Evidence Table Appendix A

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results	Comments
Gorecka, 1997	Randomized control trial, pts referred to 9 regional LTOT centers in Poland, no blinding Inclusion: FEV1/FVC ratio <70% predicted PaO_2measures 56-65 mmHg Exclusion: Diseases of other organ systems that may impact survival Avg observation time: 40.9 mos	n= 135 control group= 67 LTOT group= 68 age 40-80 mean age: 61.2 76% male *participants in the control group cont to receive usual COPD care including bronchodilators, diuretics, steroids, antibiotics at the discretion of the physician *participants in the LTOT group received oxygen to maintain PaO_2 above 65mmHg prescribed for at least 17hours per day	Mortality	70 patients died during the observation period, 32 controls and 38 in the LTOT group; the majority due to a progression of COPD Cumulative survival rate: Yr 1: 88% Yr 2: 77% Yr 3: 66% Cox regression analysis: No difference in survival between control and LTOT groups Survivors were younger, had better lung function, and higher BMI	
Hjalmarsen, 1999	Retrospective study	n =124 Group I; n=76	Mortality Subgroup	Group I survival	Patients with PaO_2

	Inclusion criteria: Group I $PaO_2 \leq$ 7.3 kPa (55 mmHg) or Group II PaO_2 up to 8.0 kPa (60 mmHg) if coexisting polycythemia or cor pulmonale Oxygen use- at least 15h per day	Group II; n= 48 mean age: 68	analysis based on lung function, gender, hospitalization	2 Yr: 73% 5 Yr: 50% Group II survival 2 Yr: 78% 5 Yr: 40% Male survival: 2 Yr: 56% 5 Yr: 30% Female survival: 2 Yr: 83% 5 Yr: 60% Group II: PaCO2 and FVC showed significant impact on survival Lower survival in patients treated in the general hospital setting **No statistically significant survival benefit when comparing LTOT users Group I and II**	of 60mmHg only comprised 39% of the population being studied
Sliwinski, 1992	Prospective cohort study Inclusion: consecutive referrals for assessment of	n=46 Responders n=7 Non responders n=39 83% male	Acute effect of oxygen on pulmonary hemodynamics Mortality	Avg oxygen use 14.6h/day 2 Yr Survival Rate	Small number of participants in the responder group; no

	eligibility for LTOT $PaO_2 \leq$ 55mmHg or $PaO_2$56-65 mmHg if accompanied by radiologic signs of pulmonary HTN, signs of RVH, or elevated hematocrit Pts underwent a 4wk probationary period to ensure they cont. to meet inclusion criteria Exclusion: any condition that may influence survival such as HTN, ischemic heart disease, left heart failure, cirrhosis, renal failure, diabetes, or malignancy Patients were divided into Responders and Nonresponders, based on changes in pulmonary artery pressure in response to LTOT Treatment			Responders : 69% Non-Responders: 57% Hospital Admissions: Responders: 0.8 Non-Responders: 1.4	information regarding the specific number of pts with PaO_2 between 56-65 mmHg

	period: 2 yrs or until death				
Sandek, 2001	Prospective cohort study Inclusion: irreversible airflow obstruction 2 PaO_2measures < 7.3 kPa or 7.3-7.9kPa with chronic right heart failure Oxygen prescription greater than 15h per day for 6 months	N=14 Male 79% Mean age: 69	Changes in pulmonary physiology	No significant changes to pulmonary physiology noted with LTOT use.	Very small sample size

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.