

# Why Are Clinicians Not Embracing the Results from Pivotal Clinical Trials in Severe Sepsis? A Bayesian Analysis

Andre C. Kalil<sup>1\*</sup>, Junfeng Sun<sup>2</sup>

**1** Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, **2** Assistant Professor, Department of Biostatistics, College of Public Health, University of Nebraska, Omaha, Nebraska, United States of America

## Abstract

**Background:** Five pivotal clinical trials (Intensive Insulin Therapy; Recombinant Human Activated Protein C [rhAPC]; Low-Tidal Volume; Low-Dose Steroid; Early Goal-Directed Therapy [EGDT]) demonstrated mortality reduction in patients with severe sepsis and expert guidelines have recommended them to clinical practice. Yet, the adoption of these therapies remains low among clinicians.

**Objectives:** We selected these five trials and asked: Question 1-What is the current probability that the new therapy is not better than the standard of care in my patient with severe sepsis? Question 2-What is the current probability of reducing the relative risk of death (RRR) of my patient with severe sepsis by meaningful clinical thresholds (RRR >15%; >20%; >25%)?

**Methods:** Bayesian methodologies were applied to this study. Odds ratio (OR) was considered for Question 1, and RRR was used for Question 2. We constructed prior distributions (enthusiastic; mild, moderate, and severe skeptic) based on various effective sample sizes of other relevant clinical trials (unfavorable evidence). Posterior distributions were calculated by combining the prior distributions and the data from pivotal trials (favorable evidence).

**Main Findings:** Answer 1-The analysis based on mild skeptic prior shows beneficial results with the Intensive Insulin, rhAPC, and Low-Tidal Volume trials, but not with the Low-Dose Steroid and EGDT trials. All trials' results become unacceptable by the analyses using moderate or severe skeptic priors. Answer 2-If we aim for a RRR>15%, the mild skeptic analysis shows that the current probability of reducing death by this clinical threshold is 88% for the Intensive Insulin, 62–65% for the Low-Tidal Volume, rhAPC, EGDT trials, and 17% for the Low-Dose Steroid trial. The moderate and severe skeptic analyses show no clinically meaningful reduction in the risk of death for all trials. If we aim for a RRR >20% or >25%, all probabilities of benefits become lower independent of the degree of skepticism.

**Conclusions:** Our clinical threshold analysis offers a new bedside tool to be directly applied to the care of patients with severe sepsis. Our results demonstrate that the strength of evidence (statistical and clinical) is weak for all trials, particularly for the Low-Dose Steroid and EGDT trials. It is essential to replicate the results of each of these five clinical trials in confirmatory studies if we want to provide patient care based on scientifically sound evidence.

**Citation:** Kalil AC, Sun J (2008) Why Are Clinicians Not Embracing the Results from Pivotal Clinical Trials in Severe Sepsis? A Bayesian Analysis. *PLoS ONE* 3(5): e2291. doi:10.1371/journal.pone.0002291

**Editor:** Adam J. Ratner, Columbia University, United States of America

**Received:** October 26, 2007; **Accepted:** April 9, 2008; **Published:** May 28, 2008

**Copyright:** © 2008 Kalil, Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: akalil@unmc.edu

## Introduction

*“If we begin with certainties, we shall end in doubts; but if we begin with doubts, and are patient with them, we shall end with certainties.”*

*Sir Francis Bacon (1605)*

More than 20 clinical trials involving over 10,000 patients have been performed in patients with sepsis and severe sepsis in the last 15 years with little success in reducing mortality [1]. More recently, five published clinical trials: Early Goal-Directed Therapy [2], Recombinant Human Activated Protein C [3], Low-Dose Steroid [4], Low-Tidal Volume-ARDS Network [5], and Intensive Insulin Therapy [6] demonstrated positive outcome

results and brought the prospect of improving the survival of patients with severe sepsis.

Ten multinational medical societies sponsored a joint statement, ‘Surviving Sepsis Campaign’, in which recommendations are made to include the results of these trials in the standard of care for patients with severe sepsis [7]. These recommendations have also been evaluated by the Joint Commission on Accreditation of Healthcare Organizations [8]. Despite these positive outcomes and recommendations, scientists and clinicians have been either slow or resistant to adopt the results of these trials at face value in order to apply them to patient care [9–25]. Still, strong endorsement by the medical societies is not coming without criticisms and opposition by the medical community [8,26]. Why

is this resistance to accept statistically significant results from large clinical trials so accentuated in the sepsis field?

We propose that the genesis for most of these issues lies in the confounding interpretation and poor translation of these results to the bedside, and the lack of formal analysis combining previous evidence and the current positive clinical trials. While controversy is necessary for the progression of science [27], when it comes to treating a patient with severe sepsis, a clinical decision is also necessary for the betterment of this patient's outcome.

In the following paragraphs, we argue that the best solution for the understanding of pivotal clinical trials in severe sepsis can only come from a friendly reunion of classic (frequentist) and Bayesian statistical methodologies [28–31]. The application of this more inclusive and robust interpretation of trial results will facilitate their application directly to the bedside, and will hopefully further improve the care of our patients with severe sepsis. Moreover, this “dualistic” approach will also empower us to better define the need for confirmatory trials in order to optimize the current standard of care.

## Methods

### A. Methods Background

The “early goal-directed therapy” (EGDT) trial [2] will be used as a practical example to describe the rationale for our methodology. This trial aimed to compare the use of early volume replacement/vasopressor use in the treatment arm against standard of care in the control arm for patients with severe sepsis. The final results showed a 42% relative reduction in the risk of death (16% absolute risk reduction) of in-hospital mortality with a 95% confidence interval (CI) (0.13–0.62),  $p=0.009$ . Common clinical interpretations are: (a) there is only a 0.9% probability of a false positive result (rejecting the null hypothesis when there is no treatment difference); (b) there is a 95% probability of the true risk ratio being somewhere between 0.13 and 0.62 (a reduction of 13–62% in the relative risk of death). These interpretations are incorrect due to the misunderstanding of the classic or frequentist (frequency based view of probability) statistical reporting used in this and most trials [32–35]. The correct interpretation of the classic method for this trial is the following: There is a 0.9% probability that results as good as or better than the ones found in this trial (42% relative reduction in the risk of death), will be observed among a large number of hypothetical repetitions of this trial under the null hypothesis of no treatment effect. In addition, the CIs generated from 95% of these hypothetical trials will cover the true mortality reduction. This rather convoluted language is the only possible interpretation which explicitly states that the classic method cannot provide the probabilities that clinicians are seeking for. In other words, clinicians are interpreting the conventional  $p$  values and CIs as ‘current probabilities’, although the classic method alone only provides us with probabilities over a large number of hypothetical repetitions in the long term.

Here is where the Bayesian methodology comes in to complement the classic interpretation of clinical trials. It allows us to think the way most clinicians are already thinking [36]! That is, what is the posterior or ‘current probability’ (not the probability in the long term of hypothetical trial repetitions) of observing this outcome in a given trial or population.

In order to use the Bayesian methodology, a prior probability is required. The prior probability can be based on all available evidence (i.e. biological rationale and its pre-clinical evaluation; observational and experimental clinical data) gathered by studies other than the new trial being currently analyzed. An analogy to the diagnostic setting states that “it is not possible to find a

probability of having a disease based on tests results without specifying the disease's prevalence” [37]. The basic idea described by Bayes demonstrates that the product between the prior probability and the evidence provided by the new trial (also called Bayes factor) will give us the posterior probability, which we call the “current probability” in our paper. This simple and clever algebraic calculation allows us to overcome the unsolvable ‘long term hypothetical repetitions’ inherent issue of the classic method. At the same time, it gives us what we are mostly looking for in our daily medical practice, i.e. what is the current probability to achieve the results of this pivotal trial in my patients? The potential “subjectivity” of these priors have made some statisticians and clinicians concerned about the use of this method. However, the complete exclusion of prior knowledge and evidence from the design and interpretation of recently completed clinical trials has been compared to a sentencing judge who overlooks the prior convictions of a habitual criminal [38]. Needless to say, the commonly used classic methodology in current clinical trials is far from objective. For example, the models assumed, the parameters and hypothesis chosen, and the experimental designs employed [37] are typical features that incorporate much subjectivity into the classic analysis. We agree with Berry [37] that “...silent subjectivities such as these (seen with classic methods) are dangerous in that they are difficult or impossible to make explicit. By contrast, subjectivity in prior distributions (as seen with Bayesian methods) is explicit and open to examination-and critique-by all”. Thus, how should we best determine the priors for this study? This so-called subjectivity is easily resolved in the severe sepsis world because we have many negative clinical trials done before the current positive trials, which set the stage for the perfect use of Bayesian methodology. We will provide the clinician with a realistic spectrum of prior distributions, so he/she can find the current probability of the new treatment being no better than the control (standard of care), and the current probability of reducing mortality by a clinically meaningful threshold. These two probabilities will allow the clinician to make the best clinical decision for the patient with severe sepsis without being entirely dependent on sponsors, regulators, editors, and experts in the field.

### B. Methods Description

**1. What is the current probability that the new therapy is not better than the standard of care in my patient with severe sepsis?** The goal of the first part of our study is to assess the current probability that the new therapy is no better than the control, i.e. standard of care. We used the log-odds ratio ( $\ln(\text{OR})$ ) of death for the new treatment group compared to the control group. We considered the new therapy to be no better than the control if the  $\ln(\text{OR})$  was found to be greater than  $-0.05$  [38,39]. We then constructed various prior distributions of  $\ln(\text{OR})$  assuming previous trials of different effective sample sizes (see Appendix S1). The prior distributions may be based on the following information: Early Goal-Directed Therapy [40–47], rhAPC [48–53], Low-Dose Steroids [54–67], Low-Tidal Volume [68–74], Intensive Insulin Therapy [75–85]. Of note, some trials were performed before and others after the pivotal positive trial of a given therapy. This inclusive approach of all evidence available is an important strength of the Bayesian technique, which does not require that the priors have a temporal order [32,37,86,87]. This list of studies for each therapy allows the reader to get his/her own effective sample size by summing up the prior unfavorable evidence (total sample size of relevant negative clinical trials). If believing there is no relevant previous negative data, one can assume the effective sample size to be 1 [88]. This gives us the non-informative prior, which is our ‘enthusiastic’ prior, since it is

**Table 1.** Current probability of new treatment being no better than control based on mortality outcomes.

CLINICAL TRIALS (Favorable Evidence)	Early Goal-Directed Therapy	Activated Protein C	Activated Protein C (APACHE II >25)	Low-Dose Steroids (Non-Responders)	Low-Tidal Volume	Intensive Insulin Therapy (>5 days in ICU)
Published P values (Frequentist)	0.009	0.005	0.0002	0.02	0.007	0.005
Sample Size of Unfavorable Evidence (Bayesian Priors)	<b>CURRENT PROBABILITY OF NEW TREATMENT BEING NO BETTER THAN CONTROL</b>					
ENTHUSIASTIC (Sample Size = 1)	0.01	0.009	0.0003	0.03	0.009	0.003
MILD SKEPTIC (Sample Size = 200)	0.05	0.02	0.002	0.10	0.02	0.02
MILD-MODERATE SKEPTIC (Sample Size = 500)	0.14	0.03	0.007	0.21	0.05	0.05
MODERATE SKEPTIC (Sample Size = 1000)	0.25	0.05	0.03	0.33	0.10	0.12
SEVERE SKEPTIC (Sample Size = 2000)	0.41	0.12	0.09	0.47	0.21	0.23

doi:10.1371/journal.pone.0002291.t001

ignoring all negative sepsis trials already published. If one is skeptical about the effects of the new therapy based on previous phase 2 and 3 trials, observational studies, or clinical experience, the effective sample size of negative evidence for the (skeptical) prior will be larger. Thus, the more skeptical one is based on the prior information, the harder it is to conclude efficacy using the same pivotal trial data. Based on the mortality rate and the required sample size to evaluate new therapies for severe sepsis in current times, the following effective sample sizes of negative evidence were analyzed for each new therapy: Enthusiastic ( $n = 1$ ); Mildly skeptical ( $n = 200$ ); Mild-moderately skeptical ( $n = 500$ ); Moderately skeptical ( $n = 1,000$ ); Severely skeptical ( $n = 2,000$ ). We consider the current probability of  $\ln(\text{OR}) > -0.05$  being less than 0.05 as sufficient evidence to conclude that the new treatment is better than the control. For example, if the clinician sums up approximately 500 subjects from 3 previous negative trials on low-tidal volume therapy [69–71] (i.e. mild to moderate skeptic about this therapy), our analysis (table 1) will provide a 0.05 current probability of low-tidal volume therapy being no better than control. As the reader can appreciate in table 1, because of the negative prior information that needs to be overcome, this current probability (0.05) is different from the classic p value reported in the original trial publication ( $p = 0.007$ ).

**2. What is the current probability of decreasing the relative risk of death of my patient with severe sepsis by a meaningful clinical threshold?** Based on the multitude of clinical trials already published [1], the recent FDA approval of a new therapy on sepsis [3], and a 28-day mortality ranging from 30–40%, no new therapy for severe sepsis will likely be accepted by clinicians or regulators if the relative risk reduction (RRR) for mortality is not greater than 15–25% (absolute risk reduction (ARR)  $\geq 5$ –10%). For example, a trial with a control arm mortality of 30% and an experimental arm mortality of 25% would result in an ARR = 5% and a RRR = 16%. Thus, we will present the current probabilities for greater than 15%, 20%, and 25% RRR for mortality in each trial (see Appendix S1). This second analysis can be thought of as analogous to ‘clinical significance’. For example, if the clinician sums up approximately 1000 subjects from 2 previous negative trials on volume replacement/vasopressor use [45,46] (i.e. moderate skeptic about this therapy) and believes that a minimum of 15% RRR needs to be demonstrated to add the EGDT to standard of care of patients with severe sepsis, our analysis (table 2) will provide a 6% current probability of reaching at least this RRR in mortality. This current probability (6%) of reducing mortality by at least 15% is different from the classic overall RRR reported in the original trial

publication (42%). This level of prediction can not be achieved with classic methods alone.

The results of the five positive clinical trials [2–6] will be discussed in the context of the two probability questions described above. Because of the non-significant overall results of the low-steroid trial for mortality [4], we will evaluate the prospectively defined “non-responders” sub-population. Due to the low overall control mortality (8%), and the inclusion of all comers (with and without severe sepsis) to a surgical ICU in the intensive insulin trial [6], we will evaluate the prospectively defined “>5 days in ICU” sub-population. This subgroup control had mortality rates (20%) closer to those of the other trials. The rhAPC trial [3] will have two analyses, since the drug was FDA approved based on the sicker (APACHE II >25) sub-population. Even though the ARDSNet Low-Tidal Volume trial [5] was not specifically designed for patients with severe sepsis, we included it because approximately 60% of the trial population had sepsis or pneumonia, and the trial results have been recommended for patients with severe sepsis by practice guidelines [7].

## Results

1. What is the current probability that the new therapy is not better than the standard of care in my patient with severe sepsis?

Table 1 shows the current probability of the new treatment being no better than the control. If we are enthusiastic about the five trials, all probabilities are small. If we are just mildly skeptic,

**Table 2.** Probability of Relative Risk Reduction (RRR) of Mortality by EGDT.

EARLY GOAL-DIRECTED THERAPY			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR>15%	RRR>20%	RRR>25%
ENTHUSIASTIC	94%	87%	78%
MILD SKEPTIC	62%	41%	22%
MILD-MODERATE SKEPTIC	27%	9%	2%
MODERATE SKEPTIC	6%	1%	0
SEVERE SKEPTIC	0.4%	0	0

doi:10.1371/journal.pone.0002291.t002

the probabilities from the EGDT [2] and Low-Dose Steroid [4] trials become 0.05 or larger, which may not provide strong enough evidence to change the standard of care. If we take into account the preceding multitude of unfavorable trials and analyze these results in the light of a mild-moderate skeptical view, the current probabilities for these two trials rise to 0.14 and 0.21, respectively. The results of rhAPC, Low-Tidal Volume, and Intensive Insulin trials remain acceptable if we assume a mild skepticism—all with current probabilities of less than 0.05. The mild-moderate skeptic analysis makes both the Low-Tidal Volume and Intensive Insulin results reach the 0.05 probability of the new treatment being no better than the control, and the moderate skeptic analysis shows the rhAPC trial with the same 0.05 probability as well. The rhAPC APACHE II sub-analysis shows probabilities below 0.05 in all prior levels except in the severe skeptic analysis. If the clinician is severely skeptical, no trial results will lead to changes in the standard of care.

## 2. What is the current probability of decreasing the relative risk of death of my patient with severe sepsis by a meaningful clinical threshold?

Table 2 illustrates the current probabilities of achieving at least a specific clinically meaningful RRR based on the chosen cut-off and prior distribution. For a clinician who is enthusiastic (i.e. choosing the enthusiastic prior) about EGDT and believes that an RRR should not be less than 20% to use this new therapy in patient care, the chance of EGDT to decrease the relative risk of death more than 20% is 87%. If  $RRR > 15\%$  is the goal, this therapy will have a 94% (enthusiastic) and 62% (mildly skeptic) current probabilities. This is a tangible and easy result to translate and apply to bedside. For the clinician who is mildly skeptical about this therapy, the chance of EGDT to decrease the relative risk of death more than 20% in a given patient will decrease to 41%, which may be too low for general clinical application. If a moderately skeptical prior is applied to this trial, the probability of having a RRR greater than 15% and 20% drops substantially to 6% and 1%, respectively.

For the rhAPC overall results (table 3), in the best case scenario (enthusiastic), the current probability of reaching a 20% RRR in mortality is 43%. If the clinician is comfortable with a 15% RRR, then the probability of reaching this cut-off with rhAPC goes up to 75% for the enthusiastic, and to 65% for the mildly skeptical, but remains low at 30% for the moderately skeptical. On the other side, if we analyze this trial based on APACHE II  $>25$  subgroup and enthusiastic prior, then there is 97%, 90%, and 72% probability of respectively reaching RRR of 15%, 20%, and 25% with rhAPC. These are more optimistic results than the overall trial analysis, but if the clinician remains a mild-moderate skeptic, the probability of achieving any of those same RRR thresholds becomes smaller; 63%, 30%, and 7%, respectively. The moderate and severe skeptic analyses show most probabilities of reaching any meaningful RRR in the single digits.

The Low-Dose Steroid trial analysis demonstrates well the importance of this type of clinical threshold analysis. In the best case-scenario for the least meaningful RRR (15%), the enthusiastic approach shows a current probability of 58% (table 4). Even for the mildly skeptic, steroids have an unacceptably low current probability (20% or less) of reaching any meaningful RRR in mortality. We also performed an additional analysis about the probability of steroids reaching a  $RRR > 10\%$ , but except for the enthusiastic prior (76%), all other probabilities remain similarly poor (41% or less).

The Low-Tidal Volume trial (table 5) shows a current probability of 83% (enthusiastic) or 65% (mild skeptic) for the physician who is looking for a  $RRR > 15\%$ . However, for the

**Table 3.** Probability of Relative Risk Reduction (RRR) of Mortality by recombinant human APC.

ACTIVATED PROTEIN C			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR >15%	RRR >20%	RRR >25%
ENTHUSIASTIC	75%	43%	14%
MILD SKEPTIC	65%	31%	7%
MILD-MODERATE SKEPTIC	50%	17%	3%
MODERATE SKEPTIC	30%	6%	0.4%
SEVERE SKEPTIC	9%	0.5%	0
ACTIVATED PROTEIN C (APACHE II >25)			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR >15%	RRR >20%	RRR >25%
ENTHUSIASTIC	97%	90%	72%
MILD SKEPTIC	88%	66%	35%
MILD-MODERATE SKEPTIC	63%	30%	7%
MODERATE SKEPTIC	27%	5%	0.2%
SEVERE SKEPTIC	3%	0.1%	0

doi:10.1371/journal.pone.0002291.t003

mildly skeptic requiring a  $RRR > 20\%$  the probability drops to 37%, and for the mild-moderate skeptic requiring any RRR, the probabilities of benefit from this therapy remain all below 40%.

The Intensive Insulin trial (table 6) shows consistent probabilities of reducing the risk of death (69–98%) for all RRR levels in both enthusiastic and mild skeptic analyses. The probability remains above 60% even in the mild-moderate skeptic analysis if the aim is a  $RRR > 15\%$ . However, for the moderate or severe skeptic approach, all results are below 42%.

## Discussion

The first analysis indicates that there is sufficient evidence to support the efficacy in all five trials only if we are enthusiastic with respect to the prior distribution of each of these therapies. If we analyze them with the mild skepticism, only the Intensive Insulin, rhAPC, and Low-Tidal Volume trials show beneficial results.

**Table 4.** Probability of Relative Risk Reduction (RRR) of Mortality by Low-Dose Steroid.

LOW-DOSE STEROID			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR >15%	RRR >20%	RRR >25%
ENTHUSIASTIC	58%	37%	19%
MILD SKEPTIC	17%	4%	0.5%
MILD-MODERATE SKEPTIC	3%	0.1%	0
MODERATE SKEPTIC	0.1%	0	0
SEVERE SKEPTIC	0	0	0

doi:10.1371/journal.pone.0002291.t004

**Table 5.** Probability of Relative Risk Reduction (RRR) of Mortality by Low-Tidal Volume.

LOW-TIDAL VOLUME			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR >15%	RRR >20%	RRR >25%
ENTHUSIASTIC	83%	62%	34%
MILD SKEPTIC	65%	37%	13%
MILD-MODERATE SKEPTIC	41%	14%	2%
MODERATE SKEPTIC	16%	2%	0.1%
SEVERE SKEPTIC	2%	0	0

doi:10.1371/journal.pone.0002291.t005

Although the rhAPC APACHE II >25 sub-analysis remains beneficial in the moderate skeptic analysis, the original results have not been validated in a prospective phase III trial yet. Based on our current results indicating high probability of new treatment being no better than controls in all skeptic analyses for the EGDP and Low-Dose Steroid trials, as well as on the numerous unfavorable trials evaluating different regimens of volume replacement/vasopressors or steroids in severe sepsis, we demonstrate that the beneficial results from these two trials are the ones with the weakest strength of evidence. All five trials become unacceptable if we are moderately or severely skeptical, because the current probability of the new therapy being no better than the control is too high for general clinical application. The fact that most overall beneficial results are not even that impressive with the mild-moderately skeptic analysis is concerning.

The second analysis based on a specific clinical cut-off of the RRR brings important light to the interpretation of these trials. Assuming an enthusiastic prior, the Intensive Insulin, the rhAPC APACHE II >25, and the EGDP trials demonstrate the highest probability (87–96%) of reaching an RRR of at least 20% for mortality. However, the absence of a majority of patients with severe sepsis in the Intensive Insulin trial, the absence of prospective validation of the rhAPC APACHE II >25 subgroup population, and the weak strength of evidence from our first analysis for the EGDP trial, all suggest that more skepticism should be taken before assuming these probabilities of RRR>20%. In this case, the moderate skeptic analysis for these 3 trials shows all current probabilities of 23% or less of reducing death risk by more

**Table 6.** Probability of Relative Risk Reduction (RRR) of Mortality by Intensive Insulin Therapy.

INTENSIVE INSULIN THERAPY			
SAMPLE SIZE OF UNFAVORABLE EVIDENCE	CURRENT PROBABILITY OF RELATIVE RISK REDUCTION		
	RRR >15%	RRR >20%	RRR >25%
ENTHUSIASTIC	98%	96%	93%
MILD SKEPTIC	88%	80%	69%
MILD-MODERATE SKEPTIC	68%	52%	35%
MODERATE SKEPTIC	42%	23%	9%
SEVERE SKEPTIC	16%	4%	0.6%

doi:10.1371/journal.pone.0002291.t006

than 20%. On the other hand, if we are mildly skeptical and aim a RRR>15%, the probability of reducing the risk of death is 88% for the Intensive Insulin trial, 62–65% for the rhAPC, Low-Tidal Volume, and EGDT trials, and just 17% for the Low-Dose Steroid trial. Unless we are less ambitious with respect to the RRR, i.e. 15%, and accept an enthusiastic to mildly skeptic prior, none of these trials have shown strong enough evidence for diminishing the risk of death in patients with severe sepsis. Of note, the consistently very low current probabilities (poor strength of evidence) of observing mortality reduction with Low-Dose Steroids in both of our analyses (questions 1 and 2) was just confirmed by a recently published phase III trial [66]. Interestingly, our statistical analysis predicting the poor results of Low-Dose Steroids was completed and submitted long before the report of this new phase III trial.

We would like to recognize some limitations of our study. The statistical approach we used for this study may appear to produce more conservative results than conventional methods, but this is, in fact, the main strength of our data. The Bayes' theorem is uncontroversial if derived from known data [88]. We strongly believe that the prior negative evidence is so abundant that we have the ethical obligation to consider and use this methodology in the analysis of any new therapy for severe sepsis. The different sample sizes of each trial may have influenced the current probabilities of our first analysis, but the consistent results found in both first and second analyses make the sample size influence less likely. Also, because some of the therapeutic interventions were not identical within a specific class, and other studies were different with respect to their trial design, we advise the reader to carefully evaluate the most appropriate priors to avoid overt pessimistic or optimistic current probabilities.

In conclusion, our study provides four clinical and research lessons with profound implications to the care of our patients with severe sepsis:

### Lesson 1—Need for confirmatory trials

Our results unambiguously demonstrate that it is important to replicate the results of each of these trials in well designed confirmatory studies if we want to provide patient care based on scientifically sound evidence. Further, the many study design issues raised, e.g. standard of care of control groups (EGDT and Low-Tidal Volume trials [10–12,17,21–24]; the rhAPC APACHE II >25 subgroup analysis without prospective validation [3] and the poor results of this subgroup in the ADDRESS trial [49,89]; the controversial definitions of adrenal insufficiency [20,90,91], and the just reported Low-Dose Steroid phase III trial with negative results [66]; and the failure of the Intensive Insulin Therapy to improve survival in the medical ICU population [75] all corroborate our conclusion for lesson 1.

### Lesson 2—Standard of care for patients with severe sepsis

The strength of evidence (statistical and clinical) is overall weak for the five trials. These results make any legitimate changes in the standard of care a very difficult task to accomplish. While we endorse the genuine need for more evidence, we are aware of the urgent need to improve the survival outcome of our patients with severe sepsis. How to reconcile this apparent conundrum? Sir Austin Hill already had the answer in his seminal paper from 1965: “All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer on us a freedom to ignore the knowledge that we already have, or to postpone the action that it appears to demand at a given time” [92]. Before we have the results of confirmatory trials, we urge clinicians to use our clinical threshold analysis of RRR for

mortality from each trial (tables 2–6) to guide the best course of action to take for their patients with severe sepsis.

### Lesson 3—No need for changes in the measurement of ICU quality of care

The most important premise to change the measurement of quality of care must be based on strong and established scientific evidence, but this is lacking at this time. When we completed our study it became obvious that, with such low strength of evidence and critical need for confirmatory trials, none of these studies' results should be applied as rigid tools to measure quality of care in patients with severe sepsis in the ICU.

### Lesson 4—Need for both classic and Bayesian interpretation of clinical trials

As we demonstrate in this paper, the dual use of these methods is powerful and synergistic to accomplish an ample interpretation of these pivotal trials. We strongly suggest that trialists, sponsors, regulators, and journal editors become more proactive with respect to the use of this dual statistical approach. Patients will be the ultimate beneficiaries from this more encompassing strategy.

## References

- Eichacker PQ, Parent C, Kalil A, Esposito C, Cui X, et al. (2002) Risk and the efficacy of antiinflammatory agents: retrospective and confirmatory studies of sepsis. *Am J Respir Crit Care Med* 166: 1197–1205.
- Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, et al. (2001) Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 345: 1368–1377.
- Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, et al. (2001) Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 344: 699–709.
- Anname D, Sebille V, Charpentier C, Bollaert PE, Francois B, et al. (2002) Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *Jama* 288: 862–871.
- (2002) Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med* 342: 1301–1308.
- van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, et al. (2001) Intensive insulin therapy in the critically ill patients. *N Engl J Med* 345: 1359–1367.
- Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, et al. (2004) Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med* 32: 858–873.
- Eichacker PQ, Natanson C, Danner RL (2006) Surviving sepsis—practice guidelines, marketing campaigns, and Eli Lilly. *N Engl J Med* 355: 1640–1642.
- Rubinfeld GD, Caldwell E, Hudson LD (2001) Publication of study results does not increase use of lung-protective strategy in patients with acute lung injury. *Am J Respir Crit Care Med*. 295 p.
- Eichacker PQ, Gerstenberger EP, Banks SM, Cui X, Natanson C (2002) Meta-analysis of acute lung injury and acute respiratory distress syndrome trials testing low tidal volumes. *Am J Respir Crit Care Med* 166: 1510–1514.
- Ricard JD (2003) Are we really reducing tidal volume—and should we? *Am J Respir Crit Care Med* 167: 1297–1298.
- Deans KJ, Minneci PC, Cui X, Banks SM, Natanson C, et al. (2005) Mechanical ventilation in ARDS: One size does not fit all. *Crit Care Med* 33: 1141–1143.
- Friedrich JO, Adhikari NK, Meade MO (2006) Drotrecogin alfa (activated): down and not out, but not really in either. *Crit Care* 10: 420.
- Wiedermann CJ (2006) When a single pivotal trial should not be enough—the case of drotrecogin-alfa (activated). *Intensive Care Med* 32: 604.
- Mackenzie AF (2005) Activated protein C: do more survive? *Intensive Care Med* 31: 1624–1626.
- Eichacker PQ, Danner RL, Suffredini AF, Cui X, Natanson C (2005) Reassessing recombinant human activated protein C for sepsis: time for a new randomized controlled trial. *Crit Care Med* 33: 2426–2428.
- Ho BC, Bellomo R, McGain F, Jones D, Naka T, et al. (2006) The incidence and outcome of septic shock patients in the absence of early-goal directed therapy. *Crit Care* 10: R80.
- Schultz MJ, Wolthuis EK, Moeniralam HS, Levi M (2005) Struggle for implementation of new strategies in intensive care medicine: anticoagulation, insulin, and lower tidal volumes. *J Crit Care* 20: 199–204.
- Vriesendorp TM, van Santen S, DeVries JH, de Jonge E, Rosendaal FR, et al. (2006) Predisposing factors for hypoglycemia in the intensive care unit. *Crit Care Med* 34: 96–101.
- Hamrahian AH, Oseni TS, Arafah BM (2004) Measurements of serum free cortisol in critically ill patients. *N Engl J Med* 350: 1629–1638.
- Natanson C, Danner RL (2002) Early goal-directed therapy reduced mortality and multiorgan dysfunction in severe sepsis or septic shock. *ACP J Club* 136: 90.
- Marik PE, Varon J (2002) Goal-directed therapy for severe sepsis. *N Engl J Med* 346: 1025–1026; author reply 1025–1026.
- Abroug F, Besbes L, Nounira S (2002) Goal-directed therapy for severe sepsis. *N Engl J Med* 346: 1025–1026; author reply 1025–1026.
- Sarkar S, Kupfer Y, Tessler S (2002) Goal-directed therapy for severe sepsis. *N Engl J Med* 346: 1025–1026; author reply 1025–1026.
- Warren HS, Suffredini AF, Eichacker PQ, Munford RS (2002) Risks and benefits of activated protein C treatment for severe sepsis. *N Engl J Med* 347: 1027–1030.
- Wiedermann CJ (2005) Bioethics, the Surviving Sepsis Campaign, and the industry. *Wien Klin Wochenschr* 117: 442–444.
- Tobin MJ (2003) The role of a journal in a scientific controversy. *Am J Respir Crit Care Med* 168: 511.
- Brophy JM, Joseph L (1995) Placing trials in context using Bayesian analysis. *GUSTO revisited by Reverend Bayes*. *Jama* 273: 871–875.
- Petersen JL, Mahaffey KW, Hasselblad V, Antman EM, Cohen M, et al. (2004) Efficacy and bleeding complications among patients randomized to enoxaparin or unfractionated heparin for antithrombin therapy in non-ST-Segment elevation acute coronary syndromes: a systematic overview. *Jama* 292: 89–96.
- Goodman SN (1999) Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 130: 1005–1013.
- Davidoff F (1999) Standing statistics right side up. *Ann Intern Med* 130: 1019–1021.
- Burton PR, Gurrin LC, Campbell MJ (1998) Clinical significance not statistical significance: a simple Bayesian alternative to p values. *J Epidemiol Community Health* 52: 318–323.
- Sterne JA, Davey Smith G (2001) Sifting the evidence—what's wrong with significance tests? *Bmj* 322: 226–231.
- Browner WS, Newman TB (1987) Are all significant P values created equal? The analogy between diagnostic tests and clinical research. *Jama* 257: 2459–2463.
- Goodman SN (2005) Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials* 2: 282–290; discussion 301–284, 364–278.
- Gill CJ, Sabin L, Schmid CH (2005) Why clinicians are natural bayesians. *Bmj* 330: 1080–1083.
- Berry DA (2006) Bayesian clinical trials. *Nat Rev Drug Discov* 5: 27–36.
- Diamond GA, Kaul S (2004) Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. *J Am Coll Cardiol* 43: 1929–1939.
- Harrell FE Jr, Shih YC (2001) Using full probability models to compute probabilities of actual interest to decision makers. *Int J Technol Assess Health Care* 17: 17–26.
- Lin SM, Huang CD, Lin HC, Liu CY, Wang CH, et al. (2006) A modified goal-directed protocol improves clinical outcomes in intensive care unit patients with septic shock: a randomized controlled trial. *Shock* 26: 551–557.
- Alia I, Esteban A, Gordo F, Lorente JA, Diaz C, et al. (1999) A randomized and controlled trial of the effect of treatment aimed at maximizing oxygen delivery in patients with severe sepsis or septic shock. *Chest* 115: 453–461.

42. Yu M, Burchell S, Hasaniya NW, Takanishi DM, Myers SA, et al. (1998) Relationship of mortality to increasing oxygen delivery in patients  $>$  or  $=$  50 years of age: a prospective, randomized trial. *Crit Care Med* 26: 1011–1019.
43. Yu M, Levy MM, Smith P, Takiguchi SA, Miyasaki A, et al. (1993) Effect of maximizing oxygen delivery on morbidity and mortality rates in critically ill patients: a prospective, randomized, controlled study. *Crit Care Med* 21: 830–838.
44. Tuchschildt J, Fried J, Astiz M, Rackow E (1992) Elevation of cardiac output and oxygen delivery improves outcome in septic shock. *Chest* 102: 216–220.
45. Hayes MA, Timmins AC, Yau EH, Palazzo M, Hinds CJ, et al. (1994) Elevation of systemic oxygen delivery in the treatment of critically ill patients. *N Engl J Med* 330: 1717–1722.
46. Gattinoni L, Brazzi L, Pelosi P, Latini R, Tognoni G, et al. (1995) A trial of goal-oriented hemodynamic therapy in critically ill patients. SvO<sub>2</sub> Collaborative Group. *N Engl J Med* 333: 1025–1032.
47. Xiao-zhi W, Chang-jun L, Fu-quan GAO, Xiao-hong LI (2007) Efficacy of goal-directed therapy in the treatment of septic shock. *Chin Crit Care Med (Engl)* 18: 661–664.
48. Bernard GR, Ely EW, Wright TJ, Fraiz J, Stasek JE Jr, et al. (2001) Safety and dose relationship of recombinant human activated protein C for coagulopathy in severe sepsis. *Crit Care Med* 29: 2051–2059.
49. Abraham E, Laterre PF, Garg R, Levy H, Talwar D, et al. (2005) Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med* 353: 1332–1341.
50. Bernard GR, Margolis BD, Shanies HM, Ely EW, Wheeler AP, et al. (2004) Extended evaluation of recombinant human activated protein C United States Trial (ENHANCE US): a single-arm, phase 3B, multicenter study of drotrecogin alfa (activated) in severe sepsis. *Chest* 125: 2206–2216.
51. Barton P, Kalil AC, Nadel S, Goldstein B, Okhuysen-Cawley R, et al. (2004) Safety, pharmacokinetics, and pharmacodynamics of drotrecogin alfa (activated) in children with severe sepsis. *Pediatrics* 113: 7–17.
52. Nadel S, Goldstein B, Williams MD, Dalton H, Peters M, et al. (2007) Drotrecogin alfa (activated) in children with severe sepsis: a multicentre phase III randomised controlled trial. *Lancet* 369: 836–843.
53. Goldstein B, Nadel S, Peters M, Barton R, Machado F, et al. (2006) ENHANCE: results of a global open-label trial of drotrecogin alfa (activated) in children with severe sepsis. *Pediatr Crit Care Med* 7: 200–211.
54. Bennett IL, Finland M, Hamburger M, Kass EH, Lepper M, et al. (1963) The effectiveness of hydrocortisone in the management of severe infection. *JAMA* 183: 462–465.
55. Bone RC, Fisher CJ, Jr, Clemmer TP, Slotman GJ, Metz CA, et al. (1987) A controlled clinical trial of high-dose methylprednisolone in the treatment of severe sepsis and septic shock. *N Engl J Med* 317: 653–658.
56. Briegel J, Forst H, Haller M, Schelling G, Kilger E, et al. (1999) Stress doses of hydrocortisone reverse hyperdynamic septic shock: a prospective, randomized, double-blind, single-center study. *Crit Care Med* 27: 723–732.
57. Bollaert PE, Charpentier C, Levy B, Debouverie M, Audibert G, et al. (1998) Reversal of late septic shock with supraphysiologic doses of hydrocortisone. *Crit Care Med* 26: 645–650.
58. Luce JM, Montgomery AB, Marks JD, Turner J, Metz CA, et al. (1988) Ineffectiveness of high-dose methylprednisolone in preventing parenchymal lung injury and improving mortality in patients with septic shock. *Am Rev Respir Dis* 138: 62–68.
59. Lucas CE, Ledgerwood AM (1984) The cardiopulmonary response to massive doses of steroids in patients with septic shock. *Arch Surg* 119: 537–541.
60. Klastersky J, Cappel R, Debusscher L (1971) Effectiveness of betamethasone in management of severe infections. A double-blind study. *N Engl J Med* 284: 1248–1250.
61. Schumer W (1976) Steroids in the treatment of clinical septic shock. *Ann Surg* 184: 333–341.
62. Sprung CL, Caralis PV, Marcial EH, Pierce M, Gelbard MA, et al. (1984) The effects of high-dose corticosteroids in patients with septic shock. A prospective, controlled study. *N Engl J Med* 311: 1137–1143.
63. (1987) Effect of high-dose glucocorticoid therapy on mortality in patients with clinical signs of systemic sepsis. The Veterans Administration Systemic Sepsis Cooperative Study Group. *N Engl J Med* 317: 659–665.
64. Yildiz O, Doganay M, Aygen B, Guven M, Kelestimur F, et al. (2002) Physiological-dose steroid therapy in sepsis [ISRCTN36253388]. *Crit Care* 6: 251–259.
65. Thompson WL, Gurley HT, Lutz BA, Jackson DL, Kyols LK, et al. (1976) Inefficacy of glucocorticoids in shock (double-blind study) [Abstract]. *Clin Res* 6: 258A.
66. Sprung CL, Annane D, Keh D, Moreno R, Singer M, et al. (2008) Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 358: 111–124.
67. Loisa P, Parviainen I, Tenhunen J, Hovilehto S, Ruokonen E (2007) Effect of mode of hydrocortisone administration on glycemic control in patients with septic shock: a prospective randomized trial. *Crit Care* 11: R21.
68. Amato MB, Barbas CS, Medeiros DM, Magaldi RB, Schettino GP, et al. (1998) Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. *N Engl J Med* 338: 347–354.
69. Brower RG, Shanholtz CB, Fessler HE, Shade DM, White P Jr, et al. (1999) Prospective, randomized, controlled clinical trial comparing traditional versus reduced tidal volume ventilation in acute respiratory distress syndrome patients. *Crit Care Med* 27: 1492–1498.
70. Brochard L, Roudot-Thoraval F, Roupie E, Delclaux C, Chastre J, et al. (1998) Tidal volume reduction for prevention of ventilator-induced lung injury in acute respiratory distress syndrome. The Multicenter Trail Group on Tidal Volume reduction in ARDS. *Am J Respir Crit Care Med* 158: 1831–1838.
71. Stewart TE, Meade MO, Cook DJ, Granton JT, Hodder RV, et al. (1998) Evaluation of a ventilation strategy to prevent barotrauma in patients at high risk for acute respiratory distress syndrome. Pressure- and Volume-Limited Ventilation Strategy Group. *N Engl J Med* 338: 355–361.
72. de Durante G, del Turco M, Rustichini L, Cosimini P, Giunta F, et al. (2002) ARDSNet lower tidal volume ventilatory strategy may generate intrinsic positive end-expiratory pressure in patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 165: 1271–1274.
73. Villar J, Kacmarek RM, Perez-Mendez L, Aguirre-Jaime A (2006) A high positive end-expiratory pressure, low tidal volume ventilatory strategy improves outcome in persistent acute respiratory distress syndrome: a randomized, controlled trial. *Crit Care Med* 34: 1311–1318.
74. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, et al. (2004) Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med* 351: 327–336.
75. Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, et al. (2006) Intensive insulin therapy in the medical ICU. *N Engl J Med* 354: 449–461.
76. Brunkhorst FM, Engel C, Bloos F, Meier-Hellmann A, Ragaller M, et al. (2008) Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med* 358: 125–139.
77. De Block C, Manuel YKB, Van Gaal L, Rogiers P (2006) Intensive insulin therapy in the intensive care unit: assessment by continuous glucose monitoring. *Diabetes Care* 29: 1750–1756.
78. Clayton SB, Mazur JE, Condren S, Hermayer KL, Strange C (2006) Evaluation of an intensive insulin protocol for septic patients in a medical intensive care unit. *Crit Care Med* 34: 2974–2978.
79. Schultz MJ, Royakkers AA, Levi M, Moeniralam HS, Spronk PE (2006) Intensive insulin therapy in intensive care: an example of the struggle to implement evidence-based medicine. *PLoS Med* 3: e456.
80. Van den Berghe G, Wouters PJ, Bouillon R, Weekers F, Verwaest C, et al. (2003) Outcome benefit of intensive insulin therapy in the critically ill: Insulin dose versus glycemic control. *Crit Care Med* 31: 359–366.
81. Ellger B, Westphal M, Stubbe HD, Van den Heuvel I, Van Aken H, et al. (2008) [Glycemic control in sepsis and septic shock: Friend or foe?]. *Anaesthesist* 57: 43–48.
82. Mitchell I, Finfer S, Bellomo R, Hignett T (2006) Management of blood glucose in the critically ill in Australia and New Zealand: a practice survey and inception cohort study. *Intensive Care Med* 32: 867–874.
83. Lacherade JC, Jabre P, Bastuji-Garin S, Grimaldi D, Fangio P, et al. (2007) Failure to achieve glycemic control despite intensive insulin therapy in a medical ICU: incidence and influence on ICU mortality. *Intensive Care Med* 33: 814–821.
84. Toschlog EA, Newton C, Allen N, Newell MA, Goettler CE, et al. (2007) Morbidity reduction in critically ill trauma patients through use of a computerized insulin infusion protocol: a preliminary study. *J Trauma* 62: 1370–1375; discussion 1375–1376.
85. Gandhi GY, Nuttall GA, Abel MD, Mullany CJ, Schaff HV, et al. (2007) Intensive intraoperative insulin therapy versus conventional glucose management during cardiac surgery: a randomized trial. *Ann Intern Med* 146: 233–243.
86. Greenland S (2006) Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 35: 765–775.
87. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR (1999) Methods in health service research. An introduction to bayesian methods in health technology assessment. *Bmj* 319: 508–512.
88. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR (2000) Bayesian methods in health technology assessment: a review. *Health Technol Assess* 4: 1–130.
89. LaRosa SP (2006) Drotrecogin alfa (activated) in severe sepsis. *N Engl J Med* 354: 94–96; author reply 94–96.
90. Annane D (2003) Time for a consensus definition of corticosteroid insufficiency in critically ill patients. *Crit Care Med* 31: 1868–1869.
91. Siraux V, De Backer D, Yalavatti G, Melot C, Gervy C, et al. (2005) Relative adrenal insufficiency in patients with septic shock: comparison of low-dose and conventional corticotropin tests. *Crit Care Med* 33: 2479–2486.
92. Hill AB (1965) The Environment and Disease: Association or Causation? *Proc R Soc Med* 58: 295–300.