Extracorporeal membrane oxygenation (ECMO) reconsidered

John L Moran, Richard P Chalwin and Petra L Graham

ABSTRACT

The role of extracorporeal membrane oxygenation (ECMO) in the treatment of the acute respiratory distress syndrome (ARDS) is controversial, notwithstanding the recent publication of the results of the CESAR (Conventional Ventilation or ECMO for Severe Adult Respiratory Failure) trial. Using Bayesian meta-analytic methods from three randomised controlled trials (RCTs) of ECMO in ARDS, we estimate the mortality odds ratio to be 0.78 (95% credible interval, 0.25–3.04), P (OR > 1) = 30%. Thus, a null effect of ECMO is not excluded and there appears only weak evidence of efficacy. We survey particular problems associated with the conduct of the “pragmatic” CESAR trial: composite endpoints, sample size estimation under uncertainty of baseline mortality rates, the generation of unbiased treatment comparisons, the impact of treatment non-compliance, and the uncertainty associated with cost-effectiveness and cost-utility analysis. We conclude that the CESAR trial is problematic in terms of both the clinical and economic outcomes, although observational series suggest plausible efficacy. We suggest that ECMO finds rationale as rescue therapy and that the current uncertainty of its role mandates a further RCT.

1. Time span

The three RCTs occurred over a span of nearly 30 years and, as the two recent ECMO publications point out, techniques and experience have undoubtedly improved over this time. Our previous estimate of the efficacy of ECMO in reducing mortality (OR, 1.28, based on two RCTs), was obviously modified by the results of the CESAR trial, the modest degree of heterogeneity reflecting this.

2. Composite endpoints

The purpose of the CESAR trial was to “further define the safety and efficacy of ECMO”. The primary outcome was composite (death or severe disability at 6 months), the trial not being powered for mortality. Composite endpoints increase the efficiency of RCTs, but recent reviews have urged caution in their interpretation. In particular, there is a recommendation that the components of a composite outcome be described and analysed individually. In the final analysis of the CESAR trial, the CESAR trialists rejected composite endpoints for ARDS, although the body of observational data suggested otherwise. We suggest that ECMO finds rationale as rescue therapy and that the current uncertainty of its role mandates a further RCT.
report of the CESAR trial, only one definite recorded case of severe disability was elicited, although the design of the trial was powered to include a 10% incidence of severe disability at 6 months. In the design phase,¹⁰ the anticipated mortality of the conventional-management group was 70%, which turned out to be an overestimate in view of the trial-outcome mortality of 50% for this group. The 70% estimate was “... based on the [National Institutes of Health] ARDS network database ... cross-referencing with the Case Mix Programme Database [of the] Intensive Care National Audit & Research Centre”.¹⁰ However, the quoted 61.6% mortality among patients with arterial partial pressure of oxygen to fraction of inspired oxygen ratio (PaO₂/FIO₂) ≤ 100 in the latter database appeared not to be specifically linked to a diagnosis of ARDS. Basing mortality estimates on lung dysfunction indices is also problematic. In a prospective observational study by Bersten et al (based on data collected in 1999), the hospital mortality of ARDS patients with a Murray lung injury score of ≥ 3 was 35%;¹³ in a 1996 review of 101 ARDS studies, Krafft et al found no relationship between initial PaO₂/FIO₂ ratio and mortality;¹⁴ and, more recently, Luecke et al¹⁵ (and other studies they referred to) reiterated these findings. Thus, the claim¹⁰ that the selection criterion of a Murray score of 3 would identify patients with an expected mortality of 70% appears problematic.

Recruitment period
The problems of estimating (mortality) outcome rates using non-concurrent controls (pilot studies) are well known, especially when mortality rates in the control group change over time.¹⁶ The recruitment of patients in the CESAR trial took place over a 5-year period (2001–2006), during which mortality from ARDS may have been expected to decrease (although this has been a matter of some debate).¹⁷ Annane has suggested that, at least in patients with sepsis, the trial recruitment period should not exceed 24–30 months.¹⁸

Sample size
Sample size may be considered a function of \(\{\text{variance} \times f(\text{error rates})\} + \{\text{minimum relevant effect}\}\),¹⁹ where \(f\) is some function and the error rates are types I and II. The variance that characterises the data-generating process is never exactly known in practice. In sample size calculation, use of the sample standard deviation (\(s^2\)) in place of the population SD (\(\sigma^2\)) has two consequences: (i) the resulting sample size is a random variable, as is the power of the consequent test, rather than the fixed value used in calculation; and (ii) the distribution of \(s^2\) is skewed; thus, more than 50% of the time a random \(s^2\) is less than \(\sigma^2\), and the sample size will be smaller than what is required.²⁰,²¹ To guard against calculating the sample size based on an unrepresentative estimate of control rates, Gould¹⁹ has recommended using the 75th to 80th percentile of the confidence distribution of the population variance.

3. Robustness of results
The question for clinicians relates to the robustness of the results of the CESAR trial, despite the assurance of the trialists that they were “confident that ECMO is a clinically effective treatment”.¹

Although the pivotal property of a randomised clinical trial is random treatment allocation, randomisation per se is not sufficient to provide an unbiased treatment comparison. Additional requirements are that the patient set provides an unbiased assessment of treatment effects and that missing data are ignorable.²² The question of to which presumed population the trial is addressed must also be considered: intention-to-treat provides valid estimates for the effect of the outcome based on original assignment to therapy in the RCT (use effectiveness) but not for the effect of actually administered therapy (method effectiveness). According to Sheiner and Rubin, method effectiveness may be “more relevant to medical decisions than is use effectiveness, and trials should be designed and analysed to provide estimates of it as well”.²³ That is, subsets of patients may be formulated (a reduced analysis set), retaining a cause-and-effect structure, and unbiased parameter estimates may be constructed using the approach of counterfactuals.²³-²⁵ Formal estimators of compliance-adjusted treatment effects have also been proposed.²⁶

That the estimate of the primary outcome lacked robustness is illustrated by the footnote to Table 3 in the published CESAR trial (page 1354¹), where changing the allocation of three patients changed the \(P\) value from 0.017 to 0.051.

4. “Non-compliance” in the treatment arm
The ECMO treatment arm had a “non-compliance” rate of 19% (ie, 17 of the patients [19%] received conventional management). These patients were adjudged, after 12 hours therapy at Glenfield Hospital (the CESAR Clinical Coordinating Centre), not to warrant ECMO, and were also considered to have “slightly less severe” lung disease.¹ This judgement appears at variance with a later comment regarding the “uncertainty in the trial data about patients’ severity of illness”.¹ The mortality of this subset was 18%, which was significantly less (\(P=0.017\)) than the 50% recorded in the non-Glenfield Hospital conventional management sites. There appears to be some contradiction when it is stated both that “participating units did not judge the ECMO unit at Glenfield Hospital to be competent
providers of conventional management or intensive care” and that the non-ECMO Glenfield Hospital patients “responded to expert conventional respiratory intensive care” (no intra-hospital site of management of these patients being given).1

The report of the CESAR trial suggested that secondary analyses were to include per-protocol analysis, but this was not formally presented for either the clinical trial or the economic evaluation. For a mortality endpoint, a naïve (frequentist) per-protocol analysis may be formulated for ECMO: RR 0.92 (95% CI, 0.65–1.29).

5. Economic evaluation

The economic evaluation of the CESAR trial reported the incremental cost-effectiveness of referral for ECMO and incorporated a primary-outcome RR estimate of 0.69 (Table 6, page 1358) that was not based on the full dataset. One of the key assumptions for estimating lifetime quality-adjusted life-years was that, at 24 months after randomisation, “all surviving trial patients attained the same average life expectancy and health state as adults of similar age in the UK population”.27 This assumption equates to the concept of “statistical cure” from a disease or illness (ie, when the relative survival curve of the cohort plateaus and parallels that of the general population.28 In a recent study by Ghelani et al, such a plateau was not observed to 9 years after hospital discharge, at least in patients with sepsis.29

The claim by the CESAR trialists that, based on the cost-effectiveness acceptability method, “consideration for ECMO has more than 50% probability of being cost effective” (page 1358) lacks meaningful precision. Ascribing such a threshold as “more than 50%” appears akin to the tossing of a coin. Although the cost-effectiveness acceptability curve may be the primary comparator of relative cost-effectiveness between two treatments,30 a strictly probabilistic interpretation is only valid in a Bayesian framework.31

Both the incremental cost-effectiveness and cost-utility ratios were contingent on the adoption of a particular use effectiveness (see point 3, above). The claim that ECMO “promises to be cost-effective” (page 1361) must be read against the immediately preceding acknowledgement that “referral to ECMO is likely to prove more efficient than conventional management” (page 1360), based on the “substantial” uncertainty of both the cost-effectiveness and cost-utility analyses (Table 7, page 1358). It thus seems implausible that the findings would be relevant outside the particular trial context.

6. Study design

The CESAR study design was described1 as a “pragmatic” trial similar to the UK trial of neonatal ECMO.32 The use of the pragmatic epithet presumably derives from the classic paper by Schwartz and Lellouch,33 who contrasted explanatory and pragmatic attitudes in clinical trials. We have touched on such a duality in point 3, above.

Between 1993 and 1995, the UK neonatal ECMO trial enrolled 185 infants, and of those allocated to ECMO, 84% received this support, a percentage similar to that of the CESAR trial. It is of note that an initial sample size of 300 was “judged to be sufficient”32 and no formal stopping criteria were specified. The trial was stopped early for efficacy at the 5th interim analysis (the primary endpoint being death or severe disability at 1 year) when 180 children had been enrolled and primary outcome status obtained in 118, but no data were reported of this analysis. The published report provided primary outcome detail on 124 children (Z statistic = 2.87; P = 0.002; RR, 0.54 [95% CI, 0.36–0.80]) and known death before age 1 year in 185 (Z statistic = 3.60; P = 0.002; RR, 0.55 [95% CI, 0.39–0.77]).

If one assumes that the stopping guidelines were Haybittle–Peto (for K analyses, stop at analysis k < K if |Zk| ≥ 3),34 then stopping at the 5th interim analysis (60% recruitment) was not strictly supported for the primary outcome.

The CESAR trial formally adopted Haybittle–Peto stopping guidelines and performed seven interim analyses, but proceeded to completion. Of interest, the data monitoring committee charged with informing the trial steering committee if proof beyond reasonable doubt suggested that “no clear outcome would be obtained with the chosen trial design”.1 The exact meaning of this requirement is unclear, as using Haybittle–Peto boundaries means there is almost no likelihood of stopping early under the null hypothesis.35

For the CESAR trial, expert conventional respiratory intensive care incorporated low-volume low-pressure ventilation, as mandated by the pivotal ARDS Network study in 2000.36 This study, using asymmetrical stopping boundaries37 with a predicted sample size of 1000 patients and a postulated 10% treatment effect,38 stopped for efficacy at the 4th interim analysis (with 80% recruitment and a final reported treatment effect of 8.8%).

Thus, for both adult and neonatal respiratory distress syndromes, “best standard practice” has been derived from early stopping of pivotal trials, one of which (the UK neonatal ECMO trial32) claimed a substantial treatment effect (27%). It is thus somewhat ironic to note both that the status of trial treatment effects reported from early stopping for efficacy has been called into question,39,40 and that one commentator has suggested that future sepsis trials should not conduct interim analyses for efficacy.18

Conclusion

What then is the role of ECMO in adult ARDS? Our analysis suggests that the CESAR trial is problematic in terms of
both clinical and economic outcomes. The most recent report of the widespread use of ECMO, as with other observational series, suggests plausible efficacy but, by definition, neither use effectiveness nor method effectiveness in the strict sense. Other non-RCT validated therapies for ARDS (nitric oxide therapy and prone positioning) are also still being used, as in the CESAR trial. The application of such therapies to the individual patient is not proscribed by the null effect of an RCT; rather, one may argue that the average treatment effect is not the effect of treatment for each individual, although some caution must be exercised in applying such a principle.

Current uncertainty about the role of ECMO in ARDS should mandate a further RCT, despite the status of ECMO as a rescue therapy. If a severe round of seasonal influenza A(H1N1) virus became apparent in 2010, there would be opportunity to initiate such a trial. Furthermore, in the context of rescue therapy, there would appear to be rationale for incorporating specific boundaries for early stopping.

**Author details**

John L Moran, Senior Consultant, and Associate Professor
Richard P Chalwin, Senior Registrar
Petra L Graham, Lecturer

1 Department of Intensive Care Medicine, The Queen Elizabeth Hospital, Adelaide, SA.
2 School of Medicine, University of Adelaide, Adelaide, SA.
3 Department of Intensive Care, Royal Adelaide Hospital, Adelaide, SA.
4 Department of Statistics, Faculty of Science, Macquarie University, Sydney, NSW.

**Correspondence:** john.moran@adelaide.edu.au

**References**

26 Loeyes T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics 2003; 59: 100-5.
27 Thalane M, Mugford M, Hibbert C, et al. Methods of data collection and analysis for the economic evaluation alongside a national, multi-centre trial in the UK: Conventional Ventilation or
ECMO for Severe Adult Respiratory Failure (CESAR). BMC Health Serv Res 2008; 8: 94.


