Appendices

Author/ Year	Study Design	Demographics	Intervention, outcome measures; instruments	Results	Comments
	Study, inclusion/exclusion	N, age, sex,			
Donohue 2006	Nebulized Treatment System/Outpatient Setting Adult Patients/COPD Randomized, multictr, double blind trial Tx duration- 6 wks 2 doses of levalbuterol (0.63 or 1.25mg)	n = 209 mean age=65 Moderate to severe COPD	Primary outcome: % change in FEV_1 over 0/2/6 wks Secondary outcomes: Disease control measurements(COPD exacerbations, rescue med use, symptom control, adverse events)	All treatment groups demonstrated improvement in FEV_1 COPD Exacerbations: Placebo- 12.7% Leval(.63)-11.3% Leval(1.25)-18.4% Albut(2.5)-21.2% Adverse Events Withdrawals: Placebo- 1.8% Leval(.63)-13.2% Leval(1.25)-8.2% Albut(2.5)-23.1% *most common reason COPD exacerbation	Rescue Med Usage: Decreased in both levalbut groups, *results not shown and did not meet statistical significance Symptom control: authors notes levalbut groups note overall symptom improvement but results did not meet statistical significance Only 171/257 enrolled subjects completed No trend favoring levalbuterol in adverse events No 1.25 mg albuterol arm to compare with 0.63 mg levalbuterol arm Sepracor authors
Truitt 2003	Nebulized Treatment System/Hospital Setting Adult	n = 231 COPD n=177 Mean age=58	Primary outcome: # of nebulized treatments required Secondary outcomes:	Levalbut group required 38% less nebulizations, 29% fewer days of nebulization therapy	*Authors do not correct for baseline differences in dosage frequency which would likely account for need for fewer neb treatments (i.e. levalbut doses q8 vs. racemic albut q4-6h)

	Patients/COPD or Asthma Retrospective Chart Review		change in FEV_1; duration of hospital stay, disposition, cost	Length of hospital stay: results not statistically significant	
Datta 2003	Nebulized Treatment System/Outpatient Setting Adult Patients/COPD Randomized, double-blind, placebo controlled, crossover 4 groups: levalbu(1.25mg), racemic albut, racemic albut + ipratroprium, placebo	n = 30 Mean age=69	Primary outcome: change in FEV_1 Secondary outcomes: FVC, HR, oxy sat, hand tremors	@ 0.5/1h post treatment- all treatments group had increased FEV_1 vs placebo *only albut + ipratropium group showed sustained improvement in FEV_1 vs placebo @ 2-6h post treatment *No significant difference in secondary outcome measures btwn tx groups	Small sample size
Lotvall 2001	Dosimeter System/Outpatient Setting Adult Patients/Asthma Randomized, double-blind, placebo controlled, cross over	n = 20 Mean age=50	Primary outcome: Change in FEV_1 Secondary outcomes: Monitoring of potassium levels, HR, and drug plasma concentrations	*Both levalbut and racemic albut improved FEV_1 vs placebo *Mild increases in HR and potassium were experienced by both groups but not assoc with serious adverse events or requiring study	Small sample size

				withdrawal	
Nelson 1998	Randomized Double-blind Parallel group Multicenter Lev or rac tid x 4 wks Serial PFTs at 0,2 4 wks	n = 362 mean age overall 36 Rac 3 yrs > Lev	Primary endpoint: Change in FEV 1 at 4 weeks compared to placebo Secondary AUC and use of rescue Rac	Mean differences 10% Use of rescue: All active arms better than placebo, all active arms > 95% Reported mean rescue puffs/day 0.63 Lev 3.5 1.25 Lev 2.7 1.25 Rac 3.6 2.5 Rac 3.8 Placebo 4.9	Only statistically significant diff with first dose Says intent-to-treat, but 424 initially enrolled Statistical discussion is about comparison of active arms to placebo, not between active arms. AE trend favors Rac AE withdrawals 0.63 Lev 3 (3 serious) 1.25 Lev 8 (1) 1.25 Rac 2 (1) 2.5 Rac 4 (1) Placebo 5 (2) Figures are selective about what data are displayed
Schreck 2005	ED, community based Retrospective observ case review	n = 786+186 Age >1	Admission rates	4.7 Lev 15.1 Rac	Retrospective, unblinded, uncontrolled "no cause and effect conclusions can be drawn…"
Nowak 2006	Multicenter Randomized Double blind Prednisone plus Lev or Rac	n = 627 Adults	Primary endpoint: Time to meet discharge criteria	No difference In time to meet discharge FEV1 improvement Lev 0.50 ± 0.47L Rac 0.43 ± 0.37 L (p = 0.02) No diff in hosp rates or relapses	Minimal clinical signif of 0.07 L since nl FEV1 is around 5.00 L

Pleskow 2004	Post hoc pairwise analysis of the 1998 Nelson study	See Nelson	See Nelson	Active treatment better that placebo except in one of the Rac groups (1.25 mg)	
Thompson 2004	Adult EMS prehospital Open label Before and after Prospective	n = 196/298 Mean age 68	Primary endpoint: Change in PF after one treatment	No difference 19.7% v 20.4%	
Nowak 2004	Prospective Open label Nonrandomized Pilot Sequential cohorts on 12-14 Tx q 20 min x 3 Lev 0.63, 1.25, 2.5, 3.75, or 5.0 Rac 2.5 or 5.0	NA	Efficacy and tolerability	Incomplete data in abstract. FEV1 change after 3 doses 1.25 Lev 0.9L 0.63 Lev 0.6 L Rac ? dose 0.5 L Lev > 1.25 had no further improvement	Only reports some groups
Qureshi 2005	Prospective Double blind Randomized ED setting	n = 129 Age 2-14	Superiority trial Asthma score and FEV1	No difference in clinical improvement More nausea with Rac. More tremulousness, headache, lightheadedness with Lev.	Moderate asthma

Scott 2003	Retrospective Lev 0.63 or Rac 2.5	Tertiary hospital Age > 18 Mean 65-68 n = 35/group	Mean heart rate change	Day 1: 1 BPM difference Day 3: 2.7 BPM Authors say not clin signif even at highest confidence interval	Retrospective Small sample sizes
Combination Inhalation Study Group	Randomized safety and efficacy trial. Nebulized treatment/outpatient setting/Adult Patients/COPD Comparison of combination ipratropium and albuterol versus single drug therapy, three times daily administration Tx duration- 85 days	n=652 mean age= 64.9 moderate to severe COPD	Primary outcomes: FEV1 response, PEFR, quality of life measures, adverse events.	All treatment groups had significant bronchodilator response of ≥ 15% from baseline on each of 4 test days Response to combination therapy was greater than either drug singly Morning PEFR did not change over the 12 weeks; comparable across treatment groups. Evening PEFR were significantly greater in the combination group versus the albuterol group Adverse events, at least 1: 56.8%- combination group	

				52.3%-ipatropium 57.4%-albuterol Most frequent event- worsening of lower respiratory tract symptoms	
Tashkin 1996	Nebulized Treatment System/Outpatient Setting/Adult Patients/COPD Randomized double-blind trial Comparison ipratropium versus ipratropium and metaproterenol Tx duration- 85 days	n=213	Primary outcomes: FEV1, FVC, PEFR Secondary outcomes: symptom control	FEV1 and FVC AUC were significantly higher for the combination group vs. ipratropium alone. No significant effects were noted on morning PEFR, subjective symptoms, or quality of life measures	
Gross 1998	Nebulized Treatment System/Outpatient Setting/Adult Patients/COPD Randomized double-blind, cross-over 3 treatment groups: Albuterol Ipratropium combination	n=863, randomized patients n=574, patients completing the trial mean age 66.3	Primary outcomes: change in FEV1 Secondary outcomes: Safety variables	FEV1 mean percent change with combination therapy greater than that achieved with either therapy alone Adverse events: similar adverse event rates across treatment groups, most common included respiratory issues	*large number of patients withdrew prematurely

Baumgartner et al 2007	Nebulized Arformoterol in Patients with COPD 3 AF arms, 1 salmeterol arm Placebo arm 12 wk multicenter double blind double dummy	917 enrolled 724 randomized n=717 receiving medication mean age 62.9 58% male	Mean % change morning trough FEV1 and AUC FEV1 over 12 weeks	Arformoterol and salmeterol more effective than placebo	

APPENDIX B

General Methodological Principles of Study Design

(Section VI of the 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.