

Review

Bench-to-bedside review: The evaluation of complex interventions in critical care

Anthony Delaney¹, Derek C Angus², Rinaldo Bellomo³, Peter Cameron⁴, D James Cooper⁵, Simon Finfer⁶, David A Harrison⁷, David T Huang², John A Myburgh⁸, Sandra L Peake⁹, Michael C Reade¹⁰, Steve AR Webb¹¹, Donald M Yealy¹² for the Australian Resuscitation in Sepsis Evaluation (ARISE), Protocolized Care for Early Septic Shock (ProCESS) and Protocolised Management In Sepsis (ProMISe) investigators

¹Northern Clinical School, Faculty of Medicine, University of Sydney, Intensive Care Unit, Royal North Shore Hospital, Pacific Highway, St Leonards, NSW, 2065, Australia

²Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

³Department of Intensive Care, Austin Hospital, Heidelberg, Victoria, Australia

⁴Department of Epidemiology and Preventive Medicine, Monash University and Alfred Hospital, Melbourne, Australia

⁵Department of Intensive Care Medicine, Monash University and Alfred Hospital, Melbourne, Australia

⁶Royal North Shore Hospital of Sydney, Sydney, Australia

⁷ICNARC, Tavistock House, Tavistock Square, London, UK

⁸Department of Intensive Care Medicine, The St George Hospital, Sydney, New South Wales, Australia

⁹Department of Intensive Care Medicine, Queen Elizabeth Hospital, Adelaide, SA, Australia

¹⁰Austin Hospital, University of Melbourne, Melbourne, Victoria, Australia

¹¹Royal Perth Hospital, University of Western Australia, Perth, Western Australia

¹²Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Corresponding author: Anthony Delaney, adelaney@med.usyd.edu.au

Published: 14 April 2008

This article is online at <http://ccforum.com/content/12/2/210>

© 2008 BioMed Central Ltd

Critical Care 2008, **12**:210 (doi:10.1186/cc6849)

Abstract

Complex interventions, such as the introduction of medical emergency teams or an early goal-directed therapy protocol, are developed from a number of components that may act both independently and inter-dependently. There is an emerging body of literature advocating the use of integrated complex interventions to optimise the treatment of critically ill patients. As with any other treatment, complex interventions should undergo careful evaluation prior to widespread introduction into clinical practice. During the development of an international collaboration of researchers investigating protocol-based approaches to the resuscitation of patients with severe sepsis, we examined the specific issues related to the evaluation of complex interventions. This review outlines some of these issues. The issues specific to trials of complex interventions that require particular attention include determining an appropriate study population and defining current treatments and outcomes in that population, defining the study intervention and the treatment to be used in the control group, and deploying the intervention in a standardised manner. The context in which the research takes place, including existing staffing levels and existing protocols and procedures, is crucial. We also discuss specific details of trial execution, in particular randomization, blinded outcome adjudication and analysis of the results, which are key to avoiding bias in the design and interpretation of such trials.

These aspects of study design impact upon the evaluation of complex interventions in critical care. Clinicians should also consider these specific issues when implementing new complex interventions into their practice.

Introduction

Management of critically ill patients is complex, involving multiple interventions and processes. Concomitant life-threatening pathologies require numerous and potentially interactive therapies delivered by a variety of health-care professionals. One simple observation exists: outcomes are improved when care is coordinated by medical teams with experience, training, or decision support [1-3]. Consequently, there is an emerging body of literature advocating the use of integrated complex interventions to optimise the treatment of critically ill patients. Examples of these complex interventions include medical emergency teams [4], early goal directed therapy for the management of patients with severe sepsis [5], educational interventions to improve compliance with guidelines for the treatment of patients with pneumonia in the emergency department [6] or even a bundle of measures to

ARISE = Australasian Resuscitation in Sepsis Evaluation; ICU = intensive care unit; ProCESS = Protocolised Care for Early Septic Shock; ProMISe = Protocolised Management in Sepsis; RCT = randomized controlled trial.

improve the management of all patients in the intensive care unit (ICU) [7].

Complex interventions are defined as interventions or therapies that may act both independently and inter-dependently [8], often more than just the sum of their components. Complex interventions may be seen as interventions where the function of the intervention remains constant (for example, to alleviate hypoperfusion in patients with severe sepsis), rather than the specific components (for example, a specific resuscitation protocol, or use of a specific fluid regime) used to achieve this function [9]. This allows tailoring of the intervention to the context in which the intervention is applied.

The evaluation of complex interventions requires a careful study of all potential benefits as well as adverse effects that could be attributed to the intervention. While observational studies may provide insight into the effectiveness of treatments, causal inferences require appropriately powered, randomized controlled trials (RCTs) [10]. RCTs, however, are typically used to test single interventions, such as the benefits of a drug compared to placebo and although the principles underlying the testing of a more complex intervention are the same, particular theoretical and practical difficulties arise for researchers conducting trials and for clinicians attempting to critically appraise their results. These difficulties include determining a representative study population, defining the intervention and deploying it in a standardised manner, and measuring appropriate outcomes.

The aim of this review is to give clinicians insight into the process of designing trials to evaluate whether a new complex intervention results in improved outcomes. It is hoped that these insights will aid clinicians when they consider implementing complex interventions into their own practice.

Methods

The concepts and themes of this paper arose out of discussions held between the ProCESS (Protocolized Care for Early Septic Shock, US based trial), ARISE (Australasian Resuscitation in Sepsis Evaluation) and ProMiSe (Protocolised Management in Sepsis, UK based trial) investigators during the development of their respective protocols. Each team of investigators is planning a multi-centre RCT of a resuscitation protocol for patients with early severe sepsis. We augmented the expert discussions with a literature search (Medline was searched using the PubMed interface), using search terms for 'complex interventions' combined with search terms to identify studies relevant to critical care. We also searched the relevant epidemiological literature, and included references pertaining to recent illustrative cases in critical care. The themes and concepts are addressed in four sections: pre-trial activities, trial design, trial execution, and trial reporting, and are summarised in Table 1.

Pretrial activities

The phased development and testing of a new single intervention (such as a monoclonal antibody for sepsis) is a well-developed and well recognized process. *In vitro* testing followed by animal studies establish a biological rationale and provide preliminary safety data, and phase I trials in healthy volunteers and phase II trials in subjects with the target condition precede definitive phase III studies. When evaluating a complex intervention, the pre-trial phase may follow a different but analogous path. Frameworks for the design and evaluation of complex interventions that outline a step-wise approach to the research process exist [8,11]. It is important to note that when evaluating a complex intervention, this process may be iterative rather than linear. It may be necessary to explore the mechanism of action of the various components of the protocol simultaneously with an investigation of interactions between components and an exploration of the best methods to implement the protocol. In comparison to the evaluation of single interventions, where the focus is generally on addressing a question such as whether this new monoclonal antibody for sepsis reduces mortality, the evaluation of complex interventions may address the question of whether this protocol for resuscitation of patients with severe sepsis reduces mortality. Alternatively, a complex intervention trial may ask an implementation question: what is the optimal way to implement this protocol for resuscitation of patients with severe sepsis?

Literature review

As with any investigation, assessing previous knowledge, successes and difficulties must occur. This allows the research question to be honed and avoids unintended duplicative efforts. In the case of complex interventions, additional aspects of the literature review are crucial. In the case of a resuscitation protocol for patients with severe sepsis, the review must examine the supportive evidence for each of the component parts. There must be a rationale for combining the components and for the choice of methods used to educate the staff and to implement the protocol. Knowing what, if anything, is understood about the way that these components interact will allow an optimal design. An understanding of the methods of organisational change may also be important.

Retrospective studies

National databases such as the Intensive Care National Audit and Research Case Mix Programme in the United Kingdom [12,13] and the Australian and New Zealand Intensive Care Society Adult Patient Database [14] provide important information regarding the intended study population, potential recruitment rates and baseline patient outcomes [15]. For example, in an early sepsis resuscitation protocol trial, knowing how many of the components of a particular resuscitation protocol for sepsis are currently being delivered, whether the delivery of the components of therapy within the protocol have changed over time, and how the mortality has

Table 1**Comparison of the methodological issues to be considered in the evaluation of single and complex interventions in critical care**

Component of the evaluation	Evaluation of a single intervention (for example, a monoclonal antibody for patients with sepsis)	Evaluation of a complex intervention (for example, a resuscitation protocol for patients with sepsis)
Pre-trial activities		
Study question	To determine whether this monoclonal antibody compared to placebo reduces mortality for patients with severe sepsis	To determine whether this resuscitation protocol compared to usual care reduces mortality for patients with severe sepsis To determine the best way to implement a new protocol for the resuscitation of patients with severe sepsis
Pre-clinical phase	Linear approach from <i>in vitro</i> studies to animal studies to phase I and phase II clinical trials	Non-linear, iterative approach is needed to examine the effectiveness of each aspect of the protocol, how these aspects interact with current practice and what methods of implementing the protocol as a whole are likely to be most successful
Pilot studies	Focussed on feasibility of recruitment, compliance with treatment and follow-up	Will help determine feasibility of implementing the protocol as a whole, which components are most commonly implemented or missed Needed to identify barriers to implementing the protocol, potential means to overcome these barriers, optimal strategies for implementing the protocol
Trial design		
Population	Will be patients with the target condition, for example, two SIRS criteria and evidence of organ dysfunction in patients with suspected or proven infection	May be patients with the target condition, or it may be health service delivery organisations. For example, attempts to determine whether the protocol works may be focussed on patients with severe sepsis or attempts to determine how best to implement the protocol may be focussed on physicians or even hospitals
Intervention	Clearly defined single drug therapy	Will contain multiple interventions, for example, increased fluids, blood transfusions, vasopressors, additional monitoring devices (arterial lines, lactate measurements, ScvO ₂ measurements), as well as specific guidance to clinicians regarding the timing of these interventions
Comparison group	Placebo	The control group could receive 'usual care' as determined by individual clinicians, a defined protocol of 'usual care', a protocol with different components, or an alternative suite of interventions (for example, computerised reminders) to enhance compliance with the protocol under investigation
Outcome	Primary outcome: all cause mortality at 90 days	Primary outcome may be mortality or compliance with the protocol may be the primary outcome of interest. As blinding may be less than optimal, well defined and robust outcomes are required
Context	May relate to the other treatments delivered in conjunction with the monoclonal antibody treatment. Generally reported in a table of co-interventions	Crucial element of trial design. Factors to consider include the existing protocols in place, staffing levels (both numbers and experience), availability of ScvO ₂ monitors, resources of the emergency department and current treatment patterns
Trial execution		
Randomization	Individual participants will be randomized	Randomization may be at the individual participant level, particularly for trials designed to determine whether the protocol is effective Randomization may also need to be at the level of the health care provider or service delivery organization when the aim of the study is to determine how best to implement the protocol
Blinding	Blinding should be possible	Blinding of the intervention is likely to be difficult or impossible, and may not be desirable if the intention is to determine the best way to implement the protocol. Attempts to blind outcome adjudication, data analysts may be possible and will enhance internal validity
Analysis	Simple statistical analysis is usually possible	Complex analysis is required for multi-arm trials and cluster-randomized trials Compliance with the protocol is likely to be of greater interest, and a per-protocol analysis may offer information regarding aspects of the protocol that did or did not add value
Trial reporting		
Reporting	Should follow CONSORT statement	Should follow CONSORT statement or the extension relating to cluster-randomized trials when appropriate

ScvO₂ = central venous oxygen saturation; SIRS, Systemic Inflammatory Response Syndrome; CONSORT, Consolidated Standards of Reporting Trials.

Table 2**Crude mortality by calendar year (1997 to 2005) for patients admitted to ICU following presentation to the emergency department in Australia and New Zealand with sepsis or septic shock**

Calendar year	Percent ICU mortality (n) ^a	Percent hospital mortality (n) ^b
1997	27.4 (96/351)	35.6 (124/348)
1998	29.1 (87/298)	37.7 (113/301)
1999	22.6 (60/266)	30.7 (80/261)
2000	27.8 (148/534)	35.2 (184/522)
2001	23.3 (196/840)	31.6 (256/809)
2002	21.7 (219/1,011)	28.1 (280/994)
2003	19.7 (241/1,223)	25.8 (311/1,209)
2004	18.5 (260/1,403)	24.9 (350/1,403)
2005	15.6 (207/1,324)	21.2 (281/1,325)

^aFor ICU mortality, total number of patients = 7,250 (data not available for 399 patients (5.2 percent)). ^bFor hospital mortality, total number of patients = 7,172 (data not available for 477 patients (6.2 percent). ICU, intensive care unit. (Reproduced with permission from [15].)

changed over time (Table 2) all aid in designing a trial. These studies also provide clinicians with the context within which the evaluation of the protocol is taking place.

Prospective observational studies and surveys

A prospective observational study may add to the previous background data, defining current care and the associated baseline outcome rate. Prospective observational studies may assist in defining 'standard care', determining the appropriate treatment for the control arm in a future process-of-care trial and the expected outcomes in the control patients. These data, combined with a realistic potential treatment effect, allow investigators to determine an appropriate sample size.

Surveys of clinician opinions may identify potential barriers to the implementation of the trial intervention. For example, if a trial required patients with severe sepsis be treated in the emergency department for six hours but a survey of emergency physicians revealed that most would not delay transfer to the ICU, such a trial might not be possible. Both prospective observational studies and surveys may also provide information regarding differences in attitudes and practice between countries.

For cluster-randomized trials (studies that randomize at the level of the centre rather than the patient), reliable estimates of practice variability and outcomes differences between centres are needed. These data can guide the design and analytic plan by noting variances that must be accounted for during the analysis.

Pilot studies

Pilot studies, by which we refer to small prospective RCTs designed to test aspects of the intervention, are vital for determining the feasibility, reproducibility or implementation problems associated with the complex intervention evaluated. Given that existing processes of care will differ between centres, pilot studies in multiple centres are preferable, particularly to ensure that a multi-centre trial is feasible in both academic (for example, tertiary referral) and non-academic (for example, community and rural) hospitals. For example, barriers to the training and implementation of a medical emergency team in a large teaching hospital will be different to those in smaller community hospitals and pilot studies may help to identify these issues. This will have particular importance if a trial is conducted in multiple countries, when the variation in practice is likely to be greater.

Trial design

All clinical research, including studies of complex interventions or processes of care, should be designed to answer a clearly articulated question [16]. This should be a terse declarative statement or focused question - the former, using a hypothesis format (research or null) is common and best expressed clearly. Clinicians will need to consider both the internal validity (the extent to which systematic error has been avoided) and the external validity (the extent to which the results of the trial provides a correct basis for generalization to other circumstances) [17] to make a judgment about whether a new complex intervention has a place in their clinical context.

The population

In most studies of simple interventions, the population under investigation is a well-defined group of patients with a similar condition, such as patients with acute myocardial infarction. Some complex interventions will involve populations with a defined condition, such as patients with early severe sepsis [5] or high-risk surgical patients [18]. In other circumstances, the population under investigation may not be as clear-cut. The aim of other studies, such as the introduction of trauma teams [19] or medical emergency teams [4], is to determine whether changing the overall healthcare system can deliver improved care. In these cases, the population may be the healthcare providers or an entire healthcare system. Studies designed to evaluate whether the process improves outcomes for individual patients may still randomize individual patients. However, in studies designed to evaluate how best to ensure healthcare providers implement a complex protocol, the unit of randomization for implementing the new protocol may need to be the healthcare providers, or even whole hospitals, to avoid contamination of the control group. It is important to realise that the unit of randomisation, observation and outcome measurement need not necessarily all be the same. For example it may be possible to randomize at the level of the health service provider and measure outcomes in individual patients.

The intervention

Care must be exercised to specify exactly what is being studied. For example, a protocol for the resuscitation of patients with severe sepsis in the emergency department [5] includes a number of individual components that could affect outcome (that is, fluid resuscitation, vasopressors, blood transfusion, ventilation). Alternatively, it may be the use of a novel monitoring device or the attention of the additional support staff necessary to co-ordinate the protocol that makes a difference. Clearly, any or all of these factors may be the essential component(s) that contribute to improved patient outcomes. In early goal-directed sepsis care, the intervention studied is the total protocol and the team that identify and alleviate early hypoperfusion. The total intervention may consist of a number of tests (such as the measurement of serum lactate), measurements (such as the measurement of central venous oxygen saturation), and interventions (for example, fluids, dobutamine and blood transfusions) that are prescribed in response to these measurements. It may not be possible to determine which particular part of the intervention is the primary reason for any observed change in mortality. This lack of clarity may be frustrating for clinicians who disagree with elements of the protocol. However, it need not impede rigorous initial evaluation of the impact of the overall protocol [9]. The analogy to studies of single interventions is clear; for example, acceptance of streptomycin as a treatment for tuberculosis occurred [20] long before a clearly defined molecular mechanism existed [21]. Similarly, a protocol that reduces mortality could be implemented with subsequent research undertaken to identify the component(s) most directly responsible for the benefit.

Alternatively, the intervention could be a suite of educational materials designed to improve compliance with a bundle of measures - for example, efforts to reduce central venous catheter related blood stream infections. In this case, the intervention may be an intensive campaign with computerised order sets, regular audits, standardised checklists and one-to-one educational sessions to ensure compliance compared to a less intensive campaign with only a routine checklist. The optimal approach is determined by the research question that is being addressed.

Context dependence. New processes of care may have different impacts on outcome depending on the background processes already in place. The current care context into which the new process of care is introduced needs to be considered to ensure that the proposed intervention will have the appropriate and desired outcome. For example, a resuscitation protocol for patients with early severe sepsis tested in a single centre may not produce results that are generalisable because of aggressive ancillary care that might be uncommon elsewhere. Such a protocol may need to be tested in a multi-centre study that includes various hospital types and locations to support the generalisability or external validity of the results.

Reproducibility. One of the major challenges in evaluating new complex interventions is ensuring that the intervention is accurately and reliably delivered. For large-scale trials, this is a continuous process. Delivery of the intervention could potentially improve over time from a 'learning curve effect'. Conversely, delivery of the intervention may degrade with time if the recruitment period is prolonged and trial fatigue develops. Pilot studies can help identify these learning curves, with data from the pilot intentionally excluded from the final analysis and serving only to prepare for the study. It is likely that a variety of methods will be required to ensure that the intervention is delivered reliably, including audits, computerized reminders, checklists, intensive one-to-one educational sessions and incentives for sites with the highest compliance, which may all ensure smooth implementation of the set protocols. These must be described so that clinicians can use (or avoid) strategies when implementing the new processes in their own practice.

The comparison group

One area of contention when evaluating new complex interventions or processes of care is how to define the comparison (control) group. This is a complicated problem and the optimal approach is unclear [22]. Is the control 'care as it happens now' (termed 'wild type') or is it 'a regimented and commonly accepted care'? The Acute Respiratory Distress Network (ARDSNet) low tidal volume trial [23] generated controversy in defining the comparator [24,25]. In that trial, it was argued that the control group received a standardized treatment that substantially differed from usual clinical practice; the trial results could be interpreted as demonstrating that a low tidal volume strategy is better than high tidal volume ventilation in patients with acute lung injury. This differs from concluding that introducing a protocol for low tidal volume ventilation reduces mortality when compared with current practice.

There are a number of ways to address this issue. First, if the requisite systematic review and observational studies are complete, researchers will be aware of what is 'usual' or standard care. Second, dual comparison groups may be included; in the example above, a low tidal volume group, a structured higher (albeit commonly used) tidal volume group, and a 'wild-type' group where the tidal volume is determined by the treating clinician would resolve this concern. This allows those performing the study and clinicians reading the results to determine whether the new intervention (for example, the low tidal volume protocol) is superior not only to the high tidal volume protocol, but also current practice.

The problem with this dual control approach is that the use of three groups complicates the analysis and increases the required sample size. Sample size may be kept lower if a sequential hypothesis testing procedure is used. For example, one first tests whether a ventilation protocol (either high volume or low volume) is better than allowing clinicians to

titrate according to clinical judgement. If the null is rejected, one can then ask if one protocol is superior to the other. Without meticulous attention to study rationale and ensuring an adequately powered study, the chance is high that the first hypothesis test will suggest no difference between therapeutic arms, thus precluding further primary hypothesis testing.

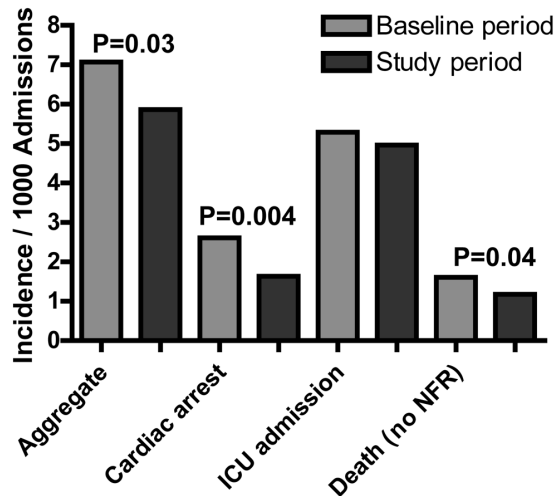
An alternative approach may be to have only two groups, that is, the new process compared to 'usual care'. This retains the advantage that the control group receives care titrated by clinicians in response to changes in patients' conditions, rather than applying a protocolized approach (a criticism of complex therapy trials in the critically ill [26]). This will simplify the analysis and place less pressure on the required sample size. If 'usual care' differs greatly from country to country or centre to centre, concerns may be raised that the control arm is uninterpretable and not representative of 'usual care' in other settings. Also, education and training for the new process may contaminate the 'usual care' control group. As the trial progresses and more medical and nursing staff are exposed to the education required to implement the new process, it is possible that standard care will drift towards that being implemented in the new process or protocol. Changes in the outcomes in the control group may be seen due to these processes even in the absence of the new intervention, a phenomenon that has been previously observed (Figure 1) [4]. Thus, differences in care between the groups will be less than otherwise anticipated, which may also reduce the apparent treatment effect and the power of the study, and may require reconsideration of the sample size. It may also cause some concern among clinicians who perceive their practice to be considerably different to that delivered in the 'usual care' arm of the trial and, therefore, question the applicability of the findings to their patients.

The outcome

Choosing a primary outcome for a complex intervention RCT is no different to choosing an outcome in any clinical research. The outcome should be robust and well defined to avoid ascertainment bias. For most trials in critical care, a clinically important outcome such as all-cause mortality at a defined point in time, such as 30 or 90 days, is appropriate.

One unique aspect of trials evaluating complex interventions is the need to collect data concerning the actual process of care to demonstrate adherence to the protocol. It may not be possible to measure compliance with every aspect of the protocol, so decisions will need to be made regarding the key components of any given protocol. When assessing compliance with a resuscitation protocol, it may be important to measure not only that each component of the protocol is delivered, but also that they are delivered in a timely manner. For example, in a resuscitation protocol that is guided by the use of continuous measurement of central venous oxygen saturation, it may be necessary to ensure that this measurement is collected in every patient in the treatment

Figure 1



MERIT study changes in outcomes over time: control hospitals. Drawn from data in [4]. ICU, intensive care unit.

arm, and that it is not used in the control arm of the trial. It may also be necessary to measure how long it took to achieve the goals set by the protocol, and how these goals were achieved. Secondary analyses, regarding which aspects of the protocol were most often delivered or neglected, may offer insights into success or failure of the intervention that may help guide future research projects. Data collection - especially that focused on adherence to protocol delivery - must continue throughout the delivery of the entire process, which has resource implications for researchers and funding agencies.

Trial execution

Randomization

The randomization process, particularly allocation concealment (preventing anyone knowing which group an individual or cluster of patients will be randomized to, until randomization occurs), is essential to limit bias [27]. Trials of complex interventions may differ from trials of simple therapies when the unit of randomization is whole practice units (for example, hospitals, ICUs) or individual providers (for example, for an educational intervention). By randomizing at the hospital level (cluster randomization), it is less likely that the educational efforts required to successfully implement the new process will affect the control group. This approach has previously been used in trials of complex interventions in critical care [4,6]. However, heterogeneity between the various healthcare services may be problematic, with very large numbers of hospitals (for example, more than 100 units) required to obtain adequate power. Identification and randomization of large numbers of suitable ICUs or hospitals may not be logistically feasible or economically viable. Moreover, cluster randomization does not ensure blinding of the clinical trial.

Blinding

Blinding, defined as attempts to keep trial participants, investigators or assessors unaware of the assigned treatment, is important to limit bias in clinical trials [28]. It may not be possible to conduct a RCT evaluating a complex intervention or process-of-care in a blinded fashion. As blinding is largely designed to avoid bias in the ascertainment of outcomes, using clear, robust and well-defined outcomes can help to limit this concern. If less objective outcomes are needed in an unblinded trial (such as assessing neurological recovery using the Glasgow outcome score or diagnosing the presence of ventilator associated pneumonia), using an outcome adjudication committee unaware of the intervention can help avoid bias. Data collectors and data analysts unaware of the intervention used in each patient can also limit bias.

Analysis

The primary analysis of most trials evaluating complex interventions does not differ from that of trials involving single interventions. In particular, while it may be tempting to exclude patients or centres where the intervention was not fully implemented, an intention-to-treat analysis (analysing all participants in the group to which they were randomized, regardless of whether they completed the protocol or not [10]) is the best primary focus of the results. Differences between the results of an intention-to-treat analysis and a per-protocol analysis or additional sensitivity analyses may point to implementation difficulties and may highlight areas for future research. If a cluster-randomized design or multiple groups have been involved, the analysis will necessarily be more complex and require advanced analytic techniques. For all trials regardless of design, a pre-analysis statistical plan will ease concerns of *post hoc* data manipulation and analytical bias.

Trials of complex interventions lend themselves to subgroup analyses. By examining each component of a new protocol, researchers may be able to demonstrate associations between these components and various outcome measures. Subgroup analyses should be prospectively determined, included in a statistical analysis plan, and limited in number (to avoid an appearance of pre-planned data mining). While subgroup analyses may help form hypotheses for future research, they should not be relied upon to provide robust evidence to guide clinical practice [29,30] and should be done with extreme caution and with stated clarity if initiated after data collection (again for concerns of data mining).

Trial reporting

There are widely accepted standards for the reporting of parallel group RCTs [31] and cluster randomized clinical trials [32]. These guidelines are equally applicable when reporting trials of complex interventions. Attention should focus on the description of the intervention, allowing it to be reproduced if so desired. A careful description of the treatments delivered to the control group(s) is needed. This will allow clinicians to determine the implementation context

and may demonstrate any deviations from usual care that may have occurred in the control group as a result of the implementation of the new intervention.

Assessment of the strength of the evidence for complex interventions

With increasing focus on evidence-based medicine, considerable attention is paid to the internal validity of studies. However, while an internally valid study is essential, it may not be sufficient to warrant a change in clinical practice. Many factors contribute to understanding the utility of a new complex intervention. For example, the guidelines proposed by Sir Austin Bradford Hill in 1965 relating to the determination of causal associations are important [33]. These include the strength of the association, consistency of the association in different contexts, the temporal relationship between the intervention and the outcome, the dose-response relationship and the underlying biological plausibility of the intervention. Multiple studies of a complex intervention may be needed to fully address these issues.

Clinicians may wonder about the inclusion criteria used in a clinical trial and if this differs from patients in their setting. Multi-centre studies or methodologically sound observational studies offer some protection from this concern. Progress in the treatment of the disease and changes in technology also need to be considered. These factors may impact on the existing processes to change mortality over time (as illustrated in Table 2) and make the interpretation of the data and predicting the impact of the new interventions more difficult. Ongoing surveillance may identify other unexpected untoward consequences arising from new interventions, such as changes in antibiotic prescribing leading to changes in bacterial resistance patterns or outcomes.

Given these difficulties, some attempts to develop simple grading systems to summarise the strength of evidence in the medical literature exist [34], none of which is ideal. While these grading systems are constantly being refined [35], most clinicians will find the use of a structured critical appraisal helps them assess the strength of evidence provided by individual trial reports.

Ongoing research

The challenge of improving the process of care for critically ill patients will not be overcome by a single research project. For example, the mortality rates for acute myocardial infarction have fallen over the past 20 years, not due to a single intervention, but due to multiple interventions that have been combined into a process of care that improves overall outcome [36]. It is not just the percutaneous coronary intervention that improves outcome, but also the combination of aspirin, beta-blockers, early recognition by the emergency medical staff, having a reperfusion team available for early angiography, ensuring that the appropriate discharge medications are given and that the patient attends an

appropriate rehabilitation program that combine together to improve outcomes. These interventions are accepted because of large suitably designed RCTs, which have consistently shown their effectiveness. The same is likely to be true in other acute care areas. For example, improvement in the outcome for patients with severe sepsis may require early identification, effective resuscitation, early and appropriate antimicrobial therapy and adequate source control, all delivered by the appropriate people at the appropriate time.

The results of RCTs of complex critical care interventions must be examined closely so that an ongoing program of research aimed at improving the care of the critically ill can be sustained. Each completed project is likely to suggest problems that need further investigation. Examples might include better ways to ensure compliance with a protocol, refined protocols with more of one component and less of others, extending the scope of the new process to other populations or moving the process into a new context, such as out of academic centres and into community hospitals. By doing so, all critically ill patients can share in improvements in the process of their care.

Conclusion

There are specific issues involved in the evaluation of complex interventions that clinicians should consider. By considering these details, both researchers and clinicians will be able to work together to improve the process by which we care for critically ill patients.

Competing interests

The authors of this manuscript are all active investigators in the ARISE (Anthony Delaney, Rinaldo Bellomo, Peter Cameron, D James Cooper, Simon Finfer, John A Myburgh, Sandra L Peake, Steve AR Webb), ProCESS (Derek C Angus, David T Huang, Michael C Reade, Donald M Yealy) and ProMiSe (David A Harrison) studies. These are international, multi-centre, randomised controlled trials of a complex intervention, early goal-directed therapy, for patients with severe sepsis.

Acknowledgements

This work was funded in part by NIH grant NIGMS 1P50 GM076659 and by NHMRC grant 491075.

References

1. Kahn JM, Goss CH, Heagerty PJ, Kramer AA, O'Brien CR, Rubenfeld GD: **Hospital volume and the outcomes of mechanical ventilation.** *N Engl J Med* 2006, **355**:41-50.
2. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, Sexton B, Hyzy R, Welsh R, Roth G, Bander J, Kepros J, Goeschel C: **An intervention to decrease catheter-related bloodstream infections in the ICU.** *N Engl J Med* 2006, **355**:2725-2732.
3. Pronovost PJ, Angus DC, Dorman T, Robinson KA, Dremsizov TT, Young TL: **Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review.** *JAMA* 2002, **288**:2151-2162.
4. Hillman K, Chen J, Cretikos M, Bellomo R, Brown D, Doig G, Finfer S, Flabouris A: **Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial.**

Lancet 2005, **365**:2091-2097.

5. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, Peterson E, Tomlanovich M: **Early goal-directed therapy in the treatment of severe sepsis and septic shock.** *N Engl J Med* 2001, **345**:1368-1377.
6. Yealy DM, Auble TE, Stone RA, Lave JR, Meehan TP, Graff LG, Fine JM, Obrosky DS, Mor MK, Whittle J, Fine MJ: **Effect of increasing the intensity of implementing pneumonia guidelines: a randomized, controlled trial.** *Ann Intern Med* 2005, **143**:881-894.
7. Vincent JL: **Give your patient a fast hug (at least) once a day.** *Crit Care Med* 2005, **33**:1225-1229.
8. **A framework for development and evaluation of RCTs for complex interventions to improve health** [<http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC003372>]
9. Hawe P, Shiell A, Riley T: **Complex interventions: how "out of control" can a randomised controlled trial be?** *BMJ* 2004, **328**:1561-1563.
10. Collins R, MacMahon S: **Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials.** *Lancet* 2001, **357**:373-380.
11. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P, Kinmonth AL: **Designing and evaluating complex interventions to improve health care.** *BMJ* 2007, **334**:455-459.
12. Harrison DA, Brady AR, Rowan K: **Case mix, outcome and length of stay for admissions to adult, general critical care units in England, Wales and Northern Ireland: the Intensive Care National Audit & Research Centre Case Mix Programme Database.** *Crit Care* 2004, **8**:R99-111.
13. Harrison DA, Welch CA, Eddleston JM: **The epidemiology of severe sepsis in England, Wales and Northern Ireland, 1996 to 2004: secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database.** *Crit Care* 2006, **10**:R42.
14. Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, Bellomo R: **Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database.** *J Crit Care* 2006, **21**:133-141.
15. ARISE; ANZICS APD Management Committee: **The outcome of patients with sepsis and septic shock presenting to emergency departments in Australia and New Zealand.** *Crit Care Resusc* 2007, **9**:8-18.
16. Richardson WS, Wilson MC, Nishikawa J, Hayward RS: **The well-built clinical question: a key to evidence-based decisions.** *ACP J Club* 1995, **123**:A12-13.
17. Juni P, Altman DG, Egger M: **Systematic reviews in health care: assessing the quality of controlled clinical trials.** *BMJ* 2001, **323**:42-46.
18. Sandham JD, Hull RD, Brant RF, Knox L, Pineo GF, Doig CJ, Laporta DP, Viner S, Passerini L, Devitt H, Kirby A, Jacka M; Canadian Critical Care Clinical Trials Group: **A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients.** *N Engl J Med* 2003, **348**:5-14.
19. Bernhard M, Becker TK, Nowe T, Mohorovicic M, Sikinger M, Brenner T, Richter GM, Radeleff B, Meeder PJ, Büchler MW, Böttiger BW, Martin E, Gries A: **Introduction of a treatment algorithm can improve the early management of emergency patients in the resuscitation room.** *Resuscitation* 2007, **73**:362-373.
20. Daniels M, Hill AB: **Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council trials.** *BMJ* 1952, **1**:1162-1168.
21. Chopra I, Brennan P: **Molecular action of anti-mycobacterial agents.** *Tuber Lung Dis* 1997, **78**:89-98.
22. **Considering usual medical care in clinical trial design: scientific and ethical issues.** In: *November 14 and 15 2005; Bethesda, Maryland*: NIH program on clinical research policy analysis and coordination; 2005 [http://crpac.od.nih.gov/Draft_UsualCareProc_06062006_cvr.pdf accessed 11 July 2007].
23. Network TARDS: **Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome.** *N Engl J Med* 2000, **342**:1301-1308.
24. Eichacker PQ, Gerstenberger EP, Banks SM, Cui X, Natanson C: **Meta-analysis of acute lung injury and acute respiratory dis-**

- ress syndrome trials testing low tidal volumes. *Am J Respir Crit Care Med* 2002, **166**:1510-1514.
25. Steinbrook R: **How best to ventilate? Trial design and patient safety in studies of the acute respiratory distress syndrome.** *N Engl J Med* 2003, **348**:1393-1401.
 26. Deans KJ, Minneci PC, Suffredini AF, Danner RL, Hoffman WD, Ciu X, Klein HG, Schechter AN, Banks SM, Eichacker PQ, Natanson C: **Randomization in clinical trials of titrated therapies: unintended consequences of using fixed treatment protocols.** *Crit Care Med* 2007, **35**:1509-1516.
 27. Schulz KF, Chalmers I, Hayes RJ, Altman DG: **Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.** *JAMA* 1995, **273**:408-412.
 28. Schulz KF, Grimes DA: **Blinding in randomised trials: hiding who got what.** *Lancet* 2002, **359**:696-700.
 29. Cook DJ, GebSKI VJ, Keech AC: **Subgroup analysis in clinical trials.** *Med J Aust* 2004, **180**:289-291.
 30. Assmann SF, Pocock SJ, Enos LE, Kasten LE: **Subgroup analysis and other (mis)uses of baseline data in clinical trials.** *The Lancet* 2000, **355**:1064-1069.
 31. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T: **The revised CONSORT statement for reporting randomized trials: explanation and elaboration.** *Ann Intern Med* 2001, **134**:663-694.
 32. Campbell MK, Elbourne DR, Altman DG: **CONSORT statement: extension to cluster randomised trials.** *BMJ* 2004, **328**:702-708.
 33. Hill AB: **The environment and disease: association or causation?** *Proc R Soc Med* 1965, **58**:295-300.
 34. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group: **Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group.** *BMC Health Serv Res* 2004, **4**:38.
 35. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group: **Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system.** *BMC Health Serv Res* 2005, **5**: 25.
 36. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, Giles WH, Capewell S: **Explaining the decrease in U.S. deaths from coronary disease, 1980-2000.** *N Engl J Med* 2007, **356**: 2388-2398.