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 health outcomes for patients. An improved 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 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. 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 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.

Separate page

<table>
<tr><th rowspan="2">Author/ Year</th><th rowspan="2">Study Design</th><th rowspan="2">Demographics</th><th rowspan="2">Outcome measures</th><th colspan="2">Results</th><th rowspan="2">Methodological Comments</th></tr>
<tr><th>Intervention group</th><th>Control group</th></tr>
<tr>
<td>Venstrom J, 2003</td>
<td>Retrospective, controlled, multicenter cohort

Transplant cohort: patients on UNOS/OPTN waiting list for a pancreas transplant (PA, PAK, SPK) from Jan 1, 1995 – Dec 31, 2000 who did receive a transplant

Control cohort: patients on wait list who did not receive a transplant

Analysis stratified by transplant type

Exclusions: on wait list for multi-organ transplant other than SPK; creatinine >2 mg/dL at time of listing; listed for PA but eventually received SPK</td>
<td>Total N= 11,572

Transplanted N= 6595

PA= 378
PAK= 838
SPK= 5379

For the PA group:

40% men
10% ≥50 yrs old</td>
<td>Unadjusted wait list and post-transplant patient survival rates

Mortality risk (hazard ratio-- average risk for PA transplant patients compared to patients on wait list for comparable amount of time) assessed for 3 clinically distinct time periods in the transplant group:

0-90 days
91-365 days
366-1460 days</td>
<td>Results presented for PA patients only:

Transplant N= 361

Baseline characteristics of transplant and wait list patients statistically indistinguishable

<table>
<tr><th>Time period (days)</th><th>Mortality Risk</th><th>95% CI</th><th>P value</th></tr>
<tr><td>0-90</td><td>2.27</td><td>0.84-6.13</td><td>0.11</td></tr>
<tr><td>91-365</td><td>0.99</td><td>0.41-2.39</td><td>0.99</td></tr>
<tr><td>366-1460</td><td>1.70</td><td>0.97-2.98</td><td>0.06</td></tr>
<tr><td>Overall (4 yrs)</td><td>1.57</td><td>0.98-2.53</td><td>0.06</td></tr>
</table>

<table>
<tr><th>Patient survival</th><th>PA</th></tr>
<tr><td>At 1 yr</td><td>96.5%</td></tr>
<tr><td>At 4 yr</td><td>85.2%</td></tr>
</table>

Mortality risk factors: donor death due to stroke; any complications post-transplant (includes pancreatitis, abscess, or anastomotic leak)</td>
<td>Results presented for PA patients only:

Wait list N= 311

<table>
<tr><th>Patient survival</th><th>Wait list</th></tr>
<tr><td>At 1 yr</td><td>97.6%</td></tr>
<tr><td>At 4 yr</td><td>92.1%</td></tr>
</table>
</td>
<td>Retrospective, observational nature of trial and small n are a negative.

Use of a relevant control cohort is a plus.

Uncertain if target population was studied-- - presume every patient added to UNOS/OPTN wait list met ADA criteria for PA

Assessed patient survival/mortality risk.

Short duration of study (i.e., 4 yrs post-transplant or on the wait list).

Majority of study population was younger than 65 yrs.

lost to follow-up and # of cross-overs not reported.</td>
</tr>
<tr>
<td>Gruessner R, 2004</td>
<td>Retrospective, controlled, multicenter cohort

Transplant cohort: patients on UNOS/IPTR waiting list for a pancreas transplant (PA, PAK, SPK) from Jan 1, 1995 – May 31, 2003 who did receive a transplant

Control cohort: patients on wait list who did not receive a transplant

Analysis stratified by transplant type

Patient with multiple listings at different transplant centers and/or who changed transplant center was counted only</td>
<td>For patients on the PA wait list:

Total N= 1207

42% men
38.8+/- 9.2 yrs
8.3% retransplant

Transplanted N= 647</td>
<td>Unadjusted wait list and post-transplant patient survival rates

Mortality hazard ratio assessed for 3 clinically distinct time periods in the transplant group:

0-90 days
91-365 days
366 days on</td>
<td>Results presented for PA patients only:

Transplant N= 647

Baseline characteristics of transplant v. wait list patients not stated

Hazard ratio for Jan 1, 1995 – May 31, 2003 database:

<table>
<tr><th>Time period (days)</th><th>Mortality</th><th>95% CI</th><th>P value</th></tr>
<tr><td>0-90</td><td>4.25</td><td>1.68-10.76</td><td>0.002</td></tr>
<tr><td>91-365</td><td>1.72</td><td>0.82-3.61</td><td>0.15</td></tr>
<tr><td>366 on</td><td>0.15</td><td>0.08-0.29</td><td>0.0001</td></tr>
<tr><td>Overall</td><td>0.66</td><td>0.39-1.12</td><td>0.12</td></tr>
</table>

Hazard ratio for Jan 1, 1995 – Dec 31, 2000 subset of database:

<table>
<tr><th>Time period (days)</th><th>Mortality</th><th>95% CI</th><th>P value</th></tr>
</table>
</td>
<td>N= 1207

Baseline characteristics of transplant v. wait list patients not stated</td>
<td>Retrospective, observational nature of trial is a negative.

Use of a relevant control cohort is a plus.

Uncertain if target population was studied-- - presume every patient added to UNOS/OPTN wait list met ADA criteria for PA

Assessed patient survival/mortality risk.

Majority of study population most likely younger than 65 yrs.

8% of patients were to receive a retransplant—</td>
</tr>
</table>

	once A patient relisted was counted only once (and all wait times summed) Exclusions: change in status from SPK/PAK/PA to KTA; listed for PA but eventually received SPK; retransplant patients			0-90: 6.40 (1.95-21.1) 0.002 91-365: 2.90 (1.1-7.63) 0.03 366 on: 0.29 (0.15-0.57) 0.0003 Overall: 1.45 (0.83-2.55) 0.19 **Patient survival** / **PA**: At 1 yr 97%; At 4 yr 90.5% Mortality risk factors: time of listing (1995-1998); use of depleting antibody induction therapy; graft failure	**Patient survival** / **Wait list**: At 1 yr 96.6%; At 4 yr 87.3% Mortality risk factors: time of listing (1995-1998); non-Caucasian (v. Caucasian)	i.e., presumably these patients were not excluded from cohort prior to analysis. # lost to follow up not reported.

Key

CsA—cyclosporine A
TAC—tacrolimus
DM—Diabetes mellitus
Pop—population
N/A—not applicable; not available
PA—pancreas transplant alone
SPK—simultaneous pancreas-kidney transplant
PAK—pancreas after kidney transplant
CI—confidence interval
KTA—kidney transplant alone