays and viewed an MRI of my lumbar spine. His recommended treatment was to have the lumbar spine fused and, at the same time, to perform decompression surgery (surgical removal of parts of the bones and ligaments around the nerve beds so they are not constantly inflamed or impinged upon). He felt I would be on pain medication for the rest of my life. I would not be able to return to work as a firefighter. I would most likely never be pain free, but he was sure I would be in "less pain than you are right now." This was not an acceptable solution to my wife or me.

While researching ADR on the Internet my wife found [redacted] Hospital in Germany. We e-mailed and talked to several people who had gone there to have ADR surgery. Everyone we talked to was positive [redacted] Hospital was the best place to have this type of surgery. All of them were thrilled with the outcome of their surgeries. We sent the doctors at [redacted] the x-rays and MRI films of my back to see if I was a candidate for ADR surgery. I was; so we made plans to go to have the surgery, hoping it would allow me to continue working or at least to be pain free and not dependant on pain medication for the rest of my life. We went back to [redacted] to see Dr. [redacted] before going to Germany. He was happy we had found out about [redacted] Hospital and had decided to go to there for the ADR surgery. He reviewed the films for my back and also looked at the neck films. He told us I would likely need ADR surgery on my neck as well. We were

shocked at this, but felt if the doctors in Germany felt the same, we would have it done at the same time, as we would likely never be able to afford to go back there again for a second surgery. When I contacted my employer's HR insurance person to see if I had to have the surgery pre-approved, I was told I did not need pre-approval, but that they may not cover it. The reason for this is that multi-level ADR surgery is considered "experimental" in the US. Mind you, the cost to have the surgery in Germany would be less than $60,000 and the cost for the spine surgeon's recommendation was around $250,000, and carried with it the likelihood of leaving me partially disabled. In our opinion ADR surgery is hardly "experimental." It has been being done successfully for over 20 years in Germany. On November 29th I had two cervical ADR done in my neck and two lumbar ADR done in my low back.

My recovery has been phenomenally successful. I have been released to return to work, as a firefighter, [redacted] (I could have likely gone back sooner but the doctors in Germany recommended I stay off for a full 6 months, if I could do so.) This would not be the case had I not had the means to go to Germany and have this surgery. I would have had to involuntarily retire from firefighting, would likely be partially disabled and would always have pain.

For my wife and I, this was only possible with a huge sacrifice. We were very fortunate to have found a way to afford to have the surgery. Our RV, which we had purchased after selling our home, would have to be used as collateral on a new loan. We are therefore "re-buying" it with money we thought we would have for our living expenses and our retirement. Due to the expense of the loan, our RV is up for sale and we will not have it for our future retirement; and that is because the insurance company has refused to cover any of the bills from the surgery. While I recognize that I used a "non-participating provider" for my surgery, I still feel the insurance should have covered the bills at the standard 60% they pay on non-participating provider's bills for other medical treatments. It seems ludicrous to me that the insurance company that gets it premium money from me has the right to tell me I can have the US treatment **at FOUR times the cost** with **no chance of a full recovery** rather than getting treated in the manner that is most likely the best hope I have for a full recovery at a hospital that specializes in this type of surgery **at ONE-FOURTH of the cost of the "standard" treatment in the US**, which is actually substandard treatment, in our opinion.

There are many people in this State suffering with back pain that can benefit from this surgery. Please make it possible for patients to get the surgery they need and require the insurance companies to cover their bills for this. Insurance companies should not be allowed to side step good treatments simply because they do not want to recognize how well they work. I have talked to several people since my surgery who have debilitating back pain. Some of these people have doctors that tell them **not to get fusion surgery done**, as it will not cure the problem, it only helps for a while and causes more problems at the adjacent disc levels in the back to where the fusion was done. These people live with pain and often have to be on drugs because ADR surgery is not available to them. ADR surgery is being done in the US in a few places. The FDA has approved ADR surgery. Please force the insurance companies to cover this type of surgery, even if it

means patients have to go overseas to have it done, so others do not have to wait needlessly in pain and continue to suffer.

Please feel free to contact me if you have any questions about what I had done, and why.

Thank you.

[redacted]

[redacted]

Washington State Health Care Authority
P.O. Box 42682
Olympia, WA 98504-2682

Health Technology Assessment
Subject: Artificial disc

This document is intended to provide information and criteria for consideration by the Washington state Health Assessment (HTA) team in their evaluation process of artificial disc arthroplasty as a preferred standard of care in the treatment of Degenerative Disc Disease (DDD). As I considered artificial disc as an option for myself, I did extensive study and research. After denial for coverage from my private insurance carrier I reviewed many more reports along with the policies and "supporting" data assessments for those policies from the insurance carriers. My discovery is that those policies were taken from limited data reports, used interpretive methods inconsistent with good scientific principals, drawing conclusions with little or no bases. The resulting policies use statements out of context of the clinic study data. While the results and recommendation of the HTA team will not be binding on the Insurance companies those findings and recommendations will most certainly influence their future policies. It is imperative that the HTA team consider all data weighing poor outcomes carefully to avoid such situations as patient selection, patient compliance, and non device related complications, i.e. Infections.

While certainly not a scientific method or statistical evaluation method consideration by the HTA an evaluation of artificial Discs must incorporate the important factor "quality of life". As with all medical treatment options for major conditions that affect quality of life pain is one of the most difficult to assess. Beyond the loss of strength, function, and numbness, pain is the motivation to seek a remedy. Yet pain is also is one of the most debilitating, affecting both physical and emotional wellbeing. Frustrated and almost unable to work or walk even a short distance I refinanced my house and had the surgery in Germany. My personal testimony is echoed by many who have found the hope offered by ADR to be a reality, restoring movement with NO pain related to my back except the joy of regained motion and few sore muscles. Evident to all that know me is my restored joy of life and purpose that goes far beyond expectations. It cannot be expressed in words the promise fulfilled by this miracle. I implore the HTA team to carefully evaluate this technology as their findings will significantly affect the lives of many.

The opening paragraph of the HTA Draft key questions and background states that important clinical questions have not been answered. I as laymen to the medical community found ample evidence to convince me that ADR was a good choice. The data included in the PMA for several of Fusion methods considered the "gold standard" and the AD clinical trials along with many other peer reviewed studies clearly established ADR as better in most cases. Further I consulted five prominent experienced surgeons (University of Washington, Cedars- Sinai, and UCLA medical center); each one articulated the same caution against fusion, with one stating, in writing, that fusion would be "disastrous". Then there are many studies and reports that analyze ADR and fusion outcomes including 10 plus years follow-up of ADR patients. The more appropriate question might be are the current evaluation methods susceptible to misinterpretation or bias?

Recognizing the HTA team to be highly qualified doctors and surgeons with many years of education and experience I am confident that each member recognizes the importance of comprehensive studies and subsequent reports. As with all research and development there will be results or situations that were not expected or desired. However a comprehensive medical study or report will present all the data collected in the expectation that by its inclusion a surgeon can better evaluate a particular treatment and avoid situations that caused undesirable results. With this in

mind the HTA team must use care to evaluate poor outcomes relative to the situation that caused them, and relative to the severity or recoverability.

One significant consideration is to acknowledge nothing is perfect there will always be unfortunate situations yet these should not unduly influence the promise Artificial Discs provides to patient and the health and well being.

There is no expectation that ADR will replace fusion in all cases, as many factors enter into the decisions made by surgeons as to treatment plan. But in most cases ADR will prove a better option than fusion. Thus the HTA team should evaluate ADR and set forth guidelines rather than limitations allowing the doctors and the patients to select the most appropriate technique that values the quality-of-life of the patient.

Sources of data

There are many sources of data on artificial discs and spinal fusion and outcomes of both. It seems inadequate to attempt to provide all the actual data. Along with the various sources available through peer reviewed journals, clinical studies and the FDA trials etc. I have included a link to the clinical report for the Maverick Lumbar Disc. I offer this for two reasons; this was the device that was successful for me and this device while still awaiting final FDA approval has a significant history in Europe. Further I will offer criteria and guidelines to be considered by the HTA team during the evaluation process.

Link to Maverick Disc report http://www.getadr.com/ADR%20Study.asp

- Because the recommendations of the HTA team will be used by the State of Washington in purchasing Health related services the HTA should carefully evaluate the sources of data to assure compliance with Washington Administrative Code (WAC). Definitions listed in WAC 388-501-0165 of "credible evidence" and "Hierarchy of evidence..."may affect how the HTA team should accept and evaluate data. As provided in WAC 388-531-0050 the definitions of "Peer reviewed medical literature" seem to restrict the use of data publications from many sources including insurance carriers. Thus, as addressed later in this document, evaluations, conclusions or recommendations from insurance companies and the affiliate organizations should not be included. Members of the HTA team will then be free to determine the efficacy of ADR based on peer reviewed documents and other sources not driven exclusively by financial gain.

- Numerous data sources are available: Use only data generated from the clinical environment at major medical or spine centers both in the US and in Europe. Use reports and studies published in peer reviewed journals submitted by credentialed board certified or recognized investigators and/or practitioners. Include data from medical centers that have participated in the long term success of ADR both in US and Europe. This may seem to bias the data however data from centers that have long term use history will mitigate the factors of training and low experience allowing device related factors to be evaluated effectively.

- Disallow data from policies or evaluations provided directly or indirectly from Insurance companies or the technology evaluation centers, affiliated with or controlled by the insurance companies. A review of one such evaluation on artificial discs provided no new data; it appeared as though the authors selected studies that supported the pre-determined conclusion sighting data, and statements taken out context of reports with otherwise complete and positive studies. As noted above, this source of data may not meet the

definition of peer reviewed document for the purposes of compliance with Washington code.

NOTE: as a reference source only, one Insurance company has compiled a comprehensive list of references to acceptable studies. See Aetna corporate policy on ADR.

- It is assumed the HTA team will use the recent HTA findings for Lumbar fusion as one comparative source for that method.

Evaluation

The following is intended to provide some general guidelines in the evaluation process of artificial disc.

- The HTA team must evaluate all artificial Discs available to patients regardless of FDA approval status. Failure to be inclusive will reduce the credibility of the HTAs findings and may limit a superior treatment from consideration and use after approval has been obtained.

- Evaluate each type of Disc separately: Examine the data for each type of artificial disc evaluating the results of Charite, ProDisc, and Maverick as used for Lumbar each separately. It may be discovered that a particular device has somewhat lower success than another. Comparison of all devices in the same "pool" may mislead and falsely lower the effectiveness of a device(s) that has superior performance. I.e. the Charite has some initial concerns that may be contributed to poor patient selection, lack of surgical instrumentation developed later for alignment of the device, un-constrained core and other factors.

- Evaluate cervical discs separately: Again these discs have various designs each with feature and limitations. Evaluate all discs both those that have FDA approval and those that are pending FDA approval. This may require the HTA team to review data from European clinics that use these devices and along with studies pending completion in the US clinical trials. Acquiring clinical trial data is restricted due to Federal laws during the approval process. The HTA panel may have the authority to request and receive that information the result of the effort will be a more comprehensive evaluation.

- Success worldwide: Artificial Disc Replacement (ADR) has been used in Europe for 15- 20 years with significant success. Data and associated reports from clinics where Artificial Disc replacement are the standard of care for DDD should be given maximum consideration. It may fall to the HTA to request and collect data that has not been specifically published in the United States.

- Evaluation should consider numbers of actual patients treated at hospitals where ADR technology has been used to treat a high number of patients along with comparative results of fusion as treatment option. This should focus on surgeons that have used fusion and ADR as treatment of DDD for many years. Limiting evaluation data to only the outcomes as presented in clinical trial reports may bias the data due to factors such as improper patient selection due to inexperience, inadequate training of surgeons, and limited numbers of actual cases at a particular location or performed by a particular surgeon. Particularly important is experience based improvement in both technique and instrumentation development.

Comment: Maverick has highly successful record in Europe and is pending US FDA approval. Do not disregard the Maverick or other devices solely because the FDA process is pending. I have the Maverick disc and I am 100% pain free!!!!! Able to do almost anything with significantly regained flexibility. Please see attached letter from a firefighter that had significant disability and very successful ADR surgery (2 lumbar 2 Cervical).

- Evaluation should consider the general success of the surgeon /medical center: Data that can be identified from a surgeon, clinic or hospital with a limited number of successes should be evaluated carefully to avoid non device related morbidity. It may be that the HTA panel should recommend ADR be performed at centers specializing in ADR, even if those treatment centers are outside the State. Not all surgeons have the experience to perform ADR without significant training. Example: the most effective cancer assessment and treatment is performed at only a few regional centers that specialize in new treatment regiments along with proven methods.

- Consider the factors related to bone graft donor site, dural damage, as presented in the HTA evaluation for fusion 11-07. The higher incidence to these complications added to the higher re- surgery for non fusion issues along with longer healing time the additional known limitation of loss of flexibility, and potential adjacent disc damage certainly raised questions of efficacy in my consideration process.

- Avoid comparing device data for non artificial disc such as the Dynesis system. This type of device can provide benefit in the proper application but should not be compared to ADR.

- Seek and use data that was generated from patient s that have had surgery many years ago. It may be necessary for the HTA team to request data from larger clinics and Spine centers to evaluate the long term viability of fusion as well as ADR.

- Consider return to work rates and quality of life. Since the well-being of the citizen of Washington should be the goal of this review. The patients overall benefit should consider quality of life along with financial interests of the patient. This is very significant since return to work rates can be important factors on the individual. Evaluation of the financial impact on the state, insurance companies, or L&I should only be a secondary consideration. However: Return to work and long term wellbeing of a patient may result in reducing the drain of L&I and welfare subsidies.

- Consider multiple levels ADR: While the FDA clinic trial limited the study to a single level as appropriate for good comparative data results, many patients primarily in Europe and at major spine treatment centers in the US have benefited from two and even three level replacement. One of the letters provided with this document is from a fire fighter that had two lumbar and two cervical disc ADR surgery performed at the same time. He is back to full time work as a fire fighter in Richland WA just six months after surgery (his doctor and the City insisted he wait that long). My roommate in Germany is a fellow that had twenty plus years of pain he received three discs. He is playing golf and riding horses etc.

- Review carefully the requirements of the manufacturers and the FDA PMA report for both methods. Whereas the FDA did not approve multi level ADR the prerogative of the surgeon allows them to provide multilevel ADR as an off label benefit. However in its findings the PMA on some of the fusion options strictly forbid the multi level option for fusion.

- Compare the results of ADR with the failure and re-surgery rate of fusion. Failure of the HTA review on fusion did not report the long term results of lumbar fusion nor did not appear to address the, failure of adjacent discs. Return to work or other factors

  Comment: While Washington State L&I Medical Treatment Guidelines do not meet the proposed requirements, noted below, for data acceptability, the guideline document on spinal fusion does provide an overall outline of the probability of failure of the fusion option. Follow link to Washington L&I web site.
  http://www.lni.wa.gov/ClaimsIns/Files/OMD/MedTreat/LumbarFusion.pdf

In reviewing the Draft Key Question and Background document the last paragraph identified a significant point, that of long term serviceability of ADR.

- When evaluating the long term efficacy of ADR Consider the longer term affects of fusion where patients experienced additional pain as a result of disc damage and failure of adjacent discs to the fusion site. Because disc height is restored by ADR pain producing impingement is this also corrected along with the pain associated directly with the degenerated disc and vertebrae.

- Consider return to work rates for each type of artificial disc and fusion methods.

- Consider the re-surgery rates for all devices.

- Consider non work related activity levels for patients from both treatment options. Consider pain short and long term pain. Consider pain management requirements for both ADR and fusion.

- Consider hospital stay: perhaps this factor should be given lower consideration however the benefit to the patient and the payee can be significant. Even one day less in the hospital can offset any differences in device costs.

- Since results of this assessment will certainly be scrutinized by the insurance industry consideration and conclusions should provide clear recommendations or guidelines in the areas of patient selection, individual device success, and medical center success.

Relative to Proposed Key Questions

Question 3

- This question suggests that efficacy or the safety of ADR is dependent upon factors other than the physical or physiological condition for Degenerative Disc Disease and the treatment option for it. The draft question identifies one special population as those seeking treatment under workers compensation. The assessment team must be especially diligent to separate the attributes and success of ADR without biasing conclusion of efficacy on those individuals that misuse a system. Data on pain return to work, or any other factors dependant on the "subjective nature" of those patients intent on benefiting from the system should be scrutinized to eliminate unreliable reporting. Further the history and credibility of the L&I system may not provide suitable source of data for any treatment method for back conditions like DDD and therefore it may advisable to exclude any data from that source from consideration.

- Another special population named under this question is the elderly. During the clinical trial process for approval of Artificial Discs this population was specifically and appropriately

excluded thus data from clinical trials is not available. However there may be patients that could benefit from ADR and that evaluation should be left to the experience of a qualified surgeon.

Comment: Certainly those that are advanced in age may not realize the greater flexibility benefit of ADR However: for those patients where some surgical intervention is required, I would submit that even a limited benefit of ADR might well be better than the slow process of bone fusion. In other words since the Artificial disc allows movement the patient is not limited during the healing process and since bone healing is slower they may well recover quicker with less risk of non-fusion.

Question 4

- Several factors should be considered in evaluation of the cost implications and cost effectiveness of ADR. Yet cost should not be the driving factor, if the goal of the HTA is the most effective benefit to the patient as is the intent of health care cost must not enter in to the evaluation until all other factors have been evaluated and then only after the options are found to be equal.

- Costs of new methods are often higher due to many factors and reduce significant after acceptance of both hospital and the medical professional.

- Consider the return to work rates. Fusions reported frequency for re-surgery alone should be a financial deterrent. Consider also patient risk and pain for second surgery. Add then quality of life, and return to work potential these costs may be indirect but must be part of the HTA's comprehensive assessment.

- Use cost analysis data from clinical institutions that perform both ADR and fusion methods and do so in sufficient numbers to establish usual and customary fee structure. These same institutions may also provide costs associated with re-surgery.

- Avoid cost data provided by Insurers: Data reported by insurance companies are likely to be skewed because insurers have not generally included ADR as a necessary treatment option thus any data will be limited. Further once ADR is an accepted option insurers can use their negation power to establish cost effective payment schedules. A few health insurers do accept ADR as medically necessary Aetna for one. A few self insured companies like Microsoft allow ADR Premera Blue Cross is the administrator of their plan. But those covered under Blue Cross's private plans are denied this treatment.

- Carefully review cost data from hospitals or medical centers that have established contracts for fusion etc. as it may bias the cost.

- One noteworthy article on patient return to work is provided in the link below.

http://www.washingtonpost.com/wp-dyn/content/article/2008/04/28/AR2008042802162.html

Comment: I received a quote from a prominent surgeon at UCLA medical center that would have performed surgery for me. (Had the insurance covered the recommended treatment) The hospital and surgeon provided surgeons fee, hospital and device costs for both fusion and ADR. Cost for ADR was slightly less due directly to anticipated shorter hospital stay.

Comments from the Washington State Agencies

Page 7: Comment [m1]
Fusion is not the standard of care for DDD in the absence of other findings.

Page 7: Comment [m2]
Theoretically intended to preserve motion...

Page 8: Comment [m3]
How is it decided to that 2 devices are similar enough to warrant pooling of outcomes data?
What is the effect of pooling data when the studies are not completely reported (database closed early, or not all randomized subjects reported at 24 months)?
It seems the treatments are not the same, the patients (at least for lumbar studies) are different enough at baseline to require some discussion of when it is or is not ok to combine populations.

Page 9: Comment [m4]
In the Charite RCT the comparator treatment was a technique no longer used.

Page 9: Comment [m5]
But there is not evidence showing mobility correlates with improved outcome or reduced ASD

Page 9: Comment [m6]
CMS assessment states “The theoretical mobility provided by the artificial disc has yet to directly correlate to a proven benefit in how the patient feels or functions, making the clinical significance of post treatment range of motion unclear. Therefore, CMS does not consider post treatment range of motion an important clinical outcome of interest in this memorandum."

Page 10: Comment [m7]
Why?

Page 12: Comment [m8]
Questions about the quality of these references.

Page 12: Comment [m9]
Do not agree and recommend deleting “standard of care” with regard to fusion.

Page 14: Comment [m10]
We do not agree with this statement. Some fusion trials have shown fusion surgery to be no better than non-operative treatments. The jury remains out on the question of ADR superiority to non-op treatment.

Page 15: Comment [m11]
This seems highly speculative. Does the investment marketing material take into account the available evidence on efficacy? What is the quality and value of this information?

Page 16: Comment [m12]
Seems highly speculative and not evidence based.

Page 16: Comment [m13]
What is the quality of the unpublished cost-effectiveness study that resulted in this statement and estimate? Why include it unless it is supportable?

Page 16: Comment [m14]
Post approval studies are required for the Prodisc-L and the cervical discs. Data from these post approval studies may help us understand the longer-term outcomes and costs.

Page 16: Comment [m15]
Perhaps more discussion of the uncertainty surrounding the diagnosis of disc pain and DDD.

Page 22: Comment [m16]
The procedure is technically more demanding, has a steeper learning curve, and requires greater precision than fusion surgery.
Especially if reoperation becomes necessary.

Page 24: Comment [m19]
Is the life span expected to differ between lumbar and cervical?

Page 32: Comment [m20]
CMS completed interal HTAs. Iniital report addressed Charite, second consideration addressed Prodisc when it was approved. Available here:
http://www.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=170 and
here: http://www.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=197

Page 48: Comment [m21]
The BCBS TEC report on Charite and the FDA statistical review note the lack of an a priori statistical plan for charite. Did they address sample size/power in the paper and SSED?

Page 49: Comment [m22]
Is there data that is not interim?

Page 51: Comment [m23]
There does not appear to be enough information in the methods section of this paper to warrant inclusion with rating as LOE II.
The discussion section is good.

Page 58: Comment [m24]
Half of the control group in this pooled analysis received an outdated form of fusion.

Page 59: Comment [m25]
The BCBS Tec assessment had this to say (in part) about the Charite trial:

The second concern is that the lack of a prespecified analysis plan, unexplained closure of the database before all patients reached completion, and lack of intent-to-treat analysis may cast some doubt on the analysis. Although the sponsor provided TEC with additional analysis that included patients that were excluded from the analysis presented to the FDA, it was unclear how many additional patients actually provided 24-month outcome data and what imputation was performed for missing or discontinued data.

Page 61: Comment [m26]
The Prodisc FDA summary shows the fusion group with and ODI of 39.8 at 24 months and the ADR group at 34.5. The entry ODI was 40. Though both groups improve the resultings disability score is not very good. The Charite study has and ODI entry of at least 30 and the baseline for these patients was about 10-12 points lower than the Prodisc. It doesn't seem reasonable to combine the results from these studies for meta-analysis given the differences in device, characteristics on key entry/outcome criteria and control
treatment. CMS included the whole ODI table from the SSED in their HTA.

Page 67: Comment [m27]
These are patients who received ADR and have ASD?

Page 80: Comment [m28]
What about the Putzier paper (ref 114), 2005- 60% of a subset that did not have disc removed for fusion had HO? Hetertropic ossification in majority reported on. Paper included in CMS analysis.

Page 80: Comment [m29]
Please see the CMS report. The section on adverse events is thorough and includes excerpts from key longterm follow-up studies.

Page 82: Comment [m30]
Mehren study not included here? Rate of ossificiation almost 70% at 1 year.

Page 86: Comment [m31]
Is this chart necessary?

Page 93: Comment [m32]
If the study is designed to be a non-inferirority trial, how is it possible to reach a conclusion of superiority?

Page 93: Comment [m33]
"If the results following completion of the trial are similar to the interim results of that same trial, the confidence in the evidence that C-ADR is superior to ACDF will increase." This comment is speculative.

Page 94: Comment [m34]
But long-term follow-up studies seem to show the potential for a high number of failures leading to fusion or spontaneous fusion. How doe s that fit with the 2 year, interim study outcome data?

Addtionally the FDA is requiring Post Approval studies for all patients in these trials. Do interim reports continue to support the original 2 year findings? When will we know the true outcomes for all people treated in these IDE studies?

Page 94: Comment [m38]
It was deemed reasonable by spectrum: Why?