PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Study Protocol: NITric oxide during cardiopulmonary bypass to improve Recovery in Infants with Congenital heart defects (NITRIC trial): A Randomised Controlled Trial.
AUTHORS	Schlapbach, Luregn; Horton, Stephen; Long, Debbie; Beca, John; Erickson, Simon; Festa, Marino; d'Udekem, Yves; Alphonso, Nelson; Winlaw, David; Johnson, Kerry; Delzoppo, Carmel; van Loon, Kim; Gannon, B; Fooken, Jonas; Blumenthal, Antje; Young, Paul; Jones, Mark; Butt, Warwick; Schibler, Andreas

VERSION 1 – REVIEW

REVIEWER	David Faraoni, MD, PhD Hospital for Sick Children, Toronto, Canada
REVIEW RETURNED	10-Jan-2019

GENERAL COMMENTS	In this manuscript, the authors are describing the study protocol of the NITRIC trial, a prospective randomized study assessing the effect of nitric oxide during cardiopulmonary bypass to improve recovery in infants with CHD. This is an interesting study, and I'm looking forward to seeing the results published. The manuscript is well written and study protocol is adequately described. Suggestions: . The idea behind the use of NO on bypass is the attenuation of the SIRS. The primary endpoint (time on the ventilator) is non-specific, and prolonged could be explained by several other factors than SIRS. . The authors mention that some centers will be measuring biomarkers of SIRS. The authors should better describe the biomarkers and that protocol, as this is certainly more important than non-specific clinical endpoints. [page 13]: Why did the authors design and power to study with a non-specific primary endpoint, while the incidence of LCOS and/or the need for ECLS are more clinically relevant (also in term of cost)? Please, discuss. [page 16]: 'The postoperative care and decisions on inotropes and other vasoactive drug delivery, fluid management, renal replacement therapy, iNO therapy or indication for ECLS will be performed as per site specific standard protocols of care.'. Again, why not

	standardized? [page 16]: I'm surprised the anesthesia protocol is not standardized. Some medications used intraoperatively have been shown to have an effect on SIRS. [page 19]: Please, describe the biomarkers that will be measured. Why not including all patients, instead of a subset?

REVIEWER	John Pappachan Southampton Children's hospital United Kingdom
REVIEW RETURNED	29-Jan-2019

GENERAL COMMENTS	Following on from their pilot single centre study, this protocol presents some very important refinements. the primary endpoint is much more clinically relevant and represents a defined robustly measurable variable (VFD-28) that has clinical and health economic importance. However I am concerned that the power calculation is extremely optimistic and strongly recommend that there is a formal statistical review. The pilot study referred to (Ref 26) showed an approximately 66 hour [2.75 day] difference (0.33SD). Assuming a normal distribution the power calculation using an expected difference of 0.2SD would mean an expected treatment related increase of 1.83 VFD-28. It would be useful to know if the VFD-28 of children <2years is normally distributed (I suspect not) and what the mean/median VFD-28 is. In a recent study of remote ischaemic preconditioning in children undergoing elective surgery for congenital heart disease involving CPB conducted in Toronto (admittedly of older children with a median age of 3.1 years and 2 years in the sham and treatment groups respectively); Remote Ischemic Preconditioning in Children Undergoing Cardiac Surgery With Cardiopulmonary Bypass: A Single-Center Double-Blinded Randomized Trial Article in Journal of the American Heart Association · June 2014, the median PICU length of stay was only 22 hours. Unless the intended study population has very much longer median lengths of stay and thus VFD-28, the present study may be very underpowered. Apart from this concern this study seems very well constructed, safe and practicable to do as well as importantly being double blinded. The incidence of LCOS which is notoriously difficult to measure subjectively has been demoted to a secondary outcome measure and its definition using a lactate>4 mmol/L, an AV DO2 >35 and a VIS of >15 (as opposed to >10, which was used in the pilot and have included children on 10mcg/kg/min Dopamine and 0.5mcg/kg/min Milrinone which is a standard post-operative inotrope combination in many centres) seems far more sensible as does the decision to collect data on the initiation of ECLS separately. The safety monitoring and interim analyses seem well planned. I suggest a major revision only in that the power analysis and a formal statistical review of it are fundamental to the potential success or failure of this study

REVIEWER	James Khan Mount Sinai Hospital, Department of Anesthesia, University of Toronto, Canada
REVIEW RETURNED	10-Apr-2019

GENERAL COMMENTS	• Inclusion criteria: in Table 1, add “Open” to “Elective Heart surgery” • Exclusion criteria: “chronic ventilator dependency” is slightly vague. Is there any criteria that you will use to define this more? • Blinding: It is described throughout the manuscript that this is a double-blinded RCT. On page 14, it states that parents and caregivers will be blinded, identifying the two groups that are blinded. However, I would go further and state who else is blinded (i.e., surgeons, anesthesiologists, research assistants, data analysts). This is also a requirement as part of CONSORT guidelines. • Page 14 line 17-24: the rationale for stratification variables is described. Are there any references that you can add to support these statements? • Page 15 line 41: Please describe whether non-invasive ventilation will be considered as ventilator-free. • Sample size calculation: it is reported that the pilot demonstrated that NO had a 66 hour increase in ventilator free days. Can you clarify whether this 66 hour increase relates to ventilator free days of the 28 day period within randomization in the pilot study. Further, reporting hours for a variable that is expressed in days is a bit confusing – for consistency, report ventilator free days in days. Also, is the 0.33 SD in hours or days, because later in the paragraph you describe using an increase in 0.2 SD as the minimally clinically small effect size. If the 0.2 SD increase is calculated on the summary measures of hours previously mentioned, then the effect size is very small. In reporting the sample size calculation, please report mean VFDs used for the intervention and control groups. • Page 19 Line 37: Please add the “U” in the Mann-Whitney U test Small edits: • Abstract Page 3 Line 31: Don’t start a new sentence with a number in numerical form

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: David Faraoni, MD, PhD
Institution and Country: Hospital for Sick Children, Toronto, Canada
In this manuscript, the authors are describing the study protocol of the NITRIC trial, a prospective randomized study assessing the effect of nitric oxide during cardiopulmonary bypass to improve recovery in infants with CHD. This is an interesting study, and I’m looking forward to seeing the results published. The manuscript is well written and study protocol is adequately described.

Suggestions:

4. The idea behind the use of NO on bypass is the attenuation of the SIRS. The primary endpoint (time on the ventilator) is non-specific, and prolonged could be explained by several other factors than SIRS.
REPLY 4: We agree that the primary endpoint is not specific to the intervention, and will be influenced by many factors. Similarly, in most large ICU trials, a multitude of variables are expected to impact on the primary outcome. Ventilator free days (VFD) was chosen as the primary outcome, as it is a) an objective and robust outcome measure significantly impacting on patients, b) directly related to severity post intervention, taking both early mortality and morbidity into account, and c) strongly associated with late mortality and late morbidity, including risks such as hospital-acquired infections and sedation-related neurotoxicity, and well known to correlate with impaired long-term outcomes. VFD fulfills SMART criteria (Specific, Measurable, Achievable, Realistic and Timely). Low cardiac output syndrome (LOCS) and factors contributing to LCOS such as SIRS, are directly related to VFD, as infants usually do not tolerate weaning or fail extubation if LCOS and organ dysfunction have not resolved. SIRS is non specific and seen in most patients admitted to ICU (Kaukonen NEJM 2016, Schlapbach ICM 2018), which is why SIRS criteria are not used as outcomes. Ongoing need for ventilation binds patients to PICU, with increasing complications and sequelae associated with longer duration – all indeed directly patient relevant outcomes. Healthcare costs are directly related to mechanical ventilation, which can only be provided in PICU (daily costs of 6,200 AUD vs 1,200 AUD for children managed on the ward in the setting of participating study sites).
The secondary composite outcome of LCOS/ECLS/death captures LCOS with requirement for ECLS (and/or LCOS with mortality). While LCOS in isolation is associated with worse patient outcomes, it is a laboratory/medical entity rather than a patient centered outcome.
Finally, we note that Reviewer 2 highlights the benefits of chosing VFD as a more robust outcome in comparison to LCOS which was used in the pilot study (see Reviewer Point 10).

5. The authors mention that some centers will be measuring biomarkers of SIRS. The authors should better describe the biomarkers and that protocol, as this is certainly more important than non-specific clinical endpoints.
REPLY 5: For the rationale of using clinical outcomes, see REPLY 4. A detailed description of the biomarkers tested and statistical analyses of inflammation and coagulation markers will be separately submitted. The primary and secondary outcomes of the trial are clinical, which is why the present protocol focussed on the clinical aspects. Biobank analyses are anxillary, which is clarified in the revised protocol. Blood at sites which perform biobanking is being aliquoted and stored using study SOPs. The laboratory analyses plan in the cohort is will be finalized once the outcome from a pending grant submission to fund more extended laboratory investigations is knowsn, which will be prior to completion of recruitment of the study.

6. [page 13]: Why did the authors design and power to study with a non-specific primary endpoint, while the incidence of LCOS and/or the need for ECLS are more clinically relevant (also in term of cost)? Please, discuss.
REPLY 6: See as well REPLY 4. We have expanded the section explaining the rationale for the choice of the primary and secondary outcomes (page 14). While LCOS in isolation is associated with worse patient outcomes it is a laboratory/medical entity rather than a patient centered outcome. Note that we are applying higher threshold to define LCOS (higher inotrope score required) than the pilot (James et al, ICM 2016), to ensure we are capturing severe disease relevant for patients and healthcare systems. VFD will be affected by ECLS too as all ECLS patients are ventilated. Similar to VFD, ECLS is not specific – many ECLS runs occur due to factors that are not necessarily related to the host response to CPB, such as surgical repair difficulties, pre-operative instability, or arrhythmia.

7. [page 16]: ‘The postoperative care and decisions on inotropes and other vasoactive drug delivery, fluid management, renal replacement therapy, iNO therapy or indication for ECLS will be performed as per site specific standard protocols of care.Again, why not standardized?
REPLY 7: This is a pragmatic multicentre trials across six sites in three countries, including the five paediatric cardiac surgical services in Australia and New Zealand. This is the first large population-based RCT in the field of paediatric cardiac surgery that we are aware of. If NO has a beneficial impact on recovery after CPB surgery in infants then we expect this effect to be measurable even if some variation in local practice outside CPB exists. This is an important aspect for future generalizability of findings.
While the inter-institutional variability of care for cardiac surgical patients in OT and PICU may be less compared to large regions such as North America, Europe, or Asia, mandating a study protocol on aspects outside the intervention onto each of the participating sites would not have been feasible.

8. [page 16]: I’m surprised the anesthesia protocol is not standardized. Some medications used intraoperatively have been shown to have an effect on SIRS.
REPLY 8: See as well REPLY 7. The purpose of pragmatic trials is to measure the effect of relatively simple interventions without much alterations to usual site specific practice outside the intervention. We agree that a multitude of practices and patient factors, such as pre-operative status and management, stability during induction, blood product management, use of steroids etc all will affect SIRS. SIRS is highly non specific and was abandoned even in the definition of sepsis (Singer et al JAMA 2016). Note that the study RedCap database captures a broad range of descriptors of support pre/during/post-operatively.

9. [page 19]: Please, describe the biomarkers that will be measured. Why not including all patients, instead of a subset?
REPLY 9: see as well REPLY 5. The New Zealand site is not collecting biomarkers, due to indigenous HREC approval restrictions on sending blood of children of Maori/Pacific Islander Background outside New Zealand, and due to related costs. All other sites are collecting biomarkers. Parents need to give separate consent for biobanking - we have added this to the Revision to make it clearer. The biobanking of serum, EDTA, and PAXgene tubes allows us to investigate both candidate and discovery studies across protein/metabolic and genetic markers. A detailed description of the biomarkers tested and the involved analyses for this anxillary study is beyond the scope of the main clinical trial and will be separately submitted prior to completion of recruitment of this study.

Reviewer: 2
Reviewer Name: John Pappachan
Institution and Country: Southampton Children's hospital
United Kingdom
10. Following on from their pilot single centre study, this protocol presents some very important refinements.The primary endpoint is much more clinically relevant and represents a defined robustly measurable variable (VFD-28) that has clinical and health economic importance.
REPLY 10: We thank the Reviewer for his comment, indeed these were the reasons why VFD was chosen as a primary outcome.

11. However I am concerned that the power calculation is extremely optimistic and strongly recommend that there is a formal statistical review. The pilot study referred to (Ref 26) showed an approximately 66 hour [2.75 day] difference (0.33SD). Assuming a normal distribution the power calculation using an expected difference of 0.2SD would mean an expected treatment related increase of 1.83 VFD-28. It would be useful to know if the VFD-28 of children <2years is normally distributed (I suspect not) and what the mean/median VFD-28 is. In a recent study of remote ischaemic preconditioning in children undergoing elective surgery for congenital heart disease

involving CPB conducted in Toronto (admittedly of older children with a median age of 3.1 years and 2 years in the sham and treatment groups respectively); Remote Ischemic Preconditioning in Children Undergoing Cardiac Surgery With Cardiopulmonary Bypass: A Single-Center Double-Blinded Randomized Trial Article in Journal of the American Heart Association · June 2014, the median PICU length of stay was only 22 hours. Unless the intended study population has very much longer median lengths of stay and thus VFD-28, the present study may be very underpowered.
REPLY 11: We have carefully reviewed the pilot study results, and the sample size calculations. These were recalculated by the study statistician, Mark Jones, and the study trialist, Paul Young from the ANZICS Clinical Trial Group.
This study is only including infants below 2 years of age - the median age in the pilot study was 5 months - and hence will include a substantial proportion of higher risk patients undergoing procedures expected to require longer duration of ventilation and longer PICU LOS (such as Norwood Stage 1, arterial switch operations etc). The power calculations are based on the pilot study of 134 patients aged < 2 years from Royal Children`s Hospital Melbourne, Australia. In the pilot study several patients died which means for these patients VFD = 0. Therefore the standard deviation for VFD was large (8.1 days in control group).
This full trial will consider patients who die as zero VFDs. We submit that a clinical difference of 1.66 VFDs (approximatively 40-hour) represents a minimally important clinical difference. Moreover, this effect size is around 60% of the magnitude of the treatment effect of NO on VFDs seen in our pilot study and thus, as highlighted by Reviewer 3, is appropriately conservative. This allows for temporal trends of reducing duration of ventilation following surgery.
We have set our minimum sample size at 1320 participants. The chosen sample size of 1320 participants will provide 90% power to detect a difference of 1.66 VFDs using a two tailed hypothesis at an alpha of 0.05. This allows for a 15% inflation in the sample size to account for the expected non-normal distribution of VFDs, and an additional 10% of participants to allow for drop-outs. In addition this sample size includes a 15% increase to account for a non-normal distribution of VFD.

The respective section (page 15/16) was changed as follows: “Sample size. A pilot study of 134 patients aged < 2 years showed an approximate 2.74 days (66 hours) increase in ventilator-free days (VFD) associated with the study intervention [26]. This includes patients who dies who were considered as zero VFD. The VFD increase associated with the intervention represents an effect size of 0.33SD based on a standard deviation of 8.1 days in the pilot study control group. Based on the primary outcome measure VFD, 1,320 patients (660 per group) would be required to demonstrate a significant increase in VFD assuming a minimally clinically significant small (0.2SD) effect size (1.66 days or 40 hours), 90% power, two-sided alpha level of significance of 5%, 10% withdrawals, and 15% increase in sample size to account for a non-normal distribution of VFD. In Australia and New Zealand approximately 800 of children < 2 years of age undergo surgery for a congenital heart defect requiring CPB each year, including patients with multiple procedures. The consent rate of eligible patients was 78% in the pilot trial[26]. With an expected conservative estimate 60% enrolment rate of eligible patients we expect a 3.5 recruitment period for the study.”

12. Apart from this concern this study seems very well constructed, safe and practicable to do as well as importantly being double blinded. The incidence of LCOS which is notoriously difficult to measure subjectively has been demoted to a secondary outcome measure and its definition using a lactate>4 mmol/L, an AV DO2 >35 and a VIS of >15 (as opposed to >10, which was used in the pilot and have included children on 10mcg/kg/min Dopamine and 0.5mcg/kg/min Milrinone which is a standard post-operative inotrope combination in many centres) seems far more sensible as does the decision to collect data on the initiation of ECLS separately.
The safety monitoring and interim analyses seem well planned.
I suggest a major revision only in that the power analysis and a formal statistical review of it are fundamental to the potential success or failure of this study

REPLY 12: We thank the reviewer for these comments.

Reviewer: 3
Reviewer Name: James Khan
Institution and Country: Mount Sinai Hospital, Department of Anesthesia, University of Toronto, Canada

Please leave your comments for the authors below
13. Inclusion criteria: in Table 1, add “Open” to “Elective Heart surgery”
REPLY 13: The change has been made as requested.

14. Exclusion criteria: “chronic ventilator dependency” is slightly vague. Is there any criteria that you will use to define this more?
REPLY 14: Patients treated with non-invasive or invasive ventilation continuously for >28 days prior to cardiopulmonary bypass are considered chronic ventilator dependency. We have added this statement.

15. Blinding: It is described throughout the manuscript that this is a double-blinded RCT. On page 14, it states that parents and caregivers will be blinded, identifying the two groups that are blinded. However, I would go further and state who else is blinded (i.e., surgeons, anesthesiologists, research assistants, data analysts). This is also a requirement as part of CONSORT guidelines.\
REPLY 15: The change has been made as requested. “Cardiologists, cardiac surgeons, anesthesiologists, intensivists, PICU nurses, research assistants, data analysts, and parents and caregivers will be blinded for the intervention.”

16. Page 14 line 17-24: the rationale for stratification variables is described. Are there any references that you can add to support these statements?
REPLY 16: The according references have been added.

17. Page 15 line 41: Please describe whether non-invasive ventilation will be considered as ventilator-free.
REPLY 17: Treatment with non-invasive ventilation and high-flow nasal cannulae will not be considered as ventilator days. We have added this to Table 2 for clarification, and to the main text (page 13).

18. Sample size calculation: it is reported that the pilot demonstrated that NO had a 66 hour increase in ventilator free days. Can you clarify whether this 66 hour increase relates to ventilator free days of the 28 day period within randomization in the pilot study. Further, reporting hours for a variable that is expressed in days is a bit confusing – for consistency, report ventilator free days in days. Also, is the 0.33 SD in hours or days, because later in the paragraph you describe using an increase in 0.2 SD as the minimally clinically small effect size. If the 0.2 SD increase is calculated on the summary measures of hours previously mentioned, then the effect size is very small. In reporting the sample size calculation, please report mean VFDs used for the intervention and control groups.
REPLY 18: We thank the reviewer for this suggestion. Indeed the 2.74 days (66 hour) increase in VFD in the pilot refers to the 28d period. We have changed the unit to “XX days” in the main text with “(XX hours)” added. In the pilot study, the mean VFD was 22.55 days in the control group versus 25.29 in the intervention group.

19. Page 19 Line 37: Please add the “U” in the Mann-Whitney U test
REPLY 19: The change has been made as requested.

20. Small edits:
Abstract Page 3 Line 31: Don’t start a new sentence with a number in numerical form
REPLY 20: The change has been made as requested.

VERSION 2 – REVIEW

REVIEWER	James Khan University of Toronto, Canada
REVIEW RETURNED	11-Jun-2019

GENERAL COMMENTS	Authors have made revisions that has led to a much improved manuscript.