Appendix A

Evidence Table

Authors/date	Study type	Demographics	Intervention	Results	Comments/Limitations
Sinclair, James, Singer, 1997	RCT	40 patients with fractures of the femoral neck; Mean age in protocol group 74, mean age in control group 75.5 Demographics between both groups were similar	Control group received conventional intra-operative fluid management (n=20), while protocol group received Additional repeated colloid fluid challenges with esophageal Doppler ultrasound to maintain stroke volume (n=20).	Patients in the protocol group also had significantly shorter hospital stays: time spent in an acute hospital bed (10 vs. 18 days), number of days needed before deemed medically fit for discharge (10 vs. 15 days), and total hospital stay (12 vs. 20 days). Mortality rates were similar between both groups.	Small sample size Term “medically fit for discharge” is an arbitrary term difficult to define, many factors (e.g., social factors) may influence this number. Data was reported using median numbers instead of mean numbers Possible bias because treating physician was not blinded to group assignment
McKendry, McGloin, Saberi, Caudwell, Brady, et al.; 2004	RCT	Participants involved in the study were patients undergoing cardiopulmonary bypass surgery Demographics between both	Control group (conventional management, n=85) or protocol group (allocated to optimization of circulatory status, n=89)	Stroke volume, cardiac index, use of colloid were well matched at baseline, but were significantly greater in the protocol group at four hours; Use of inotropes	Small sample size Study only conducted at one center, hard to generalize to other centers Disparity between mean and median results

		groups were similar; average age in control group was 66.7, while average age in protocol group was 65.6		were similar between both groups. In the protocol group, the mean number of days in the ICU was reduced from 3.2 to 2.5 (a 23% reduction), the mean duration of hospital stay was reduced from 13.9 days to 11.4 days (18% reduction), a reduction in median duration of stay from 9 to seven days. Protocol participants showed a trend toward fewer major postoperative complications compared to control group	Data was reported using median numbers instead of mean numbers Possible bias because treating physician was not blinded to group assignment
Wakeling , McFall, Jenkins, Woods, et al; 2005	Single center, RCT	128 consecutive patients undergoing colorectal resection were included in the study; Average age in	Control group (n=64) received conventional management (routine cardiovascular monitoring and central venous pressure monitoring [CVP]); experimental group (n=64) received esophageal Doppler	Median postoperative hospital stay for esophageal Doppler group was 10 days, compared to 11 days for the conventionally managed group (P<	Term “medically fit for discharge” is an arbitrary term difficult to define, many factors (e.g., social factors) may influence this number. Data was reported using

		each group was 69 Demographics between both groups were similar	guided monitoring of additional colloid administration	0.05); Median time to tolerate full diet was 6 days for the Doppler group while 7 days for the control group (P<0.01). Patients in the Doppler guided group were given a significantly greater volumes of intravenous colloid than controls, and the Doppler group achieved higher cardiac outputs and stroke volume at the end of the operation than did the control group. Nine of the patients in the Doppler group experience gastrointestinal morbidity (e.g., infections, renal, etc) compared to 29 in the control group.	median numbers instead of mean numbers Possible bias because treating physician was not blinded to group assignment
Venn, Steele, Richardson, et al. 2002	RCT	90 participants undergoing repair of femoral	Three groups: conventional operative fluid management (CON,	Greater fluid challenges occurred in the CVP group as	Small sample size Arbitrary definition of

		fractures were involved in the study. Average age of participant in control group was 84.5; the average age of participant in protocol group was 82. Base-line demographics similar between groups	n=29), and two groups receiving additional repeated colloid fluid challenges guided by central venous pressure (CVP, n=31) or esophageal Doppler ultrasonography (DOP, n=30).	well as the DOP group, compared to the CON group. As a result of this, both groups (CVP and DOP) had fewer episodes of intraoperative hypotension (P<0.048). Time to be deemed medical fit for discharge was also shorter in the DOP group (8 vs. 14 days) and the CVP group (10 vs. 14 days) compared to the conventional group. Study failed to reveal any differences in acute orthopedic hospital stay days, total number of hospital days, or mortality between the 3 groups.	"medically fit for discharge" Possible bias because treating physician was not blinded to group assignment
Gan , Soppitt, Maroof, El-Moalem, Robertson , Moretti, Dwane, Glass; 2002.	RCT	100 patients with ASA physical status I, II, were undergoing major elective	49 patients in each group Protocol group (boluses of fluid were guided by an algorithm depending on	Protocol group had a significantly higher stroke volume and cardiac output compared to the	Unable to blind anesthesiologist as to treatment group, so unable to eliminate bias

		surgery, urologic, or gynecologic surgery with an anticipated blood loss of greater than 500 ml. Average age in control group 59; average age in protocol group was 56	the Doppler estimations of stroke volume and corrected flow time) or control group (anesthesia care provider was not given results of Doppler reading, but instead relied on monitoring change in heart rate, systolic blood pressure, central venous pressure, and urine output)	control group, and a shorter hospital stay (5 +/- 3 vs. 7 +/- 3 days [mean +/- SD], 6 vs. 7 days [median] respectively (P=0.03). Also fewer protocol patients experienced severe post-operative nausea and vomiting (P=0.01), and were able to tolerate an oral solid regimen earlier than the control group.	Patients in the protocol group received larger volumes of hetastarch compared to control group; the differences between groups could be attributed to the differences in type of fluid administered. Results of study may not applicable to Medicare-age population since mean age of both groups not 65 or greater Possible bias because treating physician was not blinded to group assignment
Mythen, Webb; 1995	Prospective randomized open study	60 ASA grade III patients undergoing elective surgery for coronary artery bypass graft or single heart valve replacement Average age in control group was 64; average age in protocol group was 63	Patients were randomized to either the control group (standard practice, n=30) or to the protocol group (standard practice plus 200 ml boluses of 6% hydroxyethyl starch solution to obtain maximum stroke volume estimated by esophageal Doppler system n=30).	The incidence of gut mucosal hypoperfusion was significantly reduced in the protocol group compared to the control group (7% vs. 56%, P< 0.01); The number of complications developed (0 vs. 6 days, P=0.01) was lower in the protocol group;	Small sample sizes Possible bias because treating physician was not blinded to group assignment

				The mean number of days spent in the hospital (6.4 vs. 10.1, P=0.011), and mean number of days spent in the ICU (1 vs. 1.7, P=0.023) was also lower in protocol group compared to control group.	
Noblett, Snowden, Shenton, Horgan, 2006	Double-blinded RCT	103 patients undergoing elective colorectal resection involved in the study Average age in control group 67; average age in protocol group 62.	Patients were randomized to either the control group (n= 52) which consisted of standard treatment-peri-operative fluid at the discretion of the anesthesiologist, or randomized to the protocol group (n=51) in which additional colloid boluses were based on Doppler assessment.	Patients in the protocol group had reduced time to fitness for discharge (median 6 vs. 9 days, P=0.003), and actual discharge (7 vs. 9, P=0.005) days. No difference in lower gastrointestinal function assessed by return of bowel activity were noted, but the study did reveal that the protocol group was able to tolerate diet significantly earlier than the control group (P=0.029).	No mention of how randomization process was carried out Possible bias because treating physician was not blinded to group assignment

				Also intermediate or major complications were less frequent in the in the Doppler-guided group (1 vs. 8, P=0.043), including unplanned admission to the critical care unit (0 vs.6, P=0.012).	
Conway, Mayall, Abdul-Latif, Gilligan, Tackaberry; 2002	RCT	57 patients undergoing major bowel surgery were included in the study Average age in control group was 67.5; average age in protocol group was 66.5	Patients were randomized to either a control group (n=28) which used standard care protocol (intra-operative fluid at the discretion of a non-investigating anesthesiologist), or randomized to the protocol group (n=29) (standard care along with fluid challenges guided by esophageal Doppler monitoring).	Protocol group received more intra-operative colloid (mean 28 vs. 14.7, P=0.02); Protocol group had higher cardiac output than the control group (0.87 vs.0.31-0.1.43, P=0.003), and less morbidity (5 control participants required post-operative critical care admission vs. none in the protocol group); There were no significant differences in hospital length of stay between both groups	Small sample size Possible bias because treating physician was not blinded to group assignment

Chytra I, Pradl R, Bosman R, Pelnar P, Kasal, Zidkova A. 2007	RCT, but non-blinded	All patients involved in the study were multiple trauma patients with estimated blood loss of more than 2,000 ml	80 subjects were in Doppler group (mean age 33) and 82 subjects in the control group which received standard management (mean age 40).	After the 12-hour study period, blood lactate levels in the Doppler group was statistically lower (2.92 mmol/l versus 3.22 mmol/l, p=0.003) compared to control group; Also rate of administration of norepinephrine was lower in the Doppler group compared to control group (RR =0.56, p=0.018). The difference in lactate levels between the Doppler and control group change very little after 24 hours of ICU stay. Though no differences in SOFA levels were noted during ICU stay, and no organ dysfunction was noted between the two groups, fewer infectious complications were noted in the Doppler group compared to the control (RR=0.5491, p=0.032). There was a noted reduction in	Study limitations include a relatively small size cohort, non-blinding, the study conducted at only one center, and younger age population (average age in experimental group was 33, the average age in control group was 40) making it difficult to generalize to the Medicare population

				median duration of hospital stay (14 days versus 17.5 days, p=0.045), as well as reduction in ICU days (7 versus 8.5, p=0.031).	

APPENDIX B

General Methodological Principles of Study Design
(Section VI of the Proposed Decision Memorandum)

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 is reasonable and necessary. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve health outcomes for patients.

We divide the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the generalizability 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 potential risks and benefits.

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

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 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 (performance bias).
- 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, in general, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The design, conduct and analysis of trials are important factors as well. For example, a well designed and conducted observational study with a large sample size may provide stronger evidence than a poorly designed and conducted randomized controlled trial with a small sample size. 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 selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess and consider the evidence.

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 determinations 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. One of the goals of our determination process is to assess health outcomes. 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. Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits.

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.

Assessing the Relative Magnitude of Risks and Benefits

Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits. Health outcomes are one of several considerations in determining whether an item or service is reasonable and necessary. 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.