Best Practice & Research Clinical Anaesthesiology 23 (2009) 487–496

Contents lists available at ScienceDirect

Best Practice & Research Clinical Anaesthesiology

journal homepage: www.elsevier.com/locate/bean

11

How to weigh the current evidence for clinical practice

Andrew Padkin, BSc, MSc, MB.ChB, FRCA, MRCP, Consultant in Anaesthesia and Intensive Care Medicine *

Royal United Hospital Bath NHS Trust, Combe Park, Bath BA1 3NG, UK

Keywords:
intensive care
insulin
glucose
hyperglycaemia
internal validity
external validity
generalisability

This article presents a template for judging trials of tight glucose control in critically ill patients. It reviews threats to both internal validity and generalisability using examples from the current literature. When judging internal validity, it is important to consider factors specific to trials of glucose control (particularly the methods of glucose control, measurement and reporting) in addition to factors common to all randomised controlled trials (such as treatment allocation, losses to follow-up and protocol violations). Judging generalisability requires the identification of differences between the trial population and the population for whom the intervention is being considered. These may relate to the setting, the patients or the practical delivery of tight glucose control or other interventions. Once identified, a judgement must be made for each difference of whether it is likely to modify the effect of tight glucose control.

Since the publication of Van den Berghe's landmark trial reporting the results of tight glycaemic control (TGC) in surgical intensive care unit (SICU) patients in Leuven[1], there has been an explosion in the numbers of publications on this subject. There are now at least 26 completed randomised controlled trials (RCTs) investigating TGC in different populations of critically ill patients.[2] Integrating the results from different studies of TGC in the critically ill is currently extremely difficult. Studies with superficially similar methods have produced contradictory results and expert opinion has been polarised, with some urging caution before implementing TGC whist others have suggested it to be a standard of care and a quality marker for intensive care units (ICUs).

* Tel.: +44 1225 825056; Fax: +44 1225 825061.
E-mail address: andrew.padkin@ruh.nhs.uk

1521-6896/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bpa.2009.08.007

Practicing clinicians cannot wait for these debates to be concluded and need to decide what to do for the patients who are present in their ICUs today. To do this, they must have an understanding of how to weigh the current evidence for TGC.

This article discusses the methodological and practical issues to be considered when assessing published trials of TGC in the critically ill. Individual trials will not be considered in detail though examples will be used, principally taken from the three largest published trials: the Normoglycemia in Intensive Care Evaluation–Survival Using Glucose Algorithm Regulation (NICE-SUGAR) study[3], the Leuven SICU trial[1] and the Leuven medical ICU (MICU) trial.[4]

Key considerations

There are two key considerations when assessing published trials. These are:

1. Does the trial have internal validity, that is, was it conducted in a way which minimised bias–'Is it likely that the reported results are valid for the patients in the trial?'
2. Does the trial have external validity (generalisability), that is, 'Are the results from the trial likely to be repeated if the intervention is used for other patients?'

These will be discussed separately.

Internal validity

The principal aim of a clinical trial is to provide an unbiased estimate of the effect of using a defined intervention in a defined patient group. It is accepted that the best way of doing this is to compare the outcomes of a group of patients receiving the intervention with a similar group who did not, and to assign patients to each group so that potential confounding factors (e.g., age, sex and severity of illness) are balanced between groups.[5]

For trials of interventions where the absolute risk reduction is small, it is particularly important to avoid any bias as even relatively small biases could reduce or negate the treatment effect.

In the initial Leuven SICU trial[1], the absolute risk reduction in ICU mortality was 3.4% and the risk reduction in hospital mortality was 3.7%. These relatively small reductions in risk suggest that when considering trials of TGC only the highest quality evidence should be used. This is provided by well-conducted RCTs. Non-randomised studies that use matching, risk adjustment or stratification in the analysis should be treated with caution. Although these studies may have a place in some situations[6], they have a high risk of bias due to residual confounding.

There are now numerous publications in the evidence-based medicine literature providing general guidance on issues that can affect the internal validity of RCTs, and the CONSORT group has made reporting recommendations.[7,8] These general issues are considered prior to issues specific to TGC.

General issues affecting internal validity of RCTs

At the trial's conclusion, all patients who entered the trial should be properly accounted for and correctly attributed to the group they were randomised to.[9] This requires complete follow-up of all patients.

As seen above, TGC produces a relatively small reduction in the absolute risk of ICU and hospital death when used for mechanically ventilated patients. Complete follow-up is particularly important in this situation as even small numbers of patients being lost to follow-up or withdrawing consent can have a profound effect on the trial result if their outcomes differ from those of the patients remaining in the trial. In the NICE-SUGAR study, the largest trial of TGC, there was only a difference of 78 deaths between the treatment and the control groups and 82 patients were lost to follow-up or withdrew consent.

It has been suggested that the robustness of a trial's results can be tested firstly by re-analysing the results assuming that all losses to follow-up survived, then secondly reanalysing the results assuming that all losses to follow-up died.[9] These sensitivity analyses would show whether the trial result would

be changed by a difference in the frequency of death between those patients for whom data are available and for those unavailable, although these analyses are seldom presented.

For similar reasons, it is important that protocol violations are minimised so that patients assigned to a treatment group actually get that treatment. If they do not, then bias can be introduced to a trial. To allow judgements to be made of whether this is likely, trial reports should include the number of protocol violations and why they occurred in each group. Very few trials do this completely. The NICE-SUGAR study report is one of the best, reporting 529 patients who discontinued the allocated intervention. Of these, the 231 who were changed to palliative care during the trial are almost completely balanced between the groups and seem very unlikely to have biased the results, but the 163 who were withdrawn on physician's request are unbalanced (115 in the TGC group and 48 in the control group) and raise the question of whether this may have introduced bias.

A final general concern in TGC trials is that treating clinicians are often not blinded to group allocation as they need knowledge of a patient's blood glucose to deliver safe care. This makes the trials particularly vulnerable to bias due to the use of different interventions apart from TGC (called co-interventions) in the treatment and control groups. Trial reports should provide sufficient detail of the co-interventions used in both the treatment and control groups to allow reviewers to judge whether differences have occurred. If differences are found, then the reason for the differences and whether this affects the internal validity of the trial must be considered. For example, in the NICE-SUGAR study, an increased use of corticosteroid use was reported in the TGC group compared with the control group. The trial report provides details of the reasons for this steroid use, and the principal imbalance is in patients with septic shock. The reasons for this are currently unclear, but reviewers must decide whether it is due to an imbalance that occurred by chance during the randomisation process (which may threaten internal validity but does not appear likely from the baseline characteristics published), whether there were other differences in co-interventions that may have resulted in greater steroid use in the TGC group (which again would threaten internal validity) or whether reducing blood glucose levels by the administration of insulin has adverse effects on the cardiovascular system as suggested by the NICE-SUGAR authors[10] (which would not threaten internal validity). Whilst waiting for further studies to be published to tease apart these possibilities, knowledgeable clinicians familiar with the clinical management of patients in settings similar to those in the trial must make decisions about which possibility is most likely.

Specific issues affecting internal validity

In TGC trials, the measurement and reporting of glucose control and the method used for achieving glucose control are often incompletely reported. Bias can be introduced in any or all of these three areas:

Measurement of blood glucose. Achieving TGC requires the accurate, frequent measurement of blood glucose. Inaccuracies will result in the malfunction of algorithms used for insulin dosing (so making it more difficult to achieve blood glucose control), increase the frequency of unrecognised real hypoglycaemic episodes, increase the frequency of false hypoglycaemic episodes and reduce the accuracy of summary measures of blood glucose. These problems may not be recognised by those running RCTs of TGC unless they are actively sought.

To aid in judging whether a trial used accurate blood glucose measurements, both the site of blood sampling and the equipment used for blood glucose measurement should be reported.

Site of blood sampling for glucose measurement. Trials of tight glucose control in the critically ill have used intermittent blood sampling from arterial, venous or capillary sites. A significant body of evidence suggests that capillary blood glucose is not a reliable reflection of whole blood glucose in critically ill patients, especially those with hypoperfusion[11–18], so the internal validity of trials that have made frequent use of this sampling method may be compromised. In practice, many trials state that more than one site has been used but do not report the frequency of measurements made at each site, making judgements about internal validity almost impossible.

Equipment used for blood glucose measurement. Trials have used point-of-care testing meters, blood gas analysers and laboratory testing for measuring blood glucose. Some of the issues raised by different

glucose measurement methods have been covered in depth elsewhere in this issue of *Best Practice and Research Clinical Anaesthesiology*.

It is important to know what equipment has been used in the trials, and whether the glucose values reported in the trials that used point-of-care glucometers denote the whole blood glucose concentration or the plasma concentration. These values differ by approximately 10%, and some glucometers can display either. To get a valid trial result, all centres in a trial should be using the same measure.

The measurement method used by the glucometer should also be understood. The commonly used methods are based on glucose dehydrogenase pyrroloquinolinequinone (GDH-PQQ), glucose dehydrogenase nicotinamide adenine dinucleotide (GDH-NAD), glucose hexokinase or glucose oxidase. For each of these, there are factors that can interfere with the glucose measurement. Of particular practical importance is the potential for falsely elevated glucose readings from GDH-PQQ-based methods. This technique cannot distinguish glucose from maltose, galactose or xylose. Maltose is present in some intravenous immunoglobulin solutions and is also the metabolic product of icodextrin, present in some peritoneal dialysis solutions. Adverse outcomes[19–21] have occurred due to hypoglycaemia when falsely elevated blood glucose readings have resulted in inappropriate insulin treatment. Other substances that have been reported to interfere with glucose measurements include acetaminophen (paracetamol)[22], mannitol, dopamine and levodopa.[23] Serious adverse events (SAEs) due to hypoglycaemia are likely to occur more frequently in the tight glucose control group, and may occur due to lack of understanding of the limitations of the glucometer being used. For trials reporting frequent episodes of hypoglycaemia or SAEs due to hypoglycaemia, it is important to investigate whether GDH-PQQ-based glucometers were used and whether their limitations were understood by the participating centres.

Inaccuracy in point-of-care glucometers can also adversely affect validity. There have recently been an increasing number of publications assessing this in the critically ill.[11,14,17,24,25] Some of these have been extremely useful in pointing out the pitfalls in using point-of-care testing, although this literature should be read critically as the quality of these evaluation studies is variable. A recent review suggested that few conform to the 20 recommendations of the Standards for Reporting Diagnostic Accuracy (STARD) and the 18 recommendations from the Clinical and Laboratory Standards Institute (CLSI).[26]

Although the laboratory measurement of blood glucose provides a gold standard for accuracy, the delay involved in sending blood to the laboratory makes this method less valuable for practical use. If laboratory values alone are used for measuring blood glucose in trials, there may be unacceptable delays in treating hypo- or hyperglycaemia that may themselves introduce a threat to internal validity.

As can be seen above, for any TGC trial, the details of how the investigators performed blood glucose measurements are extremely important. It has been suggested that variability in measurement techniques may explain some of the differences seen among different studies of TGC.[27] It is vital that future research is directed towards improving the measurement of blood glucose in critically ill patients so that accurate, reliable and timely methods can be used in future studies.

Reporting blood glucose control. It is not only important to measure blood glucose accurately, but also to summarise it using a valid method. Unfortunately, there is no agreement on the gold standard. A recent systematic review identified 30 different summary indicators that have been used.[28] These measures can be categorised as follows:

a. Reports of a summary measure of a single measured blood glucose value. The value reported is often derived from a single morning blood glucose measurement from each patient as in the Leuven SICU paper.[1] Using this measure loses all information about what occurs during the other periods of the day.

b. Reports of a summary value (usually mean or median) of all measured blood glucose values. Some reports summarise all blood glucose values measured in all patients in a group being studied, some summarise blood glucose values for each individual patient and then report a summary of these. Both methods can introduce bias into trials due to

i sampling frequency (the summary measure will be biased towards those values at the time when blood glucose was sampled more frequently). This may be particularly problematic if the sampling frequency is different between the TGC and control groups.

ii unit admission policy (the more patients admitted with low severity of illness, the more the glucose summary measure will be biased towards normal).
iii unit discharge policy (if patients are kept on ICU after the period of recovery from acute physiological derangement then their blood glucose will have returned towards normal, biasing the summary measure towards normal). This may be particularly important if the use of TGC leads to changes in a patient's length of ICU stay following recovery from critical illness.

A further potential problem is that low and high blood glucose values may cancel each other out, so a patient with widely fluctuating blood glucose may have the same mean value as a patient with completely stable control.

To attempt to overcome some of the problems mentioned above, reports have used summaries of blood glucose for the times before, during or after insulin treatment, for the time after the target range was achieved, during defined time periods during which insulin was being used or for just the patients for whom insulin was used.

c. Reports using range-based measurements. The summary value reported is usually either the number/percentage of blood glucose values within a predefined range or ranges, the time within these ranges or more complex range-based measures, for example, hyperglycaemic index (the area under the curve above the upper limit of normal (glucose level 6.0 mmol l^{-1}) divided by the total length of stay)[29] or the glycaemic penalty index.[30]

In common with the summary measures above, many range-based measurements can have bias introduced by sampling frequency and unit admission and discharge policies.

d. Reporting adverse events. It is a common feature of TGC trials that the patients getting TGC have an increased frequency of hypoglycaemia. Whether this increased frequency of hypoglycaemia causes harm is still debated.[31–35] Attitudes to hypoglycaemia occurring in RCTs reflects this uncertainty: the VISEP[36] and GLUCONTROL trials were stopped early due to the increased rate of hypoglycaemia in the TGC group[36] despite this rate being lower than that seen in the Leuven MICU trial[4], which was not stopped. When reporting hypoglycaemia, a variety of definitions have been used: <40 mg dl^{-1} (2.2 mmol l^{-1}) is the most common, although others have reported a range up to <72 mg dl^{-1} (4.0 mmol l^{-1}).[28] Studies have generally reported the total number of hypoglycaemic episodes occurring or the number (%) of patients having a hypoglycaemic episode, and these values are sometimes presented per patient per day or for the whole study period.[28]

Due to the wide variety of measures that have been used and the potential for bias that exists for each, it is difficult to compare the glucose control achieved or the adverse events occurring in different studies (and in some cases it is difficult to compare between different groups in the same study).

As understanding of the biological basis for the adverse effects of hyperglycaemia improves, it is likely to become clearer which summary measure best correlates with patient outcome and this will become the standard measure used. It may be that none of the measures described above is ideal: based on observations of the adverse biological effects of rapid changes in blood glucose concentrations, it has been suggested that blood glucose variability may be better associated with adverse outcome than the indicators above, and there are empirical data to support this.[37] At least six different measures of variability have already been published[37–43]– although there is currently no evidence about which of these correlates best with outcome, whether medical interventions can reduce glucose variability or whether improving variability results in improved outcomes.

Controlling blood glucose. Exogenous insulin is needed to control hyperglycaemia. Most ICUs give this as a continuous intravenous infusion. In deciding the infusion rate of insulin, some ICUs use a nurse-led strategy based on non-prescriptive guidelines. Others have developed more didactic protocols that can be generally classified as sliding scales or dynamic protocols. Sliding scales (where a particular glucose reading leads to the infusion of insulin at a prescribed rate) generally result in moderate-to-poor glucose control.[44] Dynamic protocols use the rate of change of glucose in addition to the most recently measured value to decide on the insulin infusion rate, and generally give better results for glycaemic control.[44] More sophisticated computer-based algorithms have also been used.[45] These protocols are hard to compare because of differences between studies in evaluation strategies, measurement of blood glucose and the summary measures used for glucose control[45] (discussed above). Therefore, it is

not possible to make definitive recommendations of whether the use of any particular type of insulin protocol is likely to compromise the internal validity of a clinical trial.

This section has discussed potential threats to the internal validity of clinical trials of TGC in the critically ill. Identifying the potential for these methodological weaknesses is relatively simple. Problems occur in trying to decide to what extent the potential weaknesses have actually occurred in any specific trial. Unfortunately, published meta-analyses of TGC[2,46,47] have made no attempt to judge the specific threats to internal validity described here, instead, using a generic score (e.g., the Jadad scale[48]) to judge the quality of RCTs.

Future research work is needed to define optimum methods of controlling, measuring and reporting blood glucose. This research agenda will be helped greatly by the introduction of accurate and reliable continuous (or near-continuous) blood glucose measurement devices, which will not only allow detailed analyses of the relationship between blood glucose and outcome but also aid the introduction of semi-closed or closed-loop glucose control systems. Currently, these monitors are still at the research stage, although many different commercial systems may be available in the near future. They are likely to revolutionise our understanding of glucose control in the critically ill and are also likely to improve the internal validity of TGC trials.

External validity (Generalisability)

If a trial is considered to have adequate internal validity, the next important question is: Are the results from the trial likely to be repeated if the intervention is used for other patients? If, on the other hand, the internal validity is thought to be poor it is irrational to consider whether results should be generalised.

Judging generalisability should be a two-step process. The first step is to identify differences between the original trial and a target population for whom the intervention is being considered. The second step is considerably harder and requires consideration of whether the effect of the intervention is likely to be modified because of any of these differences.[49,50] Identified differences are frequently referred to in reviews of TGC trials, but few reviews discuss making the judgement on whether individual differences are likely to modify the effect of TGC. This is not altogether surprising as the judgement is complex and has to be made based on biological plausibility, basic science, other published work and a detailed understanding of the clinical situation. It is entirely possible that well-informed experts can hold opposing views on whether identified differences are likely to limit the generalisability of a trial's results.

Factors to be considered when identifying differences between the original trial population and a target population include [50–52]:

The setting

Differences may be present in health-care systems, ethnic, socioeconomic or other factors between the trial setting and the setting in which the target population will be treated.

Practical considerations such as nursing staffing levels, clinician motivation and education and the availability of appropriate glucometers/blood gas analysers can all affect the ability of an ICU to deliver TGC safely and effectively in routine practice.[53–56]

The patients studied

Differences may occur between trial patients and the target population in primary diagnosis, severity of illness or co-morbidities.

Eligibility and exclusion criteria limit the patients recruited into a trial. In addition to this, failure to recruit eligible patients, either because they were not given the chance to participate or because they declined to participate, can lead to patients in a trial being considerably different to the patients the trial set out to recruit and also different to patients in a target population.[57–59] Whilst this is unlikely to affect internal validity, it can make judging generalisability more difficult and requires reviewers to take careful note of the baseline characteristics of trial patients. Trials often do not report what percentage of eligible patients were recruited, and those that do can have widely varying rates: in the Leuven SICU trial[1], randomisation was achieved for 99% of eligible patients, whereas only 15% of those

assessed for eligibility in the NICE-SUGAR study were randomised.[10] The reporting of both these trials is exemplary as not only do they report the numbers of patients assessed but not randomised, but also report the reasons for which randomisation did not occur, giving reviewers further information on which to judge whether differences exist between the trial and target populations.

In the Leuven SICU trial, over 60% were cardiac surgery patients.[1] This difference in diagnosis between the Leuven SICU patients and the general ICU patients in other places has been identified frequently in the literature, although few have presented an argument about how far this should limit the generalisabity of the results to other ICU patients.

Treatments provided

To achieve results similar to those achieved in clinical trials when an intervention is used in routine patient care requires best practice to be used in non-research settings. When using TGC, it must be delivered to the target population in the same way as it was delivered to the trial population.

As discussed previously in this section and in the section on internal validity, issues associated with the practical provision of TGC can limit the ability to deliver TGC safely and effectively. Difficulties with controlling, measuring and reporting blood glucose can be challenging even in a trial setting. The practicality of delivering TGC during routine care should not be underestimated, and failure to deliver TGC during routine care will limit the generalisability of research results.

In any RCT, the treatment and intervention groups should both receive care (apart from the intervention itself) that would be considered standard practice. One of the major criticisms of the Leuven SICU study is that a feeding regimen was used that many did not consider to be the standard.[60] Intravenous glucose was administered routinely in the first 24 h followed by the early use of total parenteral nutrition. This was not used in the Leuven MICU trial or the NICE-SUGAR study.

The use of intravenous glucose does not threaten the internal validity of the Leuven SICU study as the same co-intervention was used in both the groups. However, there have been concerns that it may limit the generalisability of the results. Although follow-up publications re-analysing data from the Leuven trials[61] suggested that the benefit of intensive insulin therapy was independent of parenteral glucose load[62], doubts still remain about whether the benefits of TGC only exist when intravenous glucose is given.[63]

Efficacy and effectiveness trials

A distinction has been made in the clinical trials literature between RCTs that investigate 'efficacy' and those that investigate 'effectiveness'. Trials of efficacy often recruit highly selected participants and provide a strictly controlled intervention, often in a well-resourced setting. They are designed to answer the question of whether the intervention can work when delivered ideally. These trials tend to show the maximum benefit possible from an intervention, but give less information about generalisability or what benefits would be expected if the intervention was used widely in routine clinical practice. On the other hand, effectiveness trials (also called pragmatic trials[64]) aim to recruit a much broader range of participants with the intervention provided as it would be in normal practice.[65,66] They are designed to answer questions of benefit, risk and cost if the intervention is used routinely. Because of the broader range of participants and the possible suboptimal delivery of the intervention, the effect size seen in effectiveness trials is unlikely to be as large as that seen in efficacy trials. Effectiveness trials may fail to find significant treatment effects when, in fact, they exist for some subgroups in the trial because of this dilution effect. Thus, effectiveness trials may be negative even if the intervention is effective for some subgroups of patients.

In the TGC literature, the Leuven trials[1,4] can be considered to be efficacy trials as they were performed in a single centre where glucose control was likely to be excellent and with tightly defined feeding and other protocols. The NICE-SUGAR study[10] is closer to an effectiveness trial as it recruited from many units, so glucose control and adherence to other protocols may have been more variable. Understanding this may reconcile the different results in the two trials. It is possible that both studies are internally valid, but that there is something about the Leuven SICU setting, the patients or the treatment provided that made TGC provide an overall benefit there when it was associated with harm in the wider setting of the NICE-SUGAR study. Further analysis of both the NICE-SUGAR and the Leuven trial data may help to identify these factors, and whether there were sub-groups of patients in the NICE-SUGAR study for whom TGC was effective.[2]

Identifying potential threats to the internal validity and generalisability of trials of TGC in the critically ill can be done using the criteria above. However, partly due to the lack of standardised methods of controlling, monitoring and reporting blood glucose in critically ill patients, it remains a matter of judgement to quantify the extent to which any of these potential threats introduced bias into an individual study. This is illustrated by two recent reviews that each provided a detailed commentary on the Leuven SICU study and the NICE-SUGAR study.[67,68] Each review identifies potential threats to internal validity and generalisability, but the threats identified and the importance attached to each is completely different in each review. These debates about the significance of methodological and practical issues that occur in TGC trials suggest that there is considerable scope for further research.

In summary, when weighing the evidence presented in a published trial of TGC, initial consideration should be given to factors affecting the internal validity of the trial. As in all RCTs, it is important to consider whether biases may have been introduced, particularly in the allocation of patients to treatment or control groups, by patients being lost to follow-up or by protocol violations. Important factors specific to trials of TGC that must be addressed are the methods of controlling blood glucose, of measuring blood glucose (both the site of glucose sampling and the equipment used being important) and the method used for reporting blood glucose control.

If the internal validity is considered adequate, the external validity (generalisability) can then be considered. Factors that may affect the external validity are differences between the trial population and the target population in setting, the patients for whom it will be used and the practical details of how TGC and other co-interventions will be provided. If differences exist in any of these factors between the trial and target populations, consideration should then be given as to whether these differences may alter the effect of TGC. It is particularly important that practicing clinicians engage in the debate on external validity because so many important considerations rely on expert knowledge of the clinical situation.

Practice points

- When weighing the evidence that a research report provides for or against tight glucose control the internal validity and generalisability of the report should be considered separately.
- Internal validity depends on factors common to all RCTs (ensuring random treatment allocation and minimising losses to follow-up and protocol violations) and also factors specific to trials of glucose control (particularly the methods of glucose control, measurement and reporting).
- Generalisability depends on judging differences between the trial population and the target population (in setting, patient factors or the practical delivery of TGC or other interventions) and then considering if these differences will modify the effect of TGC.

Research agenda

- Research is urgently needed to define optimum methods of controlling, measuring and reporting blood glucose in critically ill patients.
- An accurate, reliable continuous (or near-continuous) blood glucose measurement device will not only improve our understanding of the relationship between blood glucose and outcome but will also lead to systems that allow better glucose control for individual patients.

Conflict of interest

The author has provided consultancy to 3i and Glysure.

References

*1. Van den Berghe G, Wouters P, Weekers F et al. Intensive insulin therapy in critically ill patients. *New England Journal of Medicine* 2001; **345:** 1359–1367.
2. Griesdale DE, de Souza RJ, van Dam RM et al. Intensive insulin therapy and mortality among critically ill patients: a meta-analysis including NICE-SUGAR study data. *CMAJ (Canadian Medical Association Journal)* 2009; **180:** 821–827.
3. Finfer S & Delaney A. Tight glycemic control in critically ill adults. *Journal of the American Medical Association* 2008; **300:** 963–965.
*4. Van den Berghe G, Wilmer A, Hermans G et al. Intensive insulin therapy in the medical ICU. *New England Journal of Medicine* 2006; **354:** 449–461.
5. Byar DP, Simon RM, Friedewald WT et al. Randomized clinical trials. Perspectives on some recent ideas. *New England Journal of Medicine* 1976; **295:** 74–80.
6. Black N. Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal* 1996; **312:** 1215–1218.
*7. Moher D, Schulz KF & Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association* 2001; **285:** 1987–1991.
8. CONSORT Group. CONSORT. Transparent reporting of Trials. 15/6/2009. Available from: http://www.consort-statement.org/home.
9. Guyatt GH, Sackett DL & Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1993; **270:** 2598–2601.
*10. Finfer S, Chittock DR, Su SY et al. Intensive versus conventional glucose control in critically ill patients. *New England Journal of Medicine* 2009; **360:** 1283–1297.
11. Kanji S, Buffie J, Hutton B et al. Reliability of point-of-care testing for glucose measurement in critically ill adults. *Critical Care Medicine* 2005; **33:** 2778–2785.
12. Atkin SH, Dasmahapatra A, Jaker MA et al. Fingerstick glucose determination in shock. *Annals of Internal Medicine* 1991; **114:** 1020–1024.
13. Critchell CD, Savarese V, Callahan A et al. Accuracy of bedside capillary blood glucose measurements in critically ill patients. *Intensive Care Medicine* 2007; **33:** 2079–2084.
14. Petersen JR, Graves DF, Tacker DH et al. Comparison of POCT and central laboratory blood glucose results using arterial, capillary, and venous samples from MICU patients on a tight glycemic protocol. *Clinica Chimica Acta* 2008; **396:** 10–13.
15. Kulkarni A, Saxena M, Price G et al. Analysis of blood glucose measurements using capillary and arterial blood samples in intensive care patients. *Intensive Care Medicine* 2005; **31:** 142–145.
16. Slater-MacLean L, Cembrowski G, Chin D et al. Accuracy of glycemic measurements in the critically ill. *Diabetes Technology & Therapeutics* 2008; **10:** 169–177.
17. Desachy A, Vuagnat AC, Ghazali AD et al. Accuracy of bedside glucometry in critically ill patients: influence of clinical characteristics and perfusion index. *Mayo Clinic Proceedings* 2008; **83:** 400–405.
18. Sylvain HF, Pokorny ME, English SM et al. Accuracy of fingerstick glucose values in shock patients. *The American Journal of Critical Care* 1995; **4:** 44–48.
*19. U.S. Food and Drug Administration Office of Vitro Diagnostic Device Evaluation and Safety. FDA reminders for falsely elevated glucose readings from use of inappropriate test method. Available from: http://www.fda.gov/cdrh/oivd/news/glucosefalse.html; 2005 [cited 2009 22/05/09].
20. Institute for Safe Medication Practices. Be aware of false glucose results with point-of-care testing. Available from: http://www.ismp.org/newsletters/acutecare/articles/20050908.asp; 2005 [cited 2009 22/5/2009].
21. Korsatko S, Ellmerer M, Schaupp L et al. Hypoglycaemic coma due to falsely high point-of-care glucose measurements in an ICU-patient with peritoneal dialysis: a critical incidence report. *Intensive Care Medicine* 2009; **35:** 571–572.
22. Cartier LJ, Leclerc P, Pouliot M et al. Toxic levels of acetaminophen produce a major positive interference on Glucometer Elite and Accu-chek Advantage glucose meters. *Clinical Chemistry* 1998; **44:** 893–894.
23. Tang Z, Du X, Louie RF & Kost GJ. Effects of drugs on glucose measurements with handheld glucose meters and a portable glucose analyzer. *American Journal of Clinical Pathology* 2000; **113:** 75–86.
24. Finkielman JD, Oyen LJ & Afessa B. Agreement between bedside blood and plasma glucose measurement in the ICU setting. *Chest* 2005; **127:** 1749–1751.
25. Mann EA, Pidcoke HF, Salinas J et al. Accuracy of glucometers should not be assumed. *The American Journal of Critical Care* 2007; **16:** 531–532. author reply 532.
26. Mahoney JJ & Ellison JM. Assessing glucose monitor performance–a standardized approach. *Diabetes Technology & Therapeutics* 2007; **9:** 545–552.
*27. Fahy BG & Coursin DB. Critical glucose control: the devil is in the details. *Mayo Clinic Proceedings* 2008; **83:** 394–397.
28. Eslami S, de Keizer NF, de Jonge E et al. A systematic review on quality indicators for tight glycaemic control in critically ill patients: need for an unambiguous indicator reference subset. *Critical Care* 2008; **12:** R139.
29. Vogelzang M, van der Horst IC & Nijsten MW. Hyperglycaemic index as a tool to assess glucose control: a retrospective study. *Critical Care* 2004; **8:** R122–127.
30. Van Herpe T, De Brabanter J, Beullens M et al. Glycemic penalty index for adequately assessing and comparing different blood glucose control algorithms. *Critical Care* 2008; **12:** R24.
31. Vriesendorp TM, van Santen S, DeVries JH et al. Predisposing factors for hypoglycemia in the intensive care unit. *Critical Care Medicine* 2006; **34:** 96–101.
32. Krinsley JS & Grover A. Severe hypoglycemia in critically ill patients: risk factors and outcomes. *Critical Care Medicine* 2007; **35:** 2262–2267.
33. Vriesendorp TM, DeVries JH, van Santen S et al. Evaluation of short-term consequences of hypoglycemia in an intensive care unit. *Critical Care Medicine* 2006; **34:** 2714–2718.

34. Mechanick JI, Handelsman Y & Bloomgarden ZT. Hypoglycemia in the intensive care unit. *Current Opinion in Clinical Nutrition and Metabolic Care* 2007; **10:** 193–196.
35. Vriesendorp TM, DeVries JH & Hoekstra JB. Hypoglycemia and strict glycemic control in critically ill patients. *Current Opinion in Critical Care* 2008; **14:** 397–402.
*36. Brunkhorst FM, Engel C, Bloos F et al. Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *New England Journal of Medicine* 2008; **358:** 125–139.
*37. Egi M, Bellomo R, Stachowski E et al. Variability of blood glucose concentration and short-term mortality in critically ill patients. *Anesthesiology* 2006; **105:** 244–252.
38. Wintergerst KA, Buckingham B, Gandrud L et al. Association of hypoglycemia, hyperglycemia, and glucose variability with morbidity and death in the pediatric intensive care unit. *Pediatrics* 2006; **118:** 173–179.
39. Ali NA, O'Brien Jr. JM, Dungan K et al. Glucose variability and mortality in patients with sepsis. *Critical Care Medicine* 2008; **36:** 2316–2321.
40. Kovatchev BP, Otto E, Cox D et al. Evaluation of a new measure of blood glucose variability in diabetes. *Diabetes Care* 2006; **29:** 2433–2438.
41. Krinsley JS. Glycemic variability: a strong independent predictor of mortality in critically ill patients. *Critical Care Medicine* 2008; **36:** 3008–3013.
42. Dossett LA, Cao H, Mowery NT et al. Blood glucose variability is associated with mortality in the surgical intensive care unit. *American Surgeon* 2008; **74:** 679–685. discussion 685.
43. Hirshberg E, Larsen G & Van Duker H. Alterations in glucose homeostasis in the pediatric intensive care unit: hyperglycemia and glucose variability are associated with increased mortality and morbidity. *Pediatric Critical Care Medicine* 2008; **9:** 361–366.
*44. Meijering S, Corstjens AM, Tulleken JE et al. Towards a feasible algorithm for tight glycaemic control in critically ill patients: a systematic review of the literature. *Critical Care* 2006; **10:** R19.
45. Van Herpe T, De Moor B & Van den Berghe G. Towards closed-loop glycaemic control. *Best Practice & Research. Clinical Anaesthesiology* 2009; **23:** 69–80.
46. Pittas AG, Siegel RD & Lau J. Insulin therapy and in-hospital mortality in critically ill patients: systematic review and meta-analysis of randomized controlled trials. *JPEN (Journal of Parenteral and Enteral Nutrition)* 2006; **30:** 164–172.
47. Wiener RS, Wiener DC & Larson RJ. Benefits and risks of tight glucose control in critically ill adults: a meta-analysis. *Journal of the American Medical Association* 2008; **300:** 933–944.
48. Jadad AR, Moore RA, Carroll D et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials* 1996; **17:** 1–12.
49. Horton R. Common sense and figures: the rhetoric of validity in medicine. Bradford Hill Memorial Lecture 1999. *Statistics in Medicine* 2000; **19:** 3149–3164.
*50. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005; **365:** 82–93.
51. Weiss NS, Koepsell TD & Psaty BM. Generalizability of the results of randomized trials. *Archives of Internal Medicine* 2008; **168:** 133–135.
52. Pitt B, Julian DG & Pocock SJ. *Clinical trials in cardiology*. London: Saunders, 1997.
53. Meynaar IA, Dawson L, Tangkau PL et al. Introduction and evaluation of a computerised insulin protocol. *Intensive Care Medicine* 2007; **33:** 591–596.
54. Preston S, Laver SR, Lloyd W & Padkin A. Introducing intensive insulin therapy: the nursing perspective. *Nursing in Critical Care* 2006; **11:** 75–79.
55. Goldberg PA, Siegel MD, Sherwin RS et al. Implementation of a safe and effective insulin infusion protocol in a medical intensive care unit. *Diabetes Care* 2004; **27:** 461–467.
56. Clayton SB, Mazur JE, Condren S et al. Evaluation of an intensive insulin protocol for septic patients in a medical intensive care unit. *Critical Care Medicine* 2006; **34:** 2974–2978.
57. Heiat A, Gross CP & Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Archives of Internal Medicine* 2002; **162:** 1682–1688.
58. Murthy VH, Krumholz HM & Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *Journal of the American Medical Association* 2004; **291:** 2720–2726.
59. Dhruva SS & Redberg RF. Variations between clinical trial participants and Medicare beneficiaries in evidence used for Medicare national coverage decisions. *Archives of Internal Medicine* 2008; **168:** 136–140.
60. Egi M, Bellomo R, Stachowski E et al. Intensive insulin therapy in postoperative intensive care unit patients: a decision analysis. *American Journal of Respiratory and Critical Care Medicine* 2006; **173:** 407–413.
61. Van den Berghe G, Wilmer A, Milants I et al. Intensive insulin therapy in mixed medical/surgical intensive care units: benefit versus harm. *Diabetes* 2006; **55:** 3151–3159.
62. Vanhorebeek I, Langouche L & Van den Berghe G. Tight blood glucose control: what is the evidence? *Critical Care Medicine* 2007; **35:** S496–502.
63. Marik PE & Varon J. Intensive insulin therapy in the ICU: is it now time to jump off the bandwagon? *Resuscitation* 2007; **74:** 191–193.
64. Schwartz D & Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Disease* 1967; **20:** 637–648.
65. Zwarenstein M, Treweek S, Gagnier JJ et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *British Medical Journal* 2008; **337:** a2390.
66. Tunis SR, Stryer DB & Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association* 2003; **290:** 1624–1632.
67. Van den Berghe G, Schetz M, Vlasselaers D et al. Intensive insulin therapy in critically ill patients: NICE-SUGAR or Leuven blood glucose target? *Journal of Clinical Endocrinology & Metabolism* 2009; **94:** 3163–3170.
68. Bellomo R & Egi M. What is a NICE-SUGAR for patients in the intensive care unit? *Mayo Clinic Proceedings* 2009; **84:** 400–402.