Appendix A: General Methodological Principles of Study Design

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member. The critical appraisal of the evidence enables us to determine whether: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve net health outcomes for patients. An improved net health outcome is one of several considerations in determining whether an item or service is reasonable and necessary.

CMS divides the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the relevance of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention's risks and benefits.

The issues presented here represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has unique methodological aspects.

1. Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes; and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematical assessment of factors related to outcomes.
- Larger sample sizes in studies to help ensure adequate numbers of patients are enrolled to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias)
- Co-interventions or provision of care apart from the intervention under evaluation (confounding)
- Differential assessment of outcome (detection bias)
- Occurrence and reporting of patients who do not complete the study (attrition bias)

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

- Randomized controlled trials
- Non-randomized controlled trials
- Prospective cohort studies
- Retrospective case control studies
- Cross-sectional studies
- Surveillance studies (e.g., using registries or surveys)
- Consecutive case series
- Single case reports

When there are merely associations but not causal relationships between a study's variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study's selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess the evidence.

2. Generalizability of Clinical Evidence to the Medicare Population

The applicability of the results of a study to other populations, settings, treatment regimens, and outcomes assessed is known as external validity. Even well-designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease, and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing, and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study's external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator's lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.

Given the evidence available in the research literature, some degree of generalization about an intervention's potential benefits and harms is invariably required in making coverage decisions for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation), and similarities of the intervention studied to those that would be routinely available in community practice.

A study's selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations because one of the goals of our determination process is to assess net health outcomes. We are interested in the results of changed patient management not just altered management. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention's benefits are clinically significant and durable, rather than marginal or short-lived.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

3. Assessing the Relative Magnitude of Risks and Benefits

Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits. Net health outcomes are one of several considerations in determining whether an item or service is reasonable and necessary. For most determinations, CMS evaluates whether reported benefits translate into improved net health outcomes. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology's benefits and risk of harm to Medicare beneficiaries.

Appendix B
CMS Review Table for Lumbar Artificial Disc Replacement

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results		Methodological Comments
				Intervention group	Control group	
	Study, inclusion/exclusion	**N, age, sex,**				
Blumenthal/2005	Randomized controlled trial – Charite III vs BAK cage fusion, anterior approach. Multiple inclusion/exclusion, discography included.	304 pts 14 centers Randomized 2:1 Mean age = 40 (range 19 – 60) M 157, F=147	24 mos f/u. Composite endpoint of 4 criteria. Others outcomes	Clinical success in 57% of Charite group. Noninferior, p = 0.0001	Clinical success in 46% of control.	
Caspi/2003	Case series Low back pain with/without radicular pain, under age 55.	20 24-50 years 11 men 9 women	f/u 48 months. return to work clinical results rated as fair, good, excellent, poor	Fair = 3, good = 4, excellent = 11, poor = 4. 4 completely disabled, 1 resumed physical labor, others returned to light and sedentary work.	None	
Cinotti/1996	Case series Contraindications included facet degeneration by CT or MRI and disc degeneration adjacent to a fused area and spondylolisthesis. Single surgeon series.	46 Average age 36 (27 – 44 years) 21 M, 25 W	f/u 2 yr min. Clinical results physician rated and patient satisfaction.	Excellent 24% Good 39% Fair 30% Poor 7% 8 of 17 who had unsatisfactory results underwent fusion.	None	
David/2002	Case series Chronic low back pain alone or with sciatica. 72 with previous surgery. MRI, discography.	147 Unknown ages, sex	f/u 5 years minimum. Stauffer Coventry classification	79% excellent or good	None	Abstract only available.

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results		Methodological Comments
				Intervention group	Control group	
	Study, inclusion/exclusion	**N, age, sex,**				
Geisler/2004	Same as Blumenthal/2005	Same as Blumenthal/2005	24 mos f/u Neurologic adverse events	Neurologic adverse events: Investigational group16.6%	Control group 17.2%	Same study group as Blumenthal/2005
Griffith/1994	Case series Diagnosis DDD 65.2%, postnucleotomy syndrome 15%, internal disc derangement 10.9%, others. 31% - 1 prior procedure, 10% - 2 or more prior procedures.	93 139 prosthesis Models I,II,III Age range 25-59 43 male 50 female	f/u range for III 1 to 37 mos. Work status, pain reduction, neurologic weakness, straight leg raising, ability to walk.	Significant proportion experienced pain relief $p < 0.05$. Improvements in pain intensity, walking distance, decreased SLR, weakness, no difference in work status.	None	
Lemaire/1997	Case series Average pre-op pain 6 years.	105 Average age 39.2 years. Range 24-50 years. 68 M, 37 F	f/u average 51 months. Modified Stauffer-Coventry rating scale, work status, radiographic results.	79% good results.	None	
Lemaire/2005	Case series Excluded if obese, prior fusion, suffered from radicular pain, or spondylolisthesis or facet arthrosis	100 Average age at surgery 40, Range 24-51 41 M, 59 F	f/u minimum 10 years. Modified Stauffer-Coventry rating scale, work status, radiographic results.	62% excellent 28% good 10% poor	None	

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results		Methodological Comments
				Intervention group	Control group	
	Study, inclusion/exclusion	**N, age, sex,**				
McAfee/2005	Same as Blumenthal/2005	Same as Blumenthal/2005	24 mos f/u. Radiographs	Lateral flexion/extension films: 13.6% increase from baseline in investigational group. Disc: 82.9% ideal placement.	decreased in control group	Same study group as Blumenthal/2005
Putzier/2005	Case series Indication moderate to severe DDD	53 pts 63 prosthesis Charite I,II and III Avg. age 44, (30 – 59 range). 22 M, 33 F	Average f/u 17 yrs. VAS, ODI, subjective perception of overall outcome, radiographic measurements.	60% spontaneous ankylosis. Re-operation in 11% of patients.	none	
Sott/2000	Case series Discography performed, MRI, radiographs, Bone density, pre-op long standing disabling lumbar pain.	14 31-61 years 8 male 6 female	Mean f/u 48 mos. Good: >75% pain relief, return to work, <= slight physical restriction, no analgesics.	Good: 10/14 Fair: 2/14 Poor: 2/14	None	
Van Ooij/2003	Case report From a series of 500 pts operated on in a single institution.	27 pts Mean age 40 (30 – 67) M 12, F 15	Mean f/u 53 months.	complications	none	

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results		Methodological Comments
				Intervention group	Control group	
	Study, inclusion/exclusion	**N, age, sex,**				
Zeegers/1999	Case series Failed conservative tx. 54% had previous surgery, radiographs, CT, discography, MRI. 29 pts one level, 18 at two levels, 3 had 3 levels.	56 Mean age 43 (24-59 years) 30 F, 20 M	2 year outcomes. Improved pain, Return to work.	Stauffer and Coventry criteria: 70% had good or fair, 65% improved low back pain, 81% return to some work, 83% "no regret."	None	