

# Acute Dialysis Quality Initiative

## 2<sup>nd</sup> International Consensus Conference

### Workgroup 2

#### Clinical Endpoints (Effectiveness) in Randomized Clinical Trials in Acute Renal Failure

*Philipp G.H. Metnitz*

*Paul M. Palevsky\**

*Pasquale Piccinni*

*Christophe Vinsonneau*

### Introduction

The selection and definition of outcome measures (endpoints) is a critical factor for the successful execution of clinical trials. An outcome is defined as either a measurement (i.e., serum creatinine) or an event (i.e., death, need for dialysis) that is potentially modifiable by a defined intervention. Several criteria must be considered in the selection of outcome measures including clinical importance, responsiveness to the intervention, precision of their definition, accuracy of measurement, and completeness of ascertainment. Since multiple outcomes may be affected by a single intervention, a hierarchical ranking is required.

It is critical that the primary outcome be prospectively identified. This outcome is the measure by which the success or failure of the therapeutic intervention will be judged and the predicted magnitude of the intervention on this endpoint is the primary factor determining the number of patients required for a clinical trial.

Secondary endpoints are outcomes that may be related to the primary endpoint (for example Kaplan-Meier survival in a study where 30-day all cause mortality is the primary endpoint) or may provide an independent indication of the therapeutic benefit of the intervention (for example, recovery of renal function in a study where 30-day all cause mortality is the primary endpoint). In general, however, clinically or statistically significant differences in a secondary outcome, in the absence of important changes in the primary outcome, will not be interpreted as strong evidence of a therapeutic benefit. Outcomes that are more exploratory in nature may be considered as additional (tertiary) endpoints.

In establishing this hierarchical ranking for clinical trials on the prevention or treatment of acute renal failure several key questions should be considered. Is the outcome causally related to acute renal failure? Is it clinically relevant? Can it be accurately determined? Has it been precisely defined? Has its clinical validity been established? Will it provide adequate discrimination between the experimental groups? Will the outcome be affected by its measurement?

The selection and hierarchical ranking of endpoints frequently involves tradeoffs. For example, variations in the definition of acute renal failure alter its sensitivity and specificity as an endpoint. Increasing the specificity of the definition (e.g. a 50% versus a 25% increase in the serum creatinine) will decrease the event rate and increase the number of patients required for a trial. The definition of mortality provides another example. Mortality in patients with acute renal failure is strongly influenced by factors other than either the renal failure or the therapeutic intervention. While all-cause mortality may lack a degree of specificity, it is objective and can be easily and accurately determined. In contrast, cause-specific mortality is difficult to define and introduces a need for subjective adjudication.

In prior studies of both prevention and treatment of acute renal failure the selection and definition of endpoints has varied widely. As a result, comparison between studies has been difficult. In the remainder of this article we will review the selection and definition of specific endpoints appropriate to large-scale clinical trials of acute renal failure. Intermediate (surrogate) endpoints, that may be more appropriate for smaller scale, preliminary or pilot studies, which are based on changes in physiologic parameters that are biologically closer to the disease process, are discussed elsewhere in this issue.

While the most appropriate endpoints for an individual study are dependent upon the specific hypothesis and study design, the primary endpoint in prevention trials needs to be based on an objective assessment of renal function (Table 1), while mortality and renal functional assessment may serve as the primary endpoints in studies of established acute renal failure (Table 2).

## **Assessment of Renal Function**

### ***Development of Acute Renal Failure***

A precise definition of ARF is required in order to use its development as an endpoint in clinical trials. Published studies have utilized a wide variety of definitions, based primarily on absolute or relative changes in the serum creatinine concentration. While it is accepted that variations in the definition of ARF alter the sensitivity and specificity of this endpoint, good data does not exist to correlate variations in the definition of ARF with morbidity or mortality.

### ***Measurement of Renal Function***

As an alternative to a categorical definition of acute renal failure, a continuous variable such as serum creatinine, blood urea nitrogen or determination of glomerular filtration rate may be used to assess change in renal function. Their use as clinical outcomes requires strict definition of the timing of measurement (e.g., change in serum creatinine at 48 hours, peak serum creatinine within 7 days, etc.).

### ***Requirement for Renal Replacement Therapy***

There is no consensus regarding the optimal timing for initiation of renal replacement therapy or specific criteria for the discontinuation of therapy. These issues need to be prospectively evaluated in clinical trials. If the initiation or

discontinuation of renal support is used as a clinical endpoint, strict criteria for these clinical decisions must be defined as part of the study design.

### ***Recovery of renal function***

Recovery of renal function can be defined as the lack of need for continued renal replacement therapy. This can be defined based on a GFR in excess of 15-20 mL/minute. Based on published studies, the majority of recovery of renal function occurs within 30 days of onset [1-4]. Because of the long-term implications for development of chronic progressive kidney disease, we believe it is appropriate to stratify recovery of renal function as complete and partial. There is no clear consensus, however as to what degree of recovery of renal function should be defined as “complete”. Thresholds of a serum creatinine that is no more than 0.5 mg/dL or 1.0 mg/dL greater than baseline may be reasonably selected as definitions of “complete” recovery. Alternatively, a threshold based on measured GFR could also be applied. Partial recovery is defined as a serum creatinine greater than baseline by the defined threshold, but not dialysis dependent.

### **Assessment of Survival / Mortality**

Patient survival (or its reciprocal, mortality) has been commonly used as the primary endpoint in clinical trials of renal replacement therapy in acute renal failure. The definition for timing of mortality, however, has varied widely between studies, including survival to hospital discharge [2,5], ICU-survival [5], and survival for a fixed time-period after discontinuation of renal replacement therapy [3,4]. Decisions regarding ICU or hospital discharge may be affected by factors other than disease progression, including local practice variation, cultural factors and economic issues. The use of survival for a defined time-period following study enrollment obviates the variability resulting from subjective decisions regarding ICU transfer, hospital discharge and termination of renal replacement therapy, however the optimal time-point for this assessment is not defined [6].

In critically ill patients without acute renal failure, 28-day survival may miss in excess of 20% of acute mortality [7]. In acute renal failure, a stable survival rate is not achieved until after 30-60 days [2,3,8]. The use of a later time-point, however, may result in increased numbers of patients lost to follow-up, and may result in inclusion of deaths not directly related to the acute renal failure or study intervention. For these reasons we recommend that optimal time point for evaluating all-cause mortality should be between 60 and 90 days of study entrance. In addition to evaluating mortality as a categorical endpoint, survival analysis (e.g. Kaplan-Meier) should also be performed. Evaluating mortality at multiple time-points should also be considered.

### **Assessment of Morbidity**

Several scoring systems for assessment of organ dysfunction and morbidity (e.g., MODS, LODS, and SOFA) have been validated in the general ICU population [9-13] and also been used on ARF patients, although validation studies in the ARF setting are rare [14]. There is insufficient data to recommend any single specific scoring system for patients with ARF. For this reason, morbidity scoring systems should probably not be used as primary outcomes in ARF trials, although they may be appropriate as secondary endpoints.

Organ dysfunction scoring systems have been found to be useful to assess morbidity in a variety of ways, including quantification of baseline severity of organ dysfunction, detection of organ dysfunction developing during the ICU stay, measurement of aggregate severity of organ dysfunction over the ICU stay, and measurement of organ dysfunction free intervals. Assessment of morbidity should be done on a day-to-day basis using a validated organ dysfunction scoring system. A variety of measurements may be useful as clinical endpoints, including maximum achieved score, maximal change in score, time-averaged score, and organ dysfunction free days.

In contrast to organ dysfunction scoring systems, the majority of severity of illness scoring systems (e.g., SAPS II [15], APACHE II [16]) were developed and validated for calculation on ICU admission. The assumed association between the physiologic derangement at admission and subsequent outcome is predicated on the assumption that the patient has not received treatment. This assumption does not apply at subsequent time-points. For this reason, it is not appropriate to use these severity of illness scoring systems to assess morbidity longitudinally during the ICU stay without additional validation.

## Economic Analysis

There are only a few studies that have attempted to calculate ICU costs. However, when these publications are examined in detail [17] it is clear that the methodologies employed are not comparable. Some studies included nursing staff costs or medical staff costs whereas others did not. There are also differences in the inclusion of supplies, medications and many other aspects of treatment. Economic differences and variations in care between countries or regions also confound comparisons between ICU's. Only a limited number of studies have evaluated the costs of ARF or renal replacement therapy in the ICU setting [18-22]. In order to make meaningful cost comparisons it will be essential to develop standard costing methodologies that can be applied across ICU's in any (or most) countries. Such a method is currently under development by the Working Group of Cost-Effectiveness of the European Society of Intensive Care Medicine, in cooperation with other societies [23].

An alternative approach is to calculate patient related costs using a cost proxy. Nursing costs are closely related to the severity of illness of ICU patients. The level of care, measured by systems such as TISS [24] or TISS-28 [25] has therefore often been used for that purpose. It has, however, been shown that although TISS (and other related systems) correlate with overall ICU population costs, the relationship for individual patients is weak [26]. In addition, these proxies correlate with direct costs of care, not with total cost. The limitations of these scoring systems therefore have to be considered when they are used in research projects that compare small populations [27].

## Assessment of Functional Status

There are multiple tools available for the assessment of functional status in the general ICU population. These have not been validated in ARF patients and we cannot make any statements about an optimal tool. However, long-term functional status is an important outcome in the evaluation of care of critically ill patients and is necessary in performing cost-effectiveness analyses.

## References

1. Schiff H, Lang SM, König A, Strasser T, Haider MC, Held E. Biocompatible membranes in acute renal failure: prospective case-controlled study. *Lancet* 1994; 344:570-572.
2. Himmelfarb J, Tolkoff Rubin N, Chandran P, Parker RA, Wingard RL, Hakim R. A multicenter comparison of dialysis membranes in the treatment of acute renal failure requiring dialysis. *J Am Soc Nephrol* 1998; 9:257-66.
3. Ronco C, Bellomo R, Homel P, Brendolan A, Dan M, Piccinni P, La Greca G. Effects of different doses on continuous veno-venous haemofiltration on outcomes of acute renal failure: a prospective randomised trial. *Lancet* 2000;356:26-30.
4. Schiff H; Lang SM; Fischer R. Daily hemodialysis and the outcome of acute renal failure. *N Engl J Med.* 2002; 346: 305-10.
5. Mehta RL, McDonald B, Gabbal FB, Pahl M, Pascual MTA, Farkas A, Kaplan RM. A randomized clinical trial of continuous versus intermittent dialysis for acute renal failure. *Kidney Int* 2001; 60:1154-1163.
6. Hébert PC, Cook DJ, Wells G, Marshall J. The design of randomized clinical trials in critically ill patients. *Chest* 2002; 121:1290-1300.
7. Capuzzo M, Valpondi V, Zardi S, Belin M, De Luca S, Alvisi R. Mortality of intensive care patients: at a fixed point in time or in the hospital? *Intensive Care Med* 2001; 27:Suppl 2: A 367.
8. Gastaldello K, Melot C, Kahn RJ, Vanherweghem JL, Vincent JL, Tielemans. Comparison of cellulose diacetate and polysulfone membranes in the outcome of acute renal failure. A prospective randomized study. *Nephrol Dial Transplant* 2000;15:224-30.
9. Goris RJA, Boekhorst TPA te, Nuytinck HKS et al. Multiple organ failure: generalized autodestructive inflammation? *Arch Surg* 1985;120:1109-1115.
10. Tran DD, Groeneveld ABJ, Van der Meulen J, Nauta JJP, Strack van Schijndel RJM, Thijs LG. Age, chronic disease, sepsis, organ system failure and mortality in a medical intensive care unit. *Crit Care Med* 1990 ; 18 :474-479.
11. Marshall JC, Cook DJ, Christou V, Bernard GR, Sprung CL, Sibbald WJ. Multiple Organ Dysfunction Score: a reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995;23:1638-1652.
12. Vincent JL, Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. *Crit Care Med* 1998;26:1793-1800.
13. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, Teres D. The logistic organ dysfunction system. A new way to assess organ dysfunction in the intensive care unit. *JAMA* 1996;276:802-810.
14. De Mendonca A, Vincent JL, Suter PM, Moreno R, Dearden NM, Antonelli M, Takala J, Sprung C, Cantraine F. Acute renal failure in the ICU: risk factors and outcome evaluated by the SOFA score. *Intensive Care Med* 2000; 26:915-921.
15. Le Gall JR, Lemeshow St, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-2963.
16. Knaus WA, Draper EA, Wagner DP, Zimmermann JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818-829.
17. Gyldmark M. A review of costs studies of intensive care units - problems with the cost concept. *Crit Care Med* 1995;23:964-72.
18. Hamel MB, Phillips RS, Davis RB, Desbiens N, Connors AJ, Teno MJ, Wenger N, Lynn J, Wu AW, Fulkerson W, Tsevat J. Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. *Annals Intern Med* 1997;127:195-202.
19. Korkeila M, Ruokonen E, Takala J. Costs of care, long term prognosis and quality of life in patients requiring renal replacement therapy during intensive care. *Intensive Care Med* 2000; 26:1824-1831.

20. Reeves JH, Cumming AR, Gallagher L, O'Brien JL, Santamaria JD. A controlled trial of low-molecular-weight heparin (dalteparin) versus unfractionated heparin as anticoagulant during continuous venovenous hemodialysis with filtration. *Crit Care Med*. 1999; 27:2316-7.
21. Ash SR, Bever SL. Peritoneal dialysis for acute renal failure: the safe, effective, and low-cost modality. *Adv Ren Replace Ther* 1995; 2:160-3
22. Reznik VM, Randolph G, Collins CM, Peterson BM, Lemire JM, Mendoza SA. Cost analysis of dialysis modalities for pediatric acute renal failure. *Perit Dial Int* 1993;13:311-313.
23. For further details see [www.saps3.org](http://www.saps3.org).
24. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. The therapeutic Intervention Scoring System: a method for quantitative comparison of patient care. *Crit Care Med*; 1975; 3:222-225.
25. Miranda DR, De Rijk A, Schaufeli W. Simplified Therapeutic Interventions Scoring System: The TISS-28 items - Results from a multicenter study. *Crit Care Med* 1996; 24: 64-73.
26. Dickie H, Vedio A, Dundas R, Treacher DF, Leach RM. Relationship between TISS and ICU cost. *Intensive Care Med* 1998; 24:1009-17.
27. Miranda DR. Outcome assessment - TISS as a tool to evaluate cost-effectiveness of immunological treatment. *Eur J Surg Suppl* 1999; 584:51-5.

**Table 1. Potential Endpoints for Studies Evaluating the Prevention of ARF**

Primary Endpoints

- Development of ARF (categorical)
- Quantification of renal function (continuous variable)
- Requirement for renal replacement therapy (surrogate for severity of ARF)

Secondary Endpoints

- Survival
- Morbidity (organ dysfunction measurements)
- Length of stay (hospital, ICU)
- Economic analysis
- Functional status assessment

**Table 2. Potential Endpoints for Studies Evaluating the Treatment of Established ARF**

Primary Endpoints

*These will vary depending on the setting of the trial and/or primary hypothesis (e.g. pharmacologic treatment of ARF vs renal support)*

- Survival
- Assessment of renal function
  - Recovery of renal function
  - Requirement/duration of renal replacement therapy
  - Measurement of renal function

Secondary Endpoints

- Morbidity (organ dysfunction measurements)
- Length of stay (hospital, ICU)
- Composite endpoint (dialysis free survival)
- Functional status assessment
- Economic analysis

**Authors:**

In alphabetic order.

Philipp G.H. Metnitz, MD. Department of Anesthesiology and General Intensive Care, University Hospital, Vienna Austria. Email: [philipp.metnitz@univie.ac.at](mailto:philipp.metnitz@univie.ac.at)

Paul M. Palevsky, MD. Renal Section, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA, USA. Email: [palevsky+@pitt.edu](mailto:palevsky+@pitt.edu)

Pasquale Piccinni, MD. Departments of Intensive Care and Nephrology, St. Bartolo Hospital, Vicenza, Italy. Email: [eleanna@libero.it](mailto:eleanna@libero.it)

Christophe Vinsonneau, MD. Service de Reanimation Polyvalente, Cochin Hospital, Paris, France. Email: [christophe.vinsonneau@cch.ap-hop-paris.fr](mailto:christophe.vinsonneau@cch.ap-hop-paris.fr)