APPENDIX A

Evidence Table

Studies evaluating the intermittent use of nesiritide for chronic heart failure

Authors/Year	Study Design	Demographics	Intervention, Outcome Measures, Instruments	Results	Methodological Comments (Limitations)
Yancy, Saltzberg, Berkowitz, Bertolet, Vijayaraghavan, Oren, Burnham, Walker, Horton, Silver; 2004 FUSION I Trial	A multicenter, open-label, pilot study that randomly assigned subjects to usual treatment only compared to usual treatment plus weekly infusions of nesiritide in a 1:1:1 ratio. Intention to treat analysis was followed.	Eligible patients were 18 years and older, had a NYHA class III or IV, and had 2 or more hospital admissions for ADHF within the preceding 12 months. Study involved 210 subjects (sample size was determined empirically). Study involved the use of a prospective RAS (Risk Assessment Score) based on known prognostic factors. 69 subjects received	Subjects were assigned to 1 of 3 treatment groups: (1) usual care, (2) usual care plus 0.005 µg/kg/min of nesiritide given for 4-6 hours preceded by a bolus of 1.0µg/kg bolus, (3) usual treatment plus 0.01 µg/kg/min of nesiritide given for 4-6 hours preceded by a 2.0 µg/kg bolus. Safety and tolerability were the primary endpoints, assessed by adverse events, serious adverse events, discontinuation in the study, lab assessment and vital signs. The Minnesota Living with Heart Failure	At baseline the only significant difference between treatment groups was the increased prevalence of atrial fibrillation in the usual treatment group. A total of 1,645 nesiritide infusions were administered. All treatment groups had a similar frequency of adverse events, and experienced improved quality of life. Although there was no statistically significant differences in outcomes for the 3 treatment groups, prospectively defined high risk sub-groups demonstrated	Small sample size, (sample size number not explained- no effect size stated, Insufficiently powered to detect statistical difference. Open label study is prone to investigator bias. Study had short duration. Definition of "usual care" was left to the discretion of the investigator. High-risk sub-group was defined prospectively.

		usual care, 72 subjects received usual care plus 0.005 µg/kg/min of nesiritide, and 69 0.01 µg/kg/min.	Questionnaire was also used. All-cause deaths and hospitalizations, Deaths, All cause hospitalizations, Days alive and out of hospital, and RAS scores were measured.	significant decreases in cardiovascular events. There were no statistically significant differences in deaths or hospitalizations. Subjects receiving nesiritide showed trends for more days alive and out of the hospital compared to subjects receiving usual care.	
Sheikh-Taha; 2005	A single center, nonrandomized, open label prospective study.	All subjects were 18 years and older, had NYHA class III or IV. Subjects receiving maximum oral therapy with diuretics, ACE inhibitors, ARBS, hydralazine, nitrates, β-blockers and spironolactone. Also patients intolerant of or refractory to intermittent IV inotropic therapy with	At each visit, subjects received a bolus of 2µg/kg of nesiritide, followed by 0.01 µg/kg/min of nesiritide, given over a four to six hour period. Patients also received a 4-6 hour infusion of iv dobutamine 4-6 µg/kg/min. or milrinone followed by a maintenance infusion of 0.1750-.375	At the beginning of the study, 9 subjects were in the NYHA class III, and 2 were in class IV. After 3 months of treatment, 7 patients remained in class III, and 4 patients moved to class II; no subject remained in class IV. Of the 11 subjects, 6 had improvement in NYHA class, 5 remained in the same class, and 0 regressed (p=1.0).	Open label research design prone to investigator bias Did not follow intention to treat protocol. Small sample size. Lack of randomization. Short follow-up period. No controls.

		dobutamine or milrinone. 14 patients were initially recruited for the study, but 11 remained the study (7 males, 4 females); the mean SD age was 69 +/-8 years.	µg/kg/min. Nesiritide doses were were adjusted downward in patients with renal insufficiency. Subjects were followed for three months.	The number of hospital admissions due to exacerbation of HF did decrease (11 vs 5; p=0.0253), and the number of visits to the HF clinic declined from 5.6 per month to 4 per month (p=0.0749). The intermittent administration of nesiritide along with other drugs was well tolerated by most subjects.	No effect size stated. Confounding effects of variables not adjusted for.
Josephson, Barnett; 2004	Case study	36 subjects, all with decompensated heart failure refractory to standard therapy. 475 infusions of nesiritide were administered.	Subjects received 2mcg/kg bolus, followed by 0.01 µg/kg/min of nesiritide, given over a 4 to 6 hour period.	12 weeks post infusion, 71% of patients were alive and had no hospitalization compared to 52% in the FUSION I trial. Mean hospital days for nesiritide pts 1 yr prior to nesiritide was 9 days. After treatment with nesiritide, the mean number of hospital	No comparison group. Small sample size. QOL measures not identified. Only 12-wk mortality reported. Though study divided group into high risk and low risk groups, it only

				days was 6.5 days. Mortality rate at 12 wks was 5.7 for high risk pts receiving nesiritide, compared to the FUSION I trial.	reported numbers for high-risk subjects, but not for low-risk subjects.
Peacock, Holland, Gyarmathy, et al. 2005	Multi-center, randomized, double-blind, placebo-controlled pilot study	237 emergency department/observation unit (ED/OU) patients with decompensated heart failure	Objective outcome measures included initial admission, length of stay, and inpatient re-hospitalization through 30 days. Two subjective self-assessment outcome measures were used to evaluate change in dyspnea during the study; a visual analogue scale (VAS) and a 7-point ordinal scale. Dyspnea was assessed during the first 12 hours, and also at drug discontinuation. Specific inclusion/exclusion criteria were noted, intent to treat analysis was followed, and safety measures were	Asymptomatic hypotension was higher in the treatment group than in the control group (10% vs. 3%). Nesiritide-treated subjects were more likely than placebo-treated subjects to be terminated due to adverse events (12% vs. 4%). By study day 30, six (3%) subjects were reported to have died; five (4%) were originally reported in the treatment group and one (1%) in the control group (p=0.213 [Fisher's Exact test]). Nesiritide-treated subjects had an 11%	Several mos post-publication, Scios announced two additional deaths in the nesiritide treatment group, not counted in the original report. This raises the number of deaths in the nesiritide-treated group to seven, compared to only one in the control group. Confidence intervals not included. Also, the study noted that using Fisher's Exact test statistical method, the comparable death rate between the two groups

| | | | | also defined. | reduction in hospitalization from the ED during the index period compared to control subjects (55% control, 49% nesiritide, p=0.436), and fewer nesiritide-treated subjects (30%) were hospitalized from the ED with heart failure than from the control group (38%, p=0.220).

Median initial hospital LOS was similar between the two groups, but after discharge from the index visit, fewer nesiritide-treated subjects (19%) were admitted to the hospital compared to the control group (24% p=0.43 [Fisher's Exact test].

Among patients hospitalized during the index visit, 10% of nesiritide-treated subjects were re- | was not statistically significant (p=0.213). But if measures of association are used, (Relative Risk), the nesiritide group had almost a five-fold increase in death compared to the control group (RR=4.74).

Though VAS is a validated test for dyspnea measurement, specific information on the 7-point ordinal score measure was not provided

Other limitations noted in the study by the authors include the lack of prospectively defined primary endpoints, a high number of subjects were in NYHA class I or II or had no prior history of |
|---|---|---|---|---|---|---|

				hospitalized through study day 30 compared to 23% of the control patients (p=0.058). Ordinal scale dyspnea scores were similar between treatment groups through the 12-hour collection period, as well as at drug discontinuation. Mean dyspnea scores at baseline on the VAS were the same for the two treatment groups (38.2 for the nesiritide group and 42.0 for the control group).	heart failure enrollment, and the observation unit environment may confound length of stay data.

APPENDIX B

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.