Benefits of group sequential design and sample size re-estimation for RCTs evaluating the prevention of ventilator-associated pneumonia: A simulation study informed by real world data

Holly Jackson[1], Julien Sauser[1,2], C.H. (Henri) van Werkhoven[3], Stephan Harbarth[1], Marlieke E.A. de Kraker[1]

1. Infection Control Program, Geneva University Hospitals and Faculty of Medicine, World Health Organization Collaborating Center, Geneva, Switzerland
2. Clinical Research Center, University Hospital of Lausanne and University of Lausanne, Lausanne, Switzerland
3. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

Keywords: randomised clinical trial, adaptive clinical trial design, group sequential design, sample size re-estimation, ventilator-associated pneumonia

Corresponding author: Holly Jackson, holly.jackson@hug.ch

Abstract

Background: Ventilator-associated pneumonia (VAP) is an important healthcare acquired infection. Conducting conventional randomised controlled trials (RCTs) on VAP prevention is often challenging, due to low numbers of eligible patients and events per site, especially for pathogen-specific interventions. We explored how group sequential designs (GSD), and sample size re-estimation (SSR) designs could improve RCT efficiency in simulated superiority trials to prevent VAP.

Methods: Simulations were informed using data from the prospective, observational *Hospital Network Study – Preparation for a Randomised Evaluation of anti-Pneumonia Strategies* (HONEST-PREPS). We tested the impact of different GSD and SSR designs on expected sample size (considering early stopping) and maximum sample size (no early stopping). We varied the type of stopping boundary, timepoint of interim analysis (IA), and assumed prevention effect. We included stopping boundaries for efficacy and futility within GSDs. We included efficacy boundaries within SSR, but capped the maximum sample size to be double that of an RCT using the assumed prevention effect. Thus, allowing SSR to stop for futility if the desired power would not be reached, utilising this maximum sample size increase. We applied time-to-event analyses, with effect estimates expressed as hazard ratio (HR).

Results: The estimated 28-day cumulative incidence of VAP was 15.5% in HONEST-PREPS. For a 30% reduction (HR=0.68), a standard RCT (power 80%) would require a sample size of 1291 patients. For GSD, Pocock boundaries result in a smaller expected sample size (E[N]=1128), but a larger maximum sample size (max(N)=1578) than O'Brien Fleming (OBF) boundaries (E[N]=1170 and max(N)=1389). Optimal placement of a single IA was at 48% and 64% of the maximum sample size for Pocock and OBF boundaries, respectively. SSR is more efficient compared to GSD when an incorrect prevention effect is initially used to plan the trial, as it can maintain power closer to the pre-specified desired power.

Conclusions: GSD and SSR are effective adaptive designs, preferable to fixed RCTs in a superiority trial comparing the effectiveness of an investigational intervention with a standard of care in preventing VAP. They can reduce the sample size and should be considered at the trial design stage.

Abstract word count: 345

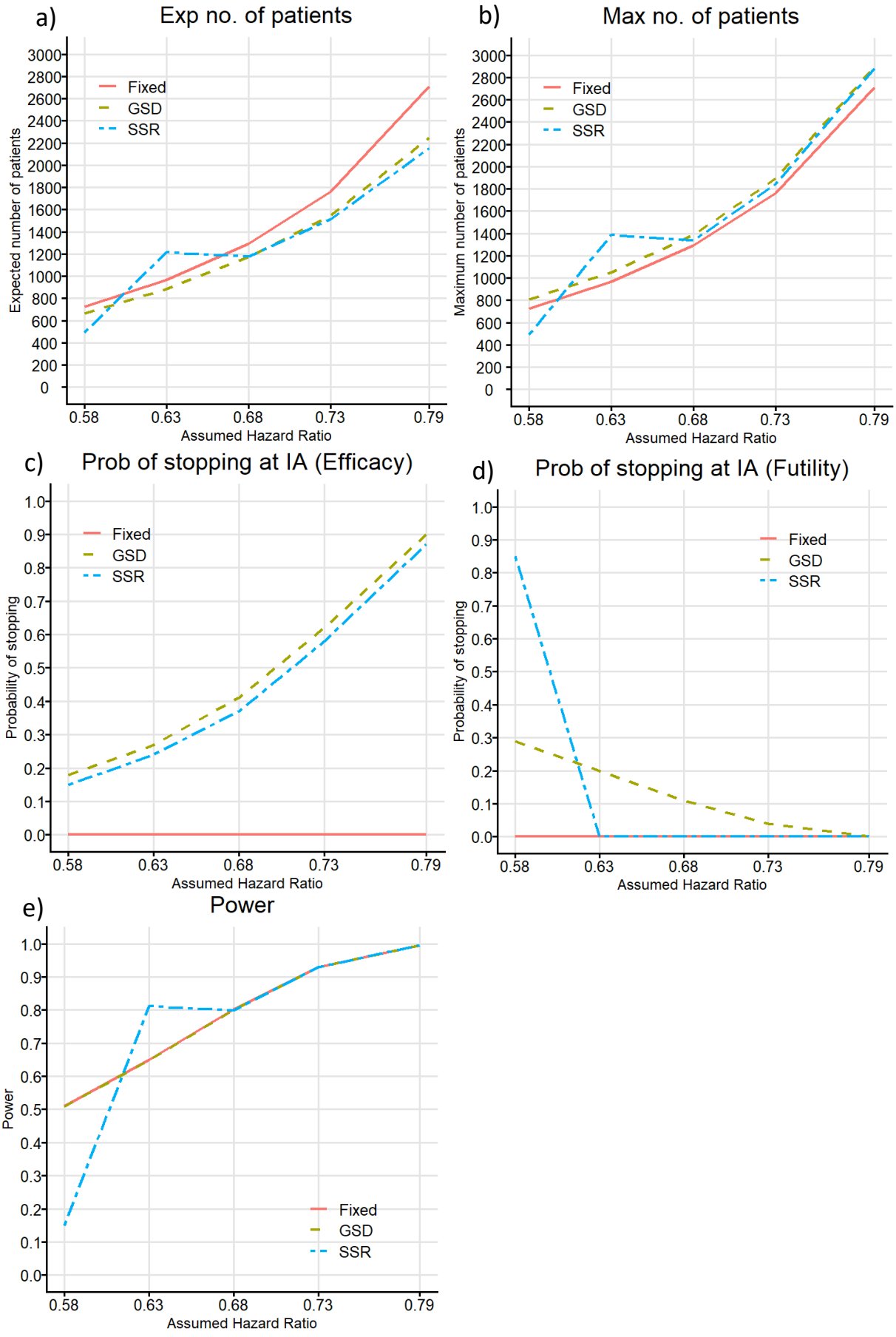
a)
Exp no. of patients
Expected number of patients
Assumed Hazard Ratio
Fixed
GSD
SSR
b)
Max no. of patients
Maximum number of patients
Assumed Hazard Ratio
c)
Prob of stopping at IA (Efficacy)
Probability of stopping
Assumed Hazard Ratio
d)
Prob of stopping at IA (Futility)
Probability of stopping
Assumed Hazard Ratio
e)
Power
Power
Assumed Hazard Ratio

Figure 1: Comparison of group sequential (GSD) and sample size re-estimation (SSR) with O'Brien-Fleming boundaries against fixed randomised controlled trial (RCT) for (a) expected (Exp) number of patients, (b) mean maximum (Max) number of patients, (c) probability (Prob) of stopping for efficacy at the interim analysis (IA), (d) Prob of stopping for futility at the IA and (e) power, for a range of assumed HRs, when the true HR=0.68, for power of 80%, with a single IA at 64% through the trial.