

Differences in outcome between the NICE-SUGAR and Leuven trials: possible methodological explanations

William R Henderson and Simon Finfer

Intensive insulin therapy (IIT), or intensive blood glucose control, is an area of intense research activity and clinical interest in critical care. The past decade has seen the publication of three large trials of IIT in critical illness — with conflicting results.

In 2001, the first large trial of IIT in critically ill patients found that targeting normoglycaemia (blood glucose concentration, 4.4–6.1 mmol/L) resulted in a significant reduction in both mortality and morbidity in patients admitted to a surgical intensive care unit in Leuven, Belgium.¹ A similar study in patients admitted to the medical ICU of the same hospital was published in 2006 and showed a beneficial effect of IIT only after subgroup analysis.² In 2009, the NICE-SUGAR Study Investigators published the largest trial of IIT to date and reported that targeting normoglycaemia increased mortality.³ Despite the similar aims, interventions and study populations of the three studies, these results are significantly different. Here we review some of the methodological differences between the trials and discuss whether these can explain the differences in reported outcomes.

Overall trial design

To reliably assess treatment effects, trials must be designed to provide strict control of bias and random error.⁴ Control of bias requires properly concealed randomisation and appropriate analysis, with the primary emphasis on the overall result rather than subgroup effects. Strict control of random error, in the current context, requires studies that report large numbers of deaths.⁴ In addition, except in very rare circumstances, trials should continue to their planned conclusion as early stopping may lead to misleading results.^{5,6}

Control of bias

Randomisation

Randomly assigning patients to one of two (or more) treatments within a clinical trial serves two important functions: firstly, if performed fastidiously, it serves to maintain allocation concealment; secondly, in large trials it ensures that both known and unknown confounders

(patient specific factors that may influence outcome) are equally distributed between the treatment groups. Allocation concealment means that those recruiting patients into a trial do not know to which treatment group a patient will be allocated until after the patient has been recruited. This is essential as knowledge of likely treatment allocation can lead to conscious or unconscious manipulation of the recruitment process.⁷

In the Leuven trials, treatments were allocated using sealed envelopes, stratified according to admission diagnosis and balanced with the use of permuted blocks of 10.^{1,2} In unblinded trials (and to date all trials of IIT have been open label), such randomisation sequences can only partially maintain allocation concealment. This is particularly so in single-centre trials where randomisation occurs on site, as in the Leuven trials.⁸ Investigators may introduce selection bias if they have influence over the number of patients eligible for enrolment, or the order in which they are randomised.^{9,10} Allocation strategies that use fixed patterns, such as the permuted blocks used in the Leuven trials, may be susceptible to deciphering by investigators or clinical staff.¹¹

A specific design feature of the NICE-SUGAR trial was to conduct randomisation via a centralised Internet server and not to stratify treatment allocation by centre. As there were 42 centres in the trial and no limit to the number of patients who could be allocated to either treatment in a given sequence at any individual centre, allocation concealment was maintained. Trials with multiple centres are at much less risk of the form of selection bias described above than trials performed at a single centre.

Appropriate analysis with emphasis on overall result

Bias can be introduced in the analysis and interpretation of trial data, particularly if undue emphasis is given to only a proportion of the patients recruited into the study— that is, to subgroup findings.^{12,13} In the Leuven trial in the surgical ICU, the beneficial effect of IIT was reported in the intention-to-treat population (all those patients allocated to one treatment or the other), although the treatment effect was of borderline statistical significance ($P=0.04$) after correction for multiple interim analyses. As considered below, it is also possible that the treatment effect was exaggerated as the trial was stopped early for benefit.⁶

In the Leuven trial in the medical ICU, there was no significant difference in outcome in the intention-to-treat

See also Myburgh and Chittock, page 178

population, and significant emphasis was given to the subgroup of patients who stayed in the ICU for 3 or more days.² In reporting their result, the authors also emphasised analyses within subgroups rather than reporting tests of interaction. This approach is considered more likely to produce biased estimates of treatment effects.¹²⁻¹⁵

The NICE-SUGAR Study Investigators broke new ground for a critical care trial by publishing their statistical analysis plan¹⁶ before completion of the trial. In doing so, they clearly identified pre-defined subgroups and placed maximum emphasis on the analysis of the treatment effect in the intention-to-treat population.

Strict control of random error

Designing trials with an expected large number of deaths

Although a randomised controlled trial is the best way to compare the relative effects of two or more treatment strategies, the results are still subject to the play of chance. The effect of chance diminishes as the number of patients suffering the outcome of interest (in this context, death) increases. Although it is widely recognised that small trials (trials that report a small number of outcomes) may have insufficient power to detect a treatment effect that exists, it is less well appreciated that small but positive trials are likely to be reporting exaggerated treatment effects or a difference in outcome that has arisen through chance (see Collins and MacMahon, "Problems with false positive results"⁴).

Continuing trials to their planned conclusion

Data monitoring committees are commonly established in large clinical trials to evaluate evidence of adverse events and unexpected benefit. Many trials are now designed with predetermined stopping rules that allow this committee to recommend the trial be stopped if there is early evidence of a large clinical benefit or of harm. The ethical rationale for this design is to allow early publication and dissemination of trial results and thereby benefit a larger pool of patients than would be included in the trial.

Unfortunately, there is substantial evidence that trials stopped early for perceived benefit are likely to overestimate clinical benefits and underestimate adverse events.⁶ These errors are more likely to occur in trials with a low rate of the outcome of interest, leading to the recommendation that trials are stopped early only if the *P* value for the comparison of the primary outcome is <0.001 and at least 200–400 events (deaths) have occurred.⁵ The Leuven surgical ICU trial, the only one of the three trials that demonstrated a clinical benefit in the intention-to-treat

population, was stopped early for benefit without meeting these criteria. The total numbers of deaths reported in the Leuven surgical ICU, medical ICU and NICE-SUGAR studies were 140, 442 and 1580, respectively. The small number of deaths in the Leuven surgical ICU study raises the concern that the reported benefit of IIT was caused by chance rather than true biological advantage.

Single-centre versus multiple-centre design: unique interventions

Did conducting the NICE-SUGAR Study in 42 centres in four countries compared with the Leuven trials, conducted in a single hospital, contribute to the difference in the results reported? Although results may be less generalisable to other settings, and the above concerns over introducing bias remain, the conduct of trials in a single centre may have advantages, particularly when applying complex treatment protocols. For example, the large numbers of patients enrolled in the Leuven trials may have allowed the bedside nurses to become highly proficient in the use of the insulin titration protocols. A concern in single-centre trials is that other treatments specific to the practice or culture of the site in question may influence the trial results.⁸ Most noticeably, there was a real difference in nutritional practice in Leuven compared with the hospitals in the NICE-SUGAR Study: the practice in Leuven is to use glucose-rich parenteral nutrition, while enteral nutrition was more commonly used in the NICE-SUGAR Study. This may mean that the results from the Leuven and NICE-SUGAR trials may be applicable in different settings dependent on the concomitant treatments used.

Conclusions

As the NICE-SUGAR Study was conducted in 42 hospitals, maintained allocation concealment, did not stratify by centre, followed a pre-defined statistical analysis plan and reported 1580 deaths, there is low risk that the results were affected by bias or random error. Of the trials in adults in Leuven, the surgical ICU trial was stopped for benefit after multiple interim analyses and reported only 140 deaths; the risk of random error may be substantial. Both trials in Leuven used a randomisation process in which allocation concealment may be only partially maintained. Together with the focus on subgroup results, this risks introducing recruitment, analysis and reporting bias. Although these methodological differences are real, whether they explain the variation in the reported results is likely to remain the subject of an argument that cannot be either proven or disproved.

Author details

William R Henderson, Assistant Professor¹

Simon Finfer, Professor,² and Senior Specialist³

1 Division of Critical Care Medicine, Faculty of Medicine, University of British Columbia, Vancouver, Canada.

2 George Institute for International Health and Sydney Medical School, University of Sydney, Sydney, NSW.

3 Intensive Care Unit, Royal North Shore Hospital, Sydney, NSW.

Correspondence: sfinfer@george.org.au

References

- 1 Van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001; 345: 1359-67.
- 2 Van den Berghe G, Wilmer A, Hermans G, et al. Intensive insulin therapy in the medical ICU. *N Engl J Med* 2006; 354: 449-61.
- 3 NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009; 360: 1283-97.
- 4 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001; 357: 373-80.
- 5 Mueller PS, Montori VM, Bassler D, et al. Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med* 2007; 146: 878-81.
- 6 Montori VM, Devereaux PJ, Adhikari NKJ, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; 294: 2203-9.
- 7 Lewis SC, Warlow CP. How to spot bias and other potential problems in randomised controlled trials. *J Neurol Neurosurg Psychiatry* 2004; 75: 181-7.
- 8 Forder PM, GebSKI VJ, Keech AC. Allocation concealment and blinding: when ignorance is bliss. *Med J Aust* 2005; 182: 87-9.
- 9 Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995; 274: 1456-8.
- 10 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; 359: 614-8.
- 11 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002; 136: 254-9.
- 12 Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine — reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007; 357: 2189-94.
- 13 Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005; 365: 1657-61.
- 14 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21: 2917-30.
- 15 Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001; 5: 1-56.
- 16 Finfer S, Heritier S; NICE Study Management Committee, Sugar Study Executive Committee. The NICE-SUGAR (Normoglycaemia in Intensive Care Evaluation and Survival Using Glucose Algorithm Regulation) Study: statistical analysis plan. *Crit Care Resusc* 2009; 11: 46-57. □