Author's response to reviews

Title: Antibiotic rotation strategies to reduce antimicrobial resistance in Gram-negative bacteria in European Intensive Care Units: A cluster-randomized cross-over trial

Authors:

Pleun J van Duijn (P.J.Duijn-3@umcutrecht.nl)
Marc JM Bonten (M.J.M.Bonten@umcutrecht.nl)

Version: 2 Date: 27 April 2014

Author's response to reviews: see over
To the Editors-in-Chief Doug Altman, Curt Furberg and Jeremy Grimshaw,

Please find the revisions to our manuscript “Antibiotic rotation strategies to reduce antimicrobial resistance in Gram-negative bacteria in European Intensive Care Units: A cluster-randomized cross-over trial” as per the comments made by Graeme MacLennan.

1. The primary hypothesis is not stated anywhere in the protocol. ClinicalTrials.gov lists the two interventions as active arms, which is fine, but is not clear what the primary comparison of interest is from the protocol. It should be a test of superiority comparing Mixing to Cycling?

**Answer:** The primary comparison is indeed the superiority of Mixing versus cycling. The null-hypothesis therefore states that there is no difference in the primary outcome, mean prevalence of antimicrobial resistance. Because it is not known which intervention is better than the other, the sample size will be tested two-tailed. The primary objective and hypothesis are now described under “study objectives”.

2. Outcomes: In the outcome section there is a description of how the primary outcome will be collected, but not an explicit statement on the primary outcome.

At http://clinicaltrials.gov/show/NCT01293071 where the trial is registered the primary outcome is described as "Mean prevalence of ICU patients colonized with antimicrobial resistant Gram-negative pathogens ". I would therefore expect to see the opening sentence in the Outcomes section along the lines of "The primary outcome is mean prevalence of ICU patients colonised with antimicrobial resistant Gram-negative pathogens. This is defined as ....". We then need a definition of mean prevalence; I am guessing the mean of the nine months of serial point prevalence measurements during each intervention period? Similarly for the secondary outcome are listed explicitly at clinicaltrials.gov, but not so in the manuscript.

**Answer:** The primary and secondary outcomes were concisely described under study objectives. They have been moved from heading “study objectives” to “outcomes”, also a more detailed description of outcome measures is added.

3. Sample size: The sample size section is very confusing and currently as written the calculation cannot be replicated. One issue is that the sample size is described as that of a cluster randomised trial where clusters are randomised but data are collected at the level of the individual. That is true here, but the analysis section (more on that below) and the primary outcome appear to be aggregated up to a cluster-level analysis. There is nothing wrong with this, but readers will notice sample size and analysis plan do not match. It is perfectly reasonable to use a more conservative sample size than required by the primary outcome, but that needs explained. However, even so, the details provided as not adequate to replicate. Firstly, there is no detail provided on the sample size calculation unadjusted for clustering. It is stated that 200 patients per intervention period are required to detect a relative decrease of 10% with 95% certainty. Because of uncertainty around the definition of the primary outcome it is difficult to understand what this means.

Is this a sample size for a difference in means (as the analysis section would suggest that primary outcome is being treated as continuous);

we need to know what the baseline value is in this calculation; where does 10% relative reduction come from; what does 95% certainty mean (do you mean power here?); does that mean that 800 patients are required in total ignoring clustering; essentially all the ingredients and assumptions underpinning the basic sample size calculation need included but are not currently.

As the authors have correctly noted, once a plain sample size is decided upon it needs adjusted for clustering, this requires an ICC (reported) and then usually one or other of the number of clusters or
The average cluster size is fixed (due to logistic or financial restraints). There is no description of what was fixed and why.

The final sample size is stated as 1476 cultures per intervention arm.

Here there is a change of terminology, now it is not cultured patients, but cultures, does that mean that the same patient could provide more than one culture if they were in the ICU long enough? There is a change from intervention period to intervention arms, this is confusing.

The last sentence in this section then states that study should be adequately powered to detect a 6% difference prevalence with 95% certainty. Is this now a 6% absolute difference? If not, why the change from 10% being important to 6%? What does adequate power mean? Again, I am not sure what 95% certainty refers to in this context?

One last point, the discussion talks about approximately 10,000 patients being recruited, which would seem to contradict the sample size section.

Answer: We thank the reviewer for addressing so carefully. After rereading the text of our initial manuscript and the original protocol, we can only admit that the description as such was not clear nor reproducible. This resulted, partly, from our own misinterpretation, but also from the unclarity of the protocol text. After ample deliberation, we have decided to perform our own “quasi-post-hoc” sample size calculation following as much as possible the original protocol text. We have assumed a binomial distribution, used an absolute reduction of 10% as in the text, and assumed 2 absolute $H_0$ and $H_1$ values of 55% to 45%. The $H_0$ and $H_1$ values were chosen to liken a worst-case scenario where the distribution was as wide as possible and the sample size, therefore, is the largest. For a binomial distribution, this is at 0.5. We used a power of 80% with a significance level of 95% tested two-tailed. From this we conclude needing 392 patients per treatment arm. See the adjustments in the text under the heading “sample size calculation”.

The average cluster size was described in the text (270 samples per ICU for both treatment arms). The number was calculated using an assumed average ICU size of 15 beds, and, following from the outcome measurement protocol, 15 screened patients for 9 monthly point-prevalence measurements. As a result, for one treatment arm, $15 \times 9 = 135$ samples are assumed.

The number of 10,000 patient inclusions is a multitude of the number of patients screened, because point-prevalence measurements are done only once a month, and the average length of stay is only around 1 week. As such, the main outcome measure of the trial is not collected from all patients, but secondary outcomes are collected over all these patients, and is therefore mentioned in the discussion.

4. Analysis: There are two analyses to be carried out: cluster-level and individual-level. It is not clear how the primary outcome will be analysed.

I would expect here to see an explicit statement on the method(s) to be used. For example, there are 8 ICUs, each getting both intervention, therefore a simple analysis would be a paired t-test on the 8 paired observation periods aggregated up over each 9 month period? Is that what is planned? Or are the serial measurements going to used here, it is just not clear. For the individual-level analyses not enough detail is provided. Merely stating “advanced regression methods” is not enough, which methods?

Answer: With regards to the primary analysis: because the outcome of the measurements is dichotomous, assumed to be binomially distributed and with the crossover design reducing inter-cluster correlation, a McNemar test for dependent pairs will be used. With this approach all the data points will be used individually (with a dependent t-test only point-prevalence averages would have been used ignoring how many patients were screened per measurement (ICU-size will differ and some beds will be vacant). Because bi-variate tests do not take into account inter-ICU differences in intervention effects and trends over time, a mixed effects model will be used with random effects per ICU and time-trends per intervention period using the 9 measurements per ICU per study period longitudinally. This will also
facilitate additional adjustment for imbalances in study arm case-mix (patient- and ICU-characteristics) not caused by the interventions themselves.

For Acquisition rates of ARGNB, bacteraemia with ARGNB and mortality, a McNemar’s test will be performed first, followed by Cox-proportional hazard regression models accounting for differences in time-at-risk and using random effects for the different ICUs. A McNemar’s test and generalized linear regression will be used for the proportion of patients receiving appropriate empirical treatment based on the same argumentation. Mean length of ICU-stay will be tested using the t-test for dependent means and if necessary with Linear mixed effects regression analyses, again with random effects for individual ICUs.

Minor points:
1. There are no page numbers.

Answer: Page numbers have been added

2. Reference 29 is the same as reference 25

Answer: Has been corrected

3. The tense that the manuscript is written is inconsistent, for example Data collection section, (my emphasis) "Dedicated staff WILL collect data..." (correct) compared to "Weekly point-prevalence measurements of antibiotic consumption ARE collected..." (incorrect). There ARE in this last sentence should read "WILL BE". This is a protocol which is pre-specifying what will take place in the study, written before the study commences. (Although from ClinicalTrials.gov recruitment is finished? Start date and end date of recruitment would be useful to know in the manuscript).

Answer: Please see corrections in the text

4. Design effect is 3.69 uses European convention of decimal comma, please change to decimal point.

Answer: Please see corrections in the text

Additional adjustments:
- Table 2 has been added stating the definitions of resistance for the primary outcome.
- Added a comment on carry-over and period effects under “Analysis”.
- “Cross-over” has been replaced by “crossover” based on spelling in the CONSORT 2010 statement: extension to cluster randomized trials.
- Added a comment on the effect of a crossover design on the intra-ICU correlation under “Study design”.
- With your permission, the final two concluding sentences were reformatted for better form

We thank you again for your comments, and hope to hear from you, at your convenience. For any additional information, correspondence can be directed to: P.J.Duijn-3@umcutrecht.nl

Best regards,

P.J. van Duijn
M.J.M Bonten