


ORIGINAL



# Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a perspective from COMBACTE's STAT-Net

Jean-François Timsit<sup>1,2\*</sup> , Marlieke E. A. de Kraker<sup>3\*</sup>, Harriet Sommer<sup>4</sup>, Emmanuel Weiss<sup>5,6</sup>, Esther Bettiol<sup>3</sup>, Martin Wolkewitz<sup>4</sup>, Stavros Nikolakopoulos<sup>7</sup>, David Wilson<sup>8</sup> and Stephan Harbarth<sup>3</sup>, on behalf of the COMBACTE-NET consortium

© 2017 The Author(s). This article is an open access publication

## Abstract

**Purpose:** In this era of rising antimicrobial resistance, slowly refilling antibiotic development pipelines, and an aging population, we need to ensure that randomized clinical trials (RCTs) determine the added benefit of new antibiotic agents effectively and in a valid way, especially for severely ill patients. Unfortunately, universally accepted endpoints for the evaluation of new drugs in severe infections are lacking.

**Methods:** We review and discuss the current practices and challenges regarding endpoints in RCTs in this field and propose novel approaches.

**Results:** Usual endpoints actually recommended for drug development suffer from important flaws. Mortality requires large sample size and only partly related to the infectious process. Clinical cure rate is highly subjective in critically ill patients where symptoms may be related to other intercurrent events. Currently, composite endpoints, hierarchical nested designs, and competing risks analysis seem to be the most promising new tools for designing and analyzing clinical trials in this area, although they require further validation.

**Conclusion:** Regulatory authorities, pharmaceutical companies, and clinicians need to agree on the most appropriate clinical endpoints for severe infections to ensure efficient approval of new, effective antibiotic agents.

**Keywords:** Randomized clinical trials, Endpoints, Antibiotic therapy, Severe infections, Consensus

## Introduction

The use of well-defined outcomes to assess clinical trial results is of major importance. Ideally, endpoints in

clinical trials for novel antibiotic agents should be objective, reproducible, have a high internal and external validity, and be clinically meaningful: a direct measure of how patients feel, function, and survive [1, 2]. Unfortunately, at the moment, there is a lack of universal, well-accepted endpoints, particularly for severe hospital-acquired infections. This has resulted in inconsistencies in how trials are designed and reported, raising questions about internal validity and making interpretations across trials difficult. A recent Delphi process (see definition in Table 1) to define standardized endpoints for trials on antibiotic therapy for bloodstream infections (BSIs) identified that no well-validated primary endpoints existed; mortality or

\*Correspondence: jean-francois.timsit@bch.aphp.fr; marlieke.dekraker@hcuge.ch

<sup>2</sup> APHP Medical and Infectious Diseases ICU, Bichat Hospital, 46 Rue Henri Huchard, 75018 Paris, France

<sup>3</sup> Infection Control Program, Geneva University Hospitals and Faculty of Medicine, Gabrielle-Perret-Gentil 4, 1205 Geneva, Switzerland  
Full author information is available at the end of the article

**Take-home message:** There is an urgent need for consensus on valid endpoints for clinical trials evaluating antibiotic therapies for severe infections. Composite endpoints, hierarchical nested designs, and competing risks analysis seem to be the most promising new tools.

**Table 1 Glossary of important concepts**

Concept	Explanation
Delphi process	A structured survey method with multiple rounds, which relies on a panel of experts. Questions are asked and responses summarized in multiple rounds with the goal of convergence to a consensus
A superiority trial	A RCT designed to test whether a new treatment is better than an old treatment with respect to a pre-specified primary endpoint
A non-inferiority trial	A RCT designed to test whether a new treatment is at least as good as the active control, which often consists of the best available treatment at that moment in time. The main goal is to find therapies with advantages in other aspects, like the safety profile, administration method, or expense
Non-inferiority margin	A pre-specified, maximum treatment difference for the primary outcome measure that is still acceptable given the possible advantages of the new treatment
Attributable mortality	The mortality in the exposed study population minus the mortality in the unexposed study population; i.e. the mortality associated with the exposure, for example VAP
Composite endpoint	An endpoint combining multiple single endpoints into one measure, often including a clinical endpoint and a safety endpoint. This increases the power of the RCT, as compared to a RCT where both endpoints would be tested separately. For example, combining mortality and kidney failure, whereby the first occurrence of either is considered a negative outcome
Hierarchical endpoint	A special type of composite endpoint, whereby the hierarchy of the individual endpoints is considered; if the most important endpoint occurs, the other endpoints lower in hierarchy are no longer considered
Hierarchical nested design	A RCT design, where the primary endpoint needs to be compared in a non-inferiority design, and if non-inferiority is confirmed, predetermined additional endpoints can be tested for superiority
Competing events	An event that either hinders the observation of the event of interest or modifies the chance that this event occurs, i.e. hospital discharge in case hospital mortality is the primary endpoint
Multistate model	A statistical method to model an ongoing random process, thereby allowing patients to move from one state to a pre-determined number of other states, for example from hospitalized to infected to death, whereby all transitions can be quantified

clinical cure, at hospital discharge or up to 12 weeks after treatment, were the most common primary endpoints [3]. As another example, the US Food and Drug Administration (FDA) recommends all-cause mortality as the primary efficacy endpoint for trials on hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP) [4], whereas the European Medicines Agency recommends clinical outcome at the test of cure visit for these type of infectious syndromes.

At the same time, both cure and mortality endpoints have challenges associated with them, especially in critically ill patients. First, in this group of patients, cure is very difficult to define since clinical signs and symptoms may vary due to the infectious process studied, but also because of many concurrent adverse events during their stay in the intensive care unit (ICU) [5]. Second, all-cause mortality in critically ill patients is often related to the underlying illnesses and severity of disease [6]. And since trials often provide inadequate data on the “standard of care”, especially for the treatment of organ dysfunctions related to severe infections [7], it is difficult to associate death with treatment failure of the infection of interest. As a result of these endpoint issues, clinical trials often become less pragmatic [8] and exclude patients with a high risk of dying, or patients in whom underlying conditions may explain, at least partly, the risk of death, even though these new antibiotics, once available for clinical use, will also be administered to these patients. Finally,

non-infection-related deaths bias the potential effectiveness of the treatment under study towards non-inferiority, which is especially relevant given that randomized clinical trials (RCTs) for antimicrobial therapies are generally non-inferiority trials (Table 1).

A further challenge represents the required sample size to detect a clinically meaningful difference in the main outcome of interest. In trials examining new therapeutic options for severe infectious diseases, it is not possible to compare new treatment to placebo for ethical reasons. The added value of a new treatment needs to be compared to the standard of care, which often results in marginal differences. Furthermore, it may not be feasible or ethical to recruit patients for whom most benefit can be expected, hence reducing the marginal differences even further. Consequently, a non-inferiority trial needs to be considered. However, a standard non-inferiority margin (Table 1) of 10% is often considered too large from a clinical point of view, especially if survival rates or cure rates are high due to the exclusion of patients at high risk of dying, which often occurs as noted above. This makes the choice for and definition of the key endpoint in this patient group critical.

In this article, we review and discuss the current practices and challenges regarding endpoints in RCTs evaluating new therapies for severe infections and propose novel approaches which could improve internal and external validity of RCT outcomes. This is based on the

expertise that was fostered within the STAT-Net group through systematic reviews [9], re-analyses of clinical trial data [10], and testing of new analytical methods and study designs, which will be reported in more detail soon.

### Currently applied endpoints

Current primary endpoints in RCTs of severe infections include all-cause mortality, attributable mortality (Table 1), improvement of clinical parameters or specific biomarkers, microbiological eradication, antibiotic- or organ-failure-free days, and quality of life evaluations [11]. Advantages and disadvantages of these different endpoints will be discussed below and are summarized in Table 2.

### Mortality

Mortality endpoints constitute the most robust outcome criteria; it is the most severe outcome and can be measured objectively. RCTs in critical care have traditionally reported ICU-, hospital-, or 28-day-mortality, partly as a regulatory requirement, partly in an attempt to balance the time needed for a drug to show its effects and the time in which other disease processes could obscure the effect. The magnitude of the effect depends on the timeliness of initiation and appropriateness of empirical antibiotic therapy [12]. The attributable risk of death has been studied for VAP [6, 13, 14], HAP [15], catheter-associated urinary tract infections [16], and nosocomial BSIs [17] using various methodologies [18]. In HAP/VAP, the most important indication for antibiotic treatment in ICU, the attributable mortality (Table 1), reported by well-designed epidemiological studies, is around 3–10% when compared to treated controls [6, 13].

However, nowadays, trials become less pragmatic, losing the match between the trial setting and the setting to which its results will be applied [8]. The application of, for example, increasingly restrictive inclusion criteria have reduced reported all-cause mortality for hospital-acquired infections from 30 to 10–15% [18]. Restrictive criteria are required to ensure that the treatment can be adequately assessed; i.e. to prevent a sizeable proportion of patients dropping out within 48 h due to death. This does, however, mean that RCTs only include part of the real-life population for whom the drug could be of benefit, and could result in new drugs being prescribed off-label after approval [19]. Moreover, recently published epidemiological trends have shown that all-cause mortality rates could become even lower in future trials due to a combination of better recognition and improved standard of care [19–21]. This would make it even more difficult for a new treatment to demonstrate higher efficacy when focusing on mortality endpoints [22]. This emphasizes the importance of selecting an endpoint that is

sensitive enough to capture relevant added benefit for the trial patients, as well as patients expected to be treated with the new drug. In Fig. 1a, we show how a small difference in mortality between the intervention and control arm influences the required sample size for a clinical trial in a non-inferiority setting; with 22% mortality in the control arm versus 20% in the intervention arm, and a non-inferiority margin of 10%, more than 370 patients should be included to have a power of 80% to be able to conclude non-inferiority. Halving of the non-inferiority margin, to a more clinically acceptable 5%, would almost triple this requirement.

### Clinical cure

Clinical cure, i.e. investigator's assessment of clinical response, is the primary endpoint used in the vast majority of studies conducted before 2010 for severe infections such as HAP/VAP [23]. This endpoint can be more sensitive than mortality to assess treatment efficacy, especially in the context of low mortality rates. However, as a consensual definition of clinical cure is still lacking, the appreciation of cure by clinicians remains subjective, raising reliability and reproducibility issues. Indeed, clinical improvement is sometimes very hard to establish in severely-ill patients [2], and this lack of objectiveness can result in variability between centers in ascertaining cure, resulting in bias on the endpoint, consequently diluting or masking a potential treatment effect. Using an adjudication committee with pre-planned charter for adjudication may be a solution to circumvent such potential bias in assessment of clinical cure, but variability in diagnostic criteria can also impact clinical cure rates. A recent study showed for example that only 27.6% of the infection-related complications in mechanically ventilated patients are related to VAP [24]. Therefore, absence of clinical cure is related to VAP in only one-quarter of the cases. Finally, while mortality is usually assessed at day 28 or at 1 month, the variability of timepoints used to assess clinical cure may impact study results [9].

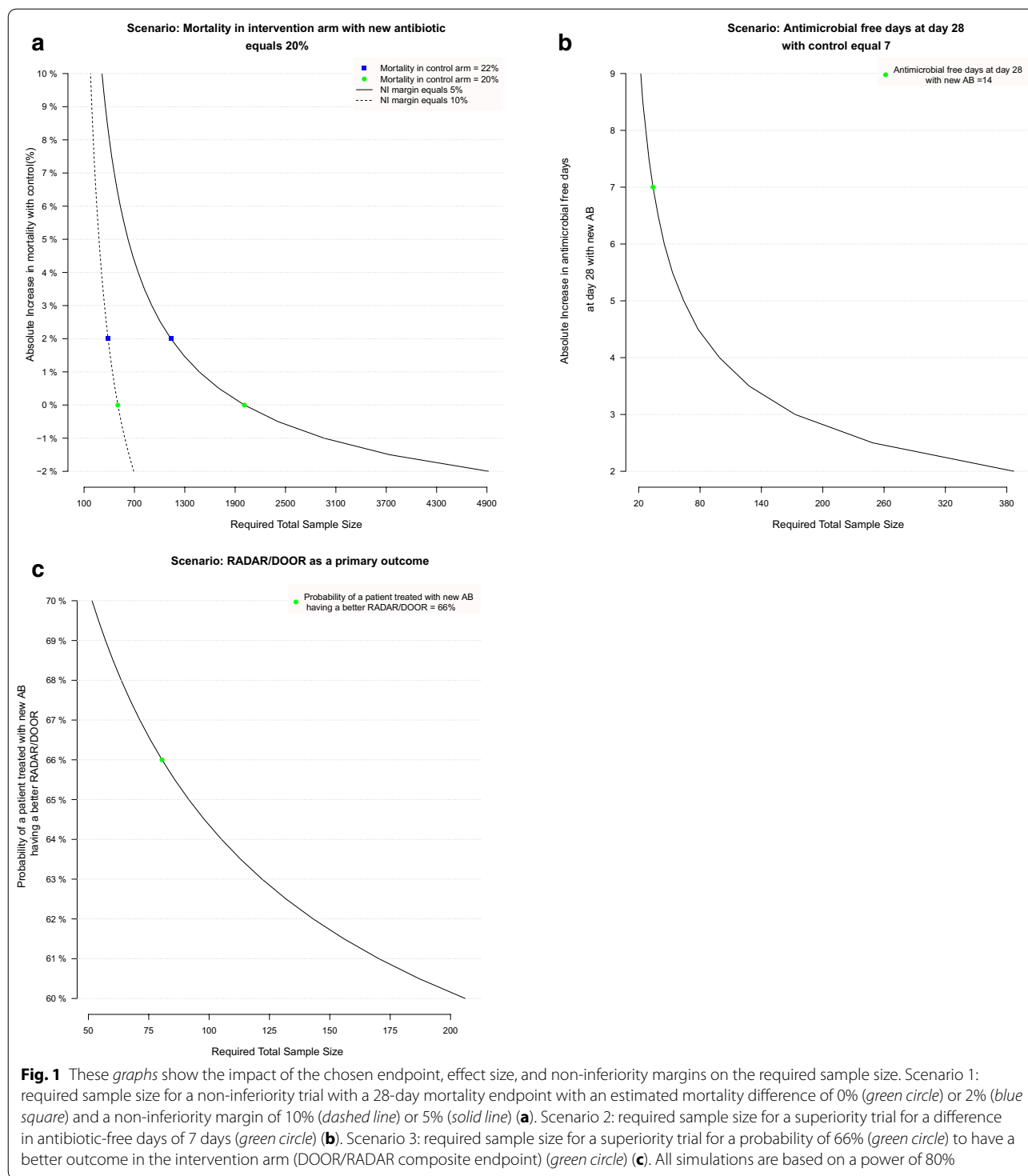
### Microbiological cure

Microbiological cure is a more objective endpoint, but it requires multiple, serial samples, which need to be adequately and reproducibly cultured. In particular, time to negativity of blood cultures and decrease in bacterial count of quantitative cultures from respiratory samples in VAP are frequently used in clinical practice. The main challenge is that the proportion of patients with a confirmed pathogen at baseline varies per infection type, but can be less than 50%, and repeated microbiological sampling is often not feasible or systematically performed. In most of the HAP/VAP studies conducted for

**Table 2 Advantages and disadvantages of clinical endpoints in randomized clinical trials evaluating antibiotic effectiveness in critically ill patients**

Endpoint	Advantages	Disadvantages
All-cause mortality [23]	Robustness: highly objective, accurate, and simple to measure Important to patients Non inferiority design allowed	Requires large sample sizes or large differences between groups Highly dependent on patient severity A substantial part is not related to infection in critically ill patients Ideal assessment time-point is debatable Recommended non-inferiority margins are high from a clinical perspective
Attributable mortality [18]	More relevant than all-cause mortality if many other causes of death or comorbidities are present	As above plus: Very difficult to assess Limited objectivity and reproducibility
Quality of life/functional status [43]	Patient-centered outcome	Lack of consensus Subjective Complex questionnaires No clinically relevant difference defined for power calculations and non-inferiority not feasible
Clinical cure (resolution of symptoms) [23]	Sensitive (especially if mortality rates are low)	No consensual definition: what symptoms should be included? Symptoms may be related to other diseases Possible subjectivity of assessment Different time-points for assessment may be necessary for each symptom
Microbiological cure Emergence of resistance [25, 26]	Objective Simple definition	Not relevant for all pathogens Requires isolation of a causative pathogen at baseline Requires multiple and sometimes invasive microbiological samples Requires homogenous and reproducible laboratory methods Does not correlate with clinical cure in some diseases (e.g., VAP)
Biomarkers [27]	Can be measured early in treatment before changes in treatment confound the effect Large differences could be detected resulting in smaller sample size	Requires a previous demonstration of surrogate properties No definition of a clinically meaningful difference for power calculations
Organ failure free survival/ventilation-free survival [28]	Combines mortality and other endpoints Improved power multiple endpoints are combined	May provide little extra meaningful data on top of mortality outcome
Antibiotic-free survival [44]	Improved power	Possible impact for the community difficult to ascertain and not directly related to individual impact May equally score patients dying at day 1 and patients alive and still under antimicrobials at end of study Difficult to define—concomitant antibiotics may be given to ensure coverage of all pathogens
Composite endpoint [30]	Improved power Can assess benefits and harms simultaneously	Difficult to interpret if there is collinearity between endpoints Clinically relevant effects could be diluted or even hidden by less important components
Hierarchical endpoint [35, 36]	Potential for providing an unified scale Potentially improves the power to detect superiority	Complexity of assigning a rank—previous databases need to explore how different it is to current endpoints Requires a previous consensual definition of the clinically meaningful difference

VAP ventilator-associated pneumonia



approval by the FDA before 2010, clinical cure without respiratory samples was also classified as microbiological cure. Moreover, in case of doubt of clinical success, additional sampling is done, i.e. the most critically ill patients are sampled more often, creating measurement

bias. Finally, the time to eradicate the organism is not always correlated with clinical response or increased mortality. For instance, Dennesen et al. and Shorr et al. failed to demonstrate a relationship between microbiological clearance and mortality for severe infections such

as HAP/VAP [25, 26]. This underlines how microbiological cure as an endpoint could readily lead to overestimation of clinical improvements and in that way disregards patients' wellbeing.

#### **Antibiotic- or organ-failure-free days**

Endpoints measuring antibiotic- or organ-failure-free days are among the most frequently used endpoints in non-registration RCTs, comparing antibiotic strategies in the ICU setting [27, 28]. This endpoint considers both mortality and other endpoints such as antimicrobial exposure or mechanical ventilation within one measure; free days are days without exposure between randomization and the pre-determined end of follow-up or death, whereby patients who die early have a higher chance of receiving a lower score. This strategy increases the chance of observing differences in outcome between treatment arms. If, for example, a new treatment increases the number of antibiotic-free days by 7 days, only around 30 patients would be required to have a power of 80% to detect this difference in a clinical trial (Fig. 1b). A disadvantage of this endpoint is, however, that patients with completely different infectious processes could receive the same score. For example, a patient under continuous antibiotic therapy and still alive at day 28 could receive the same score as another patient who died at day 7, because neither of them had any antibiotic-free days.

#### **Safety**

Safety endpoints are always measured in addition to other clinical endpoints. A new anti-infective therapy should be at least as safe as the comparator. However, in critically ill patients, the adverse events are frequent and could be due to many other drugs and invasive procedures. As such, it is almost impossible to causally attribute adverse events to the drug under study [5]. Given that treatment-limiting adverse events are part of the definition of failure on the clinical response endpoint, it is important to be able to causally assess the adverse event. Incorrect attribution of serious adverse events could jeopardize drug development and registration. Since the accuracy of the detection and attribution of serious adverse events is so low, i.e. high variability, and most RCTs are not specifically powered to detect these events, the chance of detecting any differences in safety endpoints between treatments is often very low.

#### **Emerging antimicrobial resistance**

The prevalence of multi-drug-resistant organisms (MDROs)-associated infections has increased over the years [29], and as such has increased complexities even further. If patients are randomized before microbiological results confirm the susceptibility profile of the causative

pathogen, it is impossible to distinguish resistant from susceptible infections prior to randomization. Consequently, the impact of empirical treatment of infections caused by MDROs will be diluted, and a non-inferiority trial (Table 1) would be the most feasible design, with the risk of including very few target patients. This trial therefore has a risk of providing little relevant value, especially if the overall patient group has a very high response rate, as this makes the often chosen non-inferiority margins (Table 1) around 10% clinically unacceptable. If the trial is focused on MDRO infections such that patients are randomized only after microbiological results are known, three major caveats need to be taken into account: a fully powered RCT is unlikely to be feasible as patients are rare; empirical therapy will dilute the impact of the new therapy; and possibly the optimal time frame for highest attainable impact has already passed. These challenges are present regardless of the type of endpoint studied. It does, however, make it critical that the most sensitive, relevant endpoint is chosen.

#### **Possible solutions (and their potential shortcomings)**

As outlined above, there is no consensus about the most appropriate endpoint for RCTs of antibiotic therapies in severely ill patients. Mortality endpoints are not very sensitive, while clinical cure endpoints are subjective, microbiological cure is not viable, and antibiotic- or ventilation-free days are ambiguous (Table 2). It will not be easy to reach consensus, but Delphi processes like the one recently published on endpoints for BSIs [3] are a step in the right direction. In this review, the authors concluded that it is unlikely for one endpoint to capture all relevant information and proposed to use composite endpoints. In this article, we will discuss three promising strategies: composite endpoints, which are, for example, used in several ongoing, open-label RCTs evaluating antibiotic therapies, combining mortality and clinical endpoints (NCT02634411, NCT02365493, NCT02575495). Secondly, empirical treatment trials aiming to show non-inferiority have become standard, but, given the importance of assessing patients infected with MDROs, a hierarchical nested design, combining a non-inferiority trial with a nested superiority trial (Table 1), should be considered. Finally, up-to-date statistical methods like competing event analyses (Table 1) and multistate models (Table 1) could provide more insight into how timing of events influences effect measures.

#### **Composite endpoints**

Combining multiple endpoints into one composite effect measure would overcome the need to have co-primary endpoints and thus avoids the issue of multiplicity and



the consequent need to adjust  $p$  values. It could reduce the required sample size even further, as the effect difference could be larger. The International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) has written the ICH E9 guideline [30], a guidance document about the principles of statistical methods applied to RCTs for marketing applications. In this document, composite endpoints are regarded as “a useful strategy”. In addition, defining a single endpoint that includes an efficacy component as well as a safety component is consistent with FDA’s philosophy of defining an endpoint based on how a patient feels, functions, and survives. Finally, separate analysis of the components of a composite endpoint as secondary outcome measures could preserve the possibility of comparing old and new trials.

However, construction of such an endpoint is challenging and interpretation can be misleading, especially when an intervention affects the distinct endpoints differently [31]. For example, length of stay may be reduced by a treatment only because mortality is increased. Furthermore, an inherent limitation of the conventional reporting of composite endpoints is that it emphasizes each patient’s first event, which is often the outcome of lesser importance. Therefore, Pocock et al. have suggested the win ratio as a new effect measure that takes the different priorities of the components into account [32]. Pairs are composed of patients from the new and comparator treatment, within randomization strata, and are subsequently grouped into winners/losers based on whether the treated pair member experienced the most/least favorable event first (Table 3).

Evans et al. recently proposed another way to utilize composite endpoints: the desirability of outcome ranking (DOOR) and the response adjusted for the duration of antibiotic risk (RADAR) [33]. For DOOR, a composite score is designed by assigning higher ranks to patients with better overall clinical outcomes; clinical success, clinical benefit with adverse event (AE), clinical failure without AE, clinical failure with AE, and death can be components of this composite endpoint. For RADAR, rankings are based on the duration of antibiotic use, which is tailored for trials evaluating optimal antibiotic use, and this can be combined with DOOR. In the end, DOOR/RADAR distributions can be compared between treatment arms. The generic version of DOOR is very similar to the win ratio (Table 3). This composite score could greatly reduce the required sample size to detect superiority for a novel drug. In a recent re-analysis of clinical trial data, looking at shorter duration of antimicrobial therapy for intra-abdominal infections, a DOOR/RADAR of 66% was found [34], meaning that patients in the control arm had a 66% chance of getting a better

**Table 3 Evans et al. proposed a way to utilize composite endpoints: the desirability of outcome ranking (DOOR) and the response adjusted for the duration of antibiotic risk (RADAR) [33]**

	DOOR/RADAR Evans et al. [33]	Win ratio Pocock et al. [32]
Relevance of the outcome is taken into account	Categorize patients based on overall clinical outcome as pre-specified in categories with increasing severity	Study the more serious event in matched pairs. Did the pair member with the treatment of interest experience it first: assign “loser”. Did the control member experience it first: assign “winner”
Outcome with lower priority is considered	Determine days of antibiotic use and rank patients within categories, whereby a shorter duration results in a lower rank (optional)	If no serious event occurred within a pair, study a less serious event. Who had it first? Determine whether the pair is a “winner” or “loser”
Patients are compared/ranked and effect measure is calculated	Rank all patients over all categories: Determine the number of control patients with a lower rank for each treated patient. Divide this sum by the total number of possible pairwise comparisons: Probability of a better rank for a random patient of the treatment group	Divide the number of “winners” by the number of “losers” to calculate the win ratio

Pocock et al. suggested the win ratio as a new effect measure that takes the different priorities of the components into account [32]. The generic version of DOOR/RADAR is very similar to the win ratio

DOOR/RADAR than the controls. In Fig. 1c, we show that enrolment of 75 patients would already be sufficient to detect this difference.

While Molina and Cisneros state that this design provides valuable information and could be included as a co-primary analysis [35], Phillips et al. warn that considerable evaluation of this method is necessary, as it can easily be manipulated through choice of the categories [36]. It is therefore of key importance that this kind of hierarchical endpoints are clearly defined and published in the statistical analysis plan before unblinding and analyzing the data. Importantly, interpretation of the outcomes and the difference in outcomes between the arms would require pre-trial discussions; it is a novel endpoint and a clinically meaningful difference needs to be determined.

### Components of composite endpoints

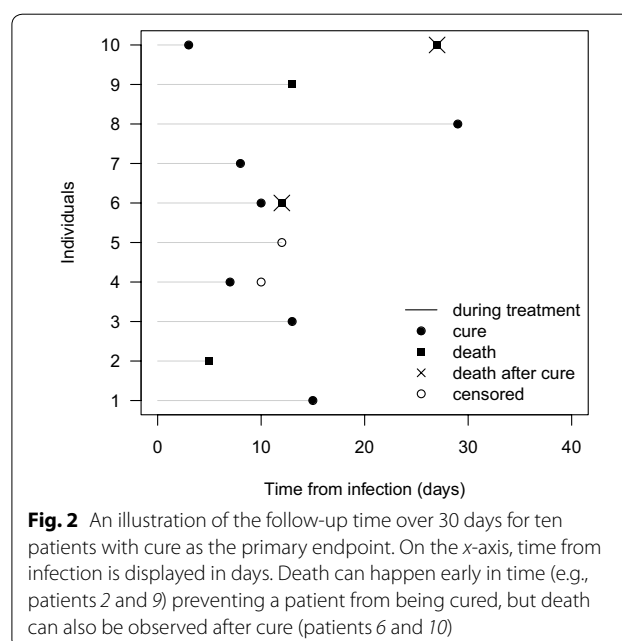
Composite endpoints seem to be a promising strategy; however, its specific components will depend on many factors, including whether it is a pivotal or pragmatic trial, or what type of indication is targeted. In two recent systematic reviews [3, 9], it is proposed to combine mortality and microbiological endpoints for BSIs, and mortality and clinical cure for HAP/VAP, respectively. However, both groups agree more debate is required to improve definitions of clinical cure and to finalize discussions on the importance of patient-centered outcomes versus microbiological or clinical response, as well as the role of safety endpoints.

### Hierarchical nested designs

A hierarchical nested design is proposed by Huque et al. in order to overcome the problems associated with RCTs specifically focused on treatment of infections caused by MDROs [37]. In this design, the primary endpoint (described as a dichotomous one in Huque et al. [37], i.e. mortality or clinical/microbiological cure) is first tested for non-inferiority in the subgroup of patients who have infections caused by pathogens susceptible to the control drug. Then, patients with infections caused by resistant pathogens can be compared using a superiority test, once non-inferiority is confirmed, on the same endpoint. However, since this design is powered for the non-inferiority comparison, the probability of achieving superiority for the patients with infections caused by MDROs is small, given they are likely to be a small proportion of the overall population. Therefore, more sensitive, non-mortality endpoints should be considered in this type of trials for testing the superiority hypothesis in the resistant subgroup [2, 38, 39]. So far, no RCT has implemented a hierarchical nested design to determine treatment efficacy for infections caused by MDROs.

### Statistical advancements in analyzing endpoints

In addition to the complexities related to study design and selection of the most appropriate endpoints, there are also some issues that need to be considered in the analytical phase. The high underlying mortality rate of critically ill patients makes an analytical evaluation of non-mortality endpoints challenging. Harhay et al. have already emphasized that special statistical methods are needed when evaluating a non-mortal clinical endpoint [38]. All non-mortal clinical endpoints are measured over time, i.e. days from randomization, such as clinical outcome at the test of cure visit. Thus, time plays a crucial role in the interpretation and analysis of the data. The main statistical challenge, when observing non-mortal clinical endpoints, is the way patients who die are handled. For instance, when cure is the endpoint of interest, patients who die before being cured, from, e.g., their underlying disease, will have zero chance of cure (Fig. 2, patients 2 and 9) as compared to patients who are simply lost to follow-up and still have a chance of cure (Fig. 2, patient 5). Death prevents observing cure as the primary endpoint and is therefore a competing event for cure [40, 41]. Competing risk methods take these effects into account and provide an appropriate estimation of how the cure probability develops over time. These methods have recently been applied for RCTs in similar settings: for instance, Ayzac et al. studied the impact of prone positioning on the incidence of VAP using a cumulative incidence function that accounts for death as a competing event [42].



**Fig. 2** An illustration of the follow-up time over 30 days for ten patients with cure as the primary endpoint. On the x-axis, time from infection is displayed in days. Death can happen early in time (e.g., patients 2 and 9) preventing a patient from being cured, but death can also be observed after cure (patients 6 and 10)



In addition, underlying all-cause mortality often remains high, even shortly after experiencing the non-mortality clinical endpoint. This requires a deeper understanding of the event dynamics after reaching clinical cure. It can be achieved by a multistate model, an extension of the competing risks model, where a non-mortality endpoint, such as cure, acts as an intermediate event. Such a model is appropriate to examine the whole time-dynamic process of the probability to be cured and alive. In particular, it accounts for the fact that patients might die from an underlying illness where the infection may or may not be contributory, either before or after the scheduled test of cure. In this model, efficacy can be evaluated at several time points simultaneously, comparing a short- with a long-term effect, potentially looking at signs and symptoms to assess how a patient is or not improving during the entire course of treatment.

### Safety endpoints

Finally, it is not just about efficacy of the new treatment, as side effects should also be considered. Traditionally, side effects of new therapies include short-term, individual effects, involving one or more organ systems. However, side effects of antimicrobial prescribing can have a far wider reach, as any use of antibiotics will inevitably have an impact on the flora of the individual patient and the microbial ecology in the environment. Therefore, collateral damage at the individual and population level, in the short and the long term, should be considered in RCTs evaluating new antimicrobial therapies. Secondary safety endpoints that could capture this wider context at an individual level include endogenous resistance development, impact on the microbiome, i.e. colonization with MDROs after treatment, incidence of *Clostridium difficile* infections or super-infections. At the population level, trends in resistance proportions or colonization rates could be important indicators. In the field of pivotal RCTs, this is still uncharted territory, and as such it is very difficult to provide any practical guidance on how to implement this type of safety endpoints in future trials. However, it is clear that assessment of this collateral damage becomes more and more critical, and including some measure of collateral damage should be considered during the design phase of new RCTs.

### Role of the clinician

As stressed previously, the commonly applied 10% non-inferiority margin for mortality is far higher than the 3–5% which would be acceptable by clinicians. Therefore, clinicians should participate in the discussion about applied non-inferiority margins in clinical trials performed in their hospitals. Secondly, it is clear that

mortality should no longer be used as a single primary outcome measure to evaluate new treatment for sepsis, especially in non-inferiority trials. A panel of clinicians, pharmacists, methodologists and patients should be involved in order to come to a clinically meaningful hierarchy of endpoints, which could be proposed to regulatory agencies. The Delphi process is one of the possible ways to come to a consensus on new endpoints. This effort has already been initiated for circulatory failure in sepsis [22] and BSI [3], and is ongoing for HAP/VAP by our group [9].

### Conclusion

Overall, there is a wide range of possible endpoints for evaluating new antibiotic therapies for patients with severe infections. In registration trials, mortality and clinical response tend to be primary endpoints. Non-registration trials have more frequently used microbiological response and antibiotic-free days. However, all of these endpoints suffer from drawbacks, as summarized in Table 2, and are inadequate in the light of improved clinical management and the increasing prevalence of MDROs. There is also a lack of consensus about which endpoint should be primary, and, as such, there is an urgent need to discuss and develop new endpoints that can capture added value of antibiotics in these specific circumstances. These endpoints need to be of importance from the patients' perspective and represent the real-world setting. Currently, composite endpoints, hierarchical nested designs, and competing risks analyses seem to be the most promising new tools for designing and analyzing clinical trials in this area.

For composite endpoints, the applicability of the win ratio and/or DOOR/RADAR should be thoroughly tested for different categorization of objective and more subjective endpoints. If a hierarchical nested design is used, the likelihood of being able to achieve superiority on the resistant pathogen subgroup should be an important consideration in the selection of the endpoint, suggesting that more sensitive endpoints, like composite scores, would be most appropriate. The performance of nested non-inferiority/superiority trials wherein the most applicable endpoint can be tested in each of the subgroups could further improve this design, but this should be tested for performance. As an additional clinical trial requirement, sensitivity analyses should be considered that assess the whole time-dynamic process of the probability to be cured and alive, without applying stringent time limits to the test for cure or determination of life status, using a multistate framework.

The current discussions on the most appropriate endpoints for RCTs in severely ill patients leave developers in an uncertain position when designing new trials. First

of all, if consensus is not reached, future reviews and meta-analyses will not be able to bring together information on new therapies, as the clinical data and evaluated endpoints will differ. Secondly, it postpones discussions about how to handle the transition period from old to new endpoints. In this era of rising resistance levels, slowly refilling antibiotic development pipelines, and an aging population, we need to make sure that RCTs effectively, and in a valid way, determine the added benefit of new antibiotic agents, especially for severely ill patients. Regulatory authorities, pharmaceutical companies, and clinical investigators need to agree soon on the most appropriate clinical endpoints for this vulnerable patient group in order to make sure that new, effective antibiotic agents will become available for medical prescription efficiently and that unnecessary deaths can be avoided.

#### Author details

<sup>1</sup> UMR 1137 IAME Inserm/Université Paris Diderot, 75018 Paris, France. <sup>2</sup> APHP Medical and Infectious Diseases ICU, Bichat Hospital, 46 Rue Henri Huchard, 75018 Paris, France. <sup>3</sup> Infection Control Program, Geneva University Hospitals and Faculty of Medicine, Gabrielle-Perret-Gentil 4, 1205 Geneva, Switzerland. <sup>4</sup> Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. <sup>5</sup> Université Paris Diderot, 75018 Paris, France. <sup>6</sup> APHP Anesthesiology and Critical Care Department, Beaujon Hospital, Paris, France. <sup>7</sup> Department of Biostatistics and Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. <sup>8</sup> AstraZeneca, Macclesfield SK10 4TG, UK.

#### Acknowledgements

We thank Carolina Fankhauser, Caroline Landelle (Geneva, Switzerland), and Martin Schumacher (Freiburg, Germany), who provided helpful input. The research leading to these results was conducted as part of the COMBACTE-NET consortium. For further information, please refer to <http://www.COMBACTE.com>.

#### Compliance with ethical standards

#### Conflicts of interests

JFT has received speaker fees from Astellas, Gilead, Novartis, Merck, and Pfizer, and is Advisory Board member for 3M, Bayer, Abbott, and Merck. SH has received consulting fees from GSK, Janssen, Abbott, and Novartis, as well as research support from Pfizer.

#### Funding

For all authors, the research leading to this manuscript has received support from the Innovative Medicines Initiative Joint Undertaking under Grant agreement No. 115523 [Combating Bacterial Resistance in Europe projects (COMBACTE)], resources of which are composed of financial contribution from the European Union's 7th Framework Programme (FP7/2007 ± 2013) and the European Federation of Pharmaceutical Industries and Associations (EFPIA) companies' in-kind contribution. MW has also received funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) under Grant No. WO 1746/1-2. DW belongs to an EFPIA member company in the IMI JU and costs related to his part in the research was carried by the respective company as an in-kind contribution under the IMI JU scheme. The funders had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

#### Open Access

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the

original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Received: 23 January 2017 Accepted: 12 April 2017

Published online: 2 May 2017

#### References

- Muscledere JG, Day A, Heyland DK (2010) Mortality, attributable mortality, and clinical events as end points for clinical trials of ventilator-associated pneumonia and hospital-acquired pneumonia. *Clin Infect Dis* 51(Suppl 1):S120–S125
- Talbot GH, Powers JH, Fleming TR, Siuciak JA, Bradley J, Boucher H (2012) Progress on developing endpoints for registrational clinical trials of community-acquired bacterial pneumonia and acute bacterial skin and skin structure infections: update from the Biomarkers Consortium of the Foundation for the National Institutes of Health. *Clin Infect Dis* 55:1114–1121
- Harris PN, McNamara JF, Lye DC, Davis JS, Bernard L, Cheng AC, Doi Y, Fowler VG, Jr, Kaye KS, Leibovici L, Lipman J, Llewelyn MJ, Munoz-Price S, Paul M, Peleg AY, Rodriguez-Bano J, Rogers BA, Seifert H, Thamlikitkul V, Thwaites G, Tong SY, Turnidge J, Utili R, Webb SA, Paterson DL (2016) Proposed primary endpoints for use in clinical trials that compare treatment options for bloodstream infection in adults: a consensus definition. *Clin Microbiol Infect*. doi:10.1016/j.cmi.2016.10.023
- U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (2014) Guidance for industry hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia: developing drugs for treatment. Center for Drug Evaluation and Research, Silver Spring
- Garroute Orgeas M, Timsit JF, Soufir L, Tafflet M, Adrie C, Philippart F, Zahar JR, Clec'h C, Goldran-Toledano D, Jamali S, Dumenil AS, Azoulay E, Carlet J (2008) Impact of adverse events on outcomes in intensive care unit patients. *Crit Care Med* 36:2041–2047
- Bekaert M, Timsit JF, Vansteelandt S, Depuydt P, Vesin A, Garroute-Orgeas M, Decruyenaere J, Clec'h C, Azoulay E, Benoit D (2011) Attributable mortality of ventilator-associated pneumonia: a reappraisal using causal analysis. *Am J Respir Crit Care Med* 184:1133–1139
- Pettila V, Hjortrup PB, Jakob SM, Wilkman E, Perner A, Takala J (2016) Control groups in recent septic shock trials: a systematic review. *Intensive Care Med* 42:1912–1921
- Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M (2015) The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 350:h2147
- Weiss E, Zahar JR, Adrie C, Wolkewitz M, Essaïd W, Timsit JF (2015) Severe hospital-acquired and ventilatory-acquired pneumonia treatment: a systematic review of inclusion and judgement criteria used in RCTs. *ECCMID, Copenhagen*
- Sommer H, Wolkewitz M, Schumacher M (2017) The time-dependent 'cure-death' model investigating two equally important endpoints simultaneously in trials treating high-risk patients with resistant pathogens. *Pharm Stat* (in press)
- Niederman MS (2010) Hospital-acquired pneumonia, health care-associated pneumonia, ventilator-associated pneumonia, and ventilator-associated tracheobronchitis: definitions and challenges in trial design. *Clin Infect Dis* 51(Suppl 1):S12–S17
- Harbarth S, Garbino J, Pugin J, Romand JA, Lew D, Pittet D (2003) Inappropriate initial antimicrobial therapy and its effect on survival in a clinical trial of immunomodulating therapy for severe sepsis. *Am J Med* 115:529–535
- Melsen WG, Rovers MM, Groenwold RH, Bergmans DC, Camus C, Bauer TT, Hanisch EW, Klarin B, Koeman M, Krueger WA, Lacherade JC, Lorente L, Memish ZA, Morrow LE, Nardi G, van Nieuwenhoven CA, O'Keefe GE, Nakos G, Scannapieco FA, Seguin P, Staudinger T, Topeli A, Ferrer M, Bonten MJ (2013) Attributable mortality of ventilator-associated pneumonia: a meta-analysis of individual patient data from randomised prevention studies. *Lancet Infect Dis* 13:665–671

14. Timsit JF, Zahar JR, Chevret S (2011) Attributable mortality of ventilator-associated pneumonia. *Curr Opin Crit Care* 17:464–471
15. Cunnion KM, Weber DJ, Broadhead WE, Hanson LC, Pieper CF, Rutala WA (1996) Risk factors for nosocomial pneumonia: comparing adult critical-care populations. *Am J Respir Crit Care Med* 153:158–162
16. Clec'h C, Schwebel C, Francais A, Toledano D, Fosse JP, Garrouste-Orgeas M, Azoulay E, Adrie C, Jamali S, Descorps-Declere A, Nakache D, Timsit JF, Cohen Y (2007) Does catheter-associated urinary tract infection increase mortality in critically ill patients? *Infect Control Hosp Epidemiol* 28:1367–1373
17. Garrouste-Orgeas M, Timsit JF, Tafflet M, Misset B, Zahar JR, Soufir L, Lazard T, Jamali S, Mourvillier B, Cohen Y, De Lasseuse A, Azoulay E, Cheval C, Descorps-Declere A, Adrie C, Costa de Beauregard MA, Carlet J (2006) Excess risk of death from intensive care unit-acquired nosocomial bloodstream infections: a reappraisal. *Clin Infect Dis* 42:1118–1126
18. Januel JM, Harbarth S, Allard R, Voirin N, Lepape A, Allaouchiche B, Guerin C, Lehot JJ, Robert MO, Fournier G, Jacques D, Chassard D, Gueugniaud PY, Artru F, Petit P, Robert D, Mohammedi I, Girard R, Centre JC, Nicolle MC, Grandjean J, Fabry J, Vanhems P (2010) Estimating attributable mortality due to nosocomial infections acquired in intensive care units. *Infect Control Hosp Epidemiol* 31:388–394
19. Bouadma L, Deslandes E, Lolom I, Le Corre B, Mourvillier B, Regnier B, Porcher R, Wolff M, Lucet JC (2010) Long-term impact of a multifaceted prevention program on ventilator-associated pneumonia in a medical intensive care unit. *Clin Infect Dis* 51:1115–1122
20. Gaieski DF, Edwards JM, Kallan MJ, Carr BG (2013) Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med* 41:1167–1174
21. Kaukonen KM, Bailey M, Suzuki S, Pilcher D, Bellomo R (2014) Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000–2012. *JAMA* 311:1308–1316
22. Mebazaa A, Laterre PF, Russell JA, Bergmann A, Gattinoni L, Gayat E, Harhay MO, Hartmann O, Hein F, Kjolbye AL, Legrand M, Lewis RJ, Marshall JC, Marx G, Radermacher P, Schroedter M, Scigalla P, Stough WG, Struck J, Van den Berghe G, Yilmaz MB, Angus DC (2016) Designing phase 3 sepsis trials: application of learned experiences from critical care trials in acute heart failure. *J Intensive Care* 4:24
23. Sorbello A, Komo S, Valappil T, Nambiar S (2010) Registration trials of antibacterial drugs for the treatment of nosocomial pneumonia. *Clin Infect Dis* 51(Suppl 1):S36–S41
24. Bouadma L, Sonnevile R, Garrouste-Orgeas M, Darmon M, Souweine B, Voiriot G, Kallel H, Schwebel C, Goldgran-Toledano D, Dumenil AS, Argaud L, Ruckly S, Jamali S, Planquette B, Adrie C, Lucet JC, Azoulay E, Timsit JF (2015) Ventilator-associated events: prevalence, outcome, and relationship with ventilator-associated pneumonia. *Crit Care Med* 43:1798–1806
25. Dennesen PJ, van der Ven AJ, Kessels AG, Ramsay G, Bonten MJ (2001) Resolution of infectious parameters after antimicrobial therapy in patients with ventilator-associated pneumonia. *Am J Respir Crit Care Med* 163:1371–1375
26. Shorr AF, Cook D, Jiang X, Muscedere J, Heyland D (2008) Correlates of clinical failure in ventilator-associated pneumonia: insights from a large, randomized trial. *J Crit Care* 23:64–73
27. Bouadma L, Luyt CE, Tubach F, Cracco C, Alvarez A, Schwebel C, Schortgen F, Lasocki S, Veber B, Dehoux M, Bernard M, Pasquet B, Regnier B, Brun-Buisson C, Chastre J, Wolff M (2010) Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet* 375:463–474
28. Brunkhorst FM, Oppert M, Marx G, Bloos F, Ludwig K, Putensen C, Nierhaus A, Jaschinski U, Meier-Hellmann A, Weyland A, Grundling M, Moerer O, Riessen R, Seibel A, Ragaller M, Buchler MW, John S, Bach F, Spies C, Reill L, Fritz H, Kiehntopf M, Kuhn E, Bogatsch H, Engel C, Loeffler M, Kollef MH, Reinhart K, Welte T, German Study Group Competence Network S (2012) Effect of empirical treatment with moxifloxacin and meropenem vs meropenem on sepsis-related organ dysfunction in patients with severe sepsis: a randomized trial. *JAMA* 307:2390–2399
29. Viale P, Giannella M, Tedeschi S, Lewis R (2015) Treatment of MDR-Gram negative infections in the 21st century: a never ending threat for clinicians. *Curr Opin Pharmacol* 24:30–37
30. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use Expert Working Group (1998) Guideline E9: "Statistical Principles for Clinical Trials". Available at: [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Efficacy/E9/Step4/E9\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf). Accessed Sept 2016
31. Pogue J, Devereaux PJ, Thabane L, Yusuf S (2012) Designing and analyzing clinical trials with composite outcomes: consideration of possible treatment differences between the individual outcomes. *PLoS ONE* 7:e34785
32. Pocock SJ, Ariti CA, Collier TJ, Wang D (2012) The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 33:176–182
33. Evans SR, Rubin D, Follmann D, Pennello G, Huskins WC, Powers JH, Schoenfeld D, Chuang-Stein C, Cosgrove SE, Fowler VG Jr, Lautenbach E, Chambers HF (2015) Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). *Clin Infect Dis* 61:800–806
34. Celestin A, Odom S, Sawyer R, Cook C (2016) Novel method allows proof of superiority in antibiotic trials using smaller cohorts. 36th annual SIS Meeting. *Surgical infections*, 17, S21
35. Molina J, Cisneros JM (2015) Editorial commentary: a chance to change the paradigm of outcome assessment of antimicrobial stewardship programs. *Clin Infect Dis* 61:807–808
36. Phillips PP, Morris TP, Walker AS (2016) DOOR/RADAR: a gateway into the unknown? *Clin Infect Dis* 62:814–815
37. Huque MF, Valappil T, Soon GG (2014) Hierarchical nested trial design (HNTD) for demonstrating treatment efficacy of new antibacterial drugs in patient populations with emerging bacterial resistance. *Stat Med* 33:4321–4336
38. Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, Cooney E, Mikkelsen ME, Kerlin MP, Small DS, Halpern SD (2014) Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med* 189:1469–1478
39. Price DL, Rubin DB, Valappil T (2015) Antimicrobial products: statistical challenges and opportunities. *Stat Biopharm Res* 7:325–330
40. Beyersmann J, Allignol A, Schumacher M (2011) Competing risks and multistate models with R. Springer, New York
41. Schumacher M, Allignol A, Beyersmann J, Binder N, Wolkewitz M (2013) Hospital-acquired infections—appropriate statistical treatment is urgently needed! *Int J Epidemiol* 42:1502–1508
42. Ayzac L, Girard R, Baboi L, Beuret P, Rabilloud M, Richard JC, Guerin C (2016) Ventilator-associated pneumonia in ARDS patients: the impact of prone positioning. A secondary analysis of the PROSEVA trial. *Intensive Care Med* 42:871–878
43. Fiteni F, Pam A, Anota A, Vernerey D, Paget-Bailly S, Westeel V, Bonnetain F (2015) Health-related quality-of-life as co-primary endpoint in randomized clinical trials in oncology. *Expert Rev Anticancer Ther* 15:885–891
44. Pugh R, Grant C, Cooke RP, Dempsey G (2015) Short-course versus prolonged-course antibiotic therapy for hospital-acquired pneumonia in critically ill adults. *Cochrane Database Syst Rev*: CD007577. doi:10.1002/14651858.CD007577