PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	PARROT Ireland: (Placental growth factor in Assessment of women with suspected pre-eclampsia to Reduce maternal morbidity: a Stepped Wedge Cluster Randomised Control Trial) Research Study Protocol
AUTHORS	Hayes Ryan, Deirdre; Hemming, Karla; Breathnach, Fionnaula; Cotter, Amanda; Devane, Declan; Hunter, Alyson; McAuliffe, Fionnuala; Morrison, John; Murphy, Deirdre; Khashan, Ali; McElroy, Brendan; Murphy, A; Dempsey, E; O'Donoghue, Keelin; Kenny, Louise

VERSION 1 – REVIEW

REVIEWER	Joao Guilherme Bezerra Alves Instituto de Medicina Integral Prof. Fernando Figueira (IMIP), Brazil
REVIEW RETURNED	30-Apr-2018

GENERAL COMMENTS	Congratulations for the project. I only have few questions. Background: Second paragraph, third and fourth sentences need references. VEGF-A was not previously introduced. Health economic assessment was only approached in objectives. Methods and Design: Study design - first sentence: "...prior to 37 weeks gestation" ; 20-37 weeks ? Trial registration is in page 4 and not page 1 (SPIRIT page 34). Secondary outcome measure - Quality of life is included as an economic avaluation (economic impact) (?). By the size of the study and possible benefits to the intervention group I suggest a data and safety monitoring boarder (DSMB) to ensure the bioethical principle of beneficence. Discussion: Possible limitations of the study were not described; it is also missing in the abstract.

REVIEWER	Sarah Stock University of Edinburgh, UK
REVIEW RETURNED	06-May-2018

GENERAL COMMENTS	1. Overall a considered and clear study protocol for a pragmatic trial 2. It was not clear what the pathway of care is for a woman who does not consent for participation in the trial (presumably usual care)

	3. Is there any risk of contamination through off protocol use of the test in clusters randomised to control - and how will this be assessed/managed? 4. What evidence base is there for the estimates of the baseline incidence of the primary outcome(s), and the anticipated reduction in this/these with use of PlGF? 5. I am not qualified to comment on the statistics in detail. 7 clusters seems a small number to detect a big effect, but I will defer to specialist statistical reviewer regarding this. I do note 80% power (with a reduction if the ICC estimates are too conservative), thus a reasonable chance of failing to detect a real effect. Perhaps this could be discussed in discussion.

REVIEWER	Professor Lucy Chappell King's College London, UK I am Co-Chief Investigator on the PARROT-UK study, and provided the PARROT-Ireland co-investigators with a copy of the PARROT-UK protocol, database and trial handbook on their request.
REVIEW RETURNED	09-May-2018

GENERAL COMMENTS	I acknowledge that this is a study protocol of a trial underway, that has been through external peer-review and ethics committee scrutiny, and therefore there is limited scope to amend at this stage. The following are therefore intended to increase clarity only and have no substantial impact on the trial design. 1. Study design a. The authors state: 'The PlGF measurement is reported as 12 the absolute value in pg/ml within 30 minutes of sampling.' This would suggest that results reporting beyond 30 mins is a protocol deviation, but I would be surprised if the investigators can achieve this in all sites. Please could the authors clarify in the protocol whether a) this time frame is essential and b) how they would account (statistically) for protocol deviations if the time frame is shown to be longer than this (e.g. for the scenarios where reporting is within 2 hours of sampling, or within 12 hours etc.). b. Please could the authors clarify whether they anticipate any impact of the algorithm for Belfast being different in the control arm and what steps they might need to take to mitigate this (e.g. in their analysis)? 2. Outcomes a. The authors state 'For maternal morbidity assessment, the fullPIERS score is used with the addition of severe hypertension (Table 3).' The authors should clarify more clearly that they have adapted the fullPIERS composite, in particular by including ICU admission (which is a process outcome that depends on the health service setting, rather than a clinical outcome), and that they are using a different definition of hepatic dysfunction (based on ALT rather than INR) that will result in a much higher proportion of cases identified as having maternal morbidity. Please could the authors clarify these adaptations in the protocol and also confirm that there will be no double counting (i.e. individuals not events will be presented for the composite). b. The authors say 'For neonatal morbidity assessment, babies are dichotomised into having or not having objectively identified

	neonatal morbidity by means of a composite neonatal score (Table 4).' I am aware (from neonatology colleagues and inclusion in my own trials) that a number of these outcomes can be difficult to adjudicate, and/ or have various stagings, where the early stage is often not included in a definition of morbidity but later stage disease would (e.g. IVH and NEC), and/or where definitions have evolved with new international consensus definitions now available (e.g. FGR). Please could the authors clarify a) whether all stages of each disease included will be part of the composite outcome; b) which definitions they are using (e.g. for FGR/ customisation of BW centile); c) whether they plan to have an adjudication committee or use local case note review of clinical diagnoses by midwives or doctors only. Depending on the clarification, the authors may wish to modify 'objectively identified' in the phrase above, as many would consider a number of the morbidity measures to have an understandable degree of subjectivity. However, I am surprised by their inclusion of a single antenatal component: 'Fetal growth restriction identified on antenatal ultrasound', as any cases that are labelled as such but do NOT have any other of the listed perinatal components of the composite might well be a test false positive, that drives clinical intervention, without associated morbidity. It would be useful to see justification of this criterion. The authors should clarify that there will be no double counting, particularly as there is a mix of fetal, perinatal and neonatal components. 3. Statistical methods: Test performance statistics: a. I might have missed it, but please could the authors clarify the source (and add a reference) of their event rate of 35% for their maternal morbidity primary outcome measure. b. Please could the authors clarify if they will they analyse test performance statistics in their trial group(s)? It appears that no blood test is being taken in the pre-intervention arm, and therefore test performance statistics would be confounded by treatment paradox. However, it would be useful to ascertain whether the test has similar diagnostic performance in this (Irish health service) setting (which might be an important explanatory aspect of the trial), and therefore it would be helpful if the authors could include comment on this e.g. in their data analysis plan. 4. SPIRIT checklist: It appears that the authors have completed the SPIRIT checklist for the protocol upload not for the main BMJ Open manuscript. Could the Editors clarify whether they require the elements in the SPIRIT checklist to be in the main manuscript or whether they are happy for the authors to refer to the attached protocol – I am unclear. If only the BMJ Open manuscript will be published, it would be more useful for the SPIRIT checklist to be completed for the manuscript, as I have not personally checked that all checklist items are present in the manuscript. I note that the CF and PIL (SPIRIT notes pp41-46) do not appear to be attached. 5. Health economic analysis: I have not commented on this (as not my main skillset) but I note with interest that the authors say: 'A systematic literature review will be conducted, the results of which will be used to inform a meta-analysis so as to estimate fetal quality of life outcomes for the estimation of QALYs.' My understanding was that this aspect is particularly lacking in the

	literature and that health economists struggled with fetal QALYs. This statement implies that deriving fetal QALYs will be feasible – do the authors still think that this will be possible? 6. It would be useful if the authors could clarify the nature of any links (or not) to the similarly named PARROT UK study, as this is frequently asked. 7. Minor point: In formal writing, 'PET' should be avoided as an abbreviation for pre-eclampsia. My preference is that the word should be spelled out in full for clarity.

REVIEWER	Elena Ricci Fondazione IRCCS Ca' Granda - Ospedale Maggiore - Policlinico, Milan, Italy
REVIEW RETURNED	14-Jul-2018

GENERAL COMMENTS	I've been invited to review the statistics of this protocol; other aspects of this study are out of my field of expertise, although I think that it deals with an interesting issue and aims to provide a much needed test. The stepped wedge design is adequate for the study question. Clusters are clearly defined; randomization is concealed, start dates at 2-3 month intervals; transition immediate from control to intervention arm. Sample size is comprehensively explained, based on pragmatic consideration and potential for enrollment in the study group. The SS estimate is robust to internal coefficient correlation variation (0.01 to 0.05). There is an unclear step in the third paragraph of SS section (p.32): that is why 8,500 women eligible for the study represent 13% of the combined annual delivery rate, since 13% of 44,000 is about 5,720 and 13% of 90,000 is about 11,700. I gather that this percentage is derived from the number of eligible women in 7 centers over the study duration (18 months): it needs to be explicitly stated. Maybe it's clear for an English native speaker, not for me, among other readers. (first question) As usual, the statistical analysis plan will be finalized prior to the commencement of data analysis. Data analysis is not detailed, as stated, but in the corresponding section main adjustment factors are considered and proposed analysis models are appropriate. Multiple testing will be managed using the Bonferroni correction: this could lower the study power, however repeat enrollment is possible, thus increasing the SS. Missing data >5% will be managed using multiple imputation methods. There's no allowance for loss to follow-up: the authors state that a realistic assumption is no missing data for primary outcome(s). I think that the readers need an explanation of why no enrolled women will be lost to follow-up. Is impossible that a woman delivers her baby in an Obstetric ward other than that where she was enrolled? (second question) Reported analyses are appropriate both for primary and secondary objectives.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Joao Guilherme Bezerra Alves Institution and Country: Instituto de Medicina Integral Prof. Fernando Figueira (IMIP), Brazil Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below Congratulations for the project. I only have few questions.
Background: Second paragraph, third and fourth sentences need references.
References now added

VEGF-A was not previously introduced.
Noted, has now been fully introduced at first use

Health economic assessment was only approached in objectives.
The importance of evaluating both the clinical as well as the health economic impact of the introduction of this test is discussed in the Background to the study. In the Methods and Design section we discuss how health economic data will captured and analysed.

Methods and Design: Study design - first sentence: "...prior to 37 weeks gestation" ; 20-37 weeks ?
Amended to "from 20 weeks and prior to 37 weeks' gestation"

Trial registration is in page 4 and not page 1 (SPIRIT page 34).
Trial Registration details now moved to page 1.

Secondary outcome measure - Quality of life is included as an economic evaluation (economic impact) (?).
The economic evaluation is a secondary analysis which will be informed by the quality of life measures (which are a secondary outcome). The text has been amended to reflect this as follows " Secondary analyses will examine further clinical outcomes (such as mode of delivery, antenatal detection of growth restriction and use of antihypertensive agents) as well as a health economic analysis, of incorporation of placental growth factor testing into routine care".

By the size of the study and possible benefits to the intervention group I suggest a data and safety monitoring boarder (DSMB) to ensure the bioethical principle of beneficence.
Agreed this is important aspect of the trial. The following has been added to the manuscript to reflect this;
"Data Monitoring
To provide protection for study participants an independent data monitoring committee (DMC) has been appointed for this trial. The DMC comprises of 4 members who are not involved with any other aspect of the trial. They include an Obstetrician, a neonatologist, a statistician and a midwife. The DMC met and ratified their charter and have advised that all serious adverse events such as stillbirth/neonatal death or profound maternal morbidity in the Intervention arm of the study be reported to them immediately. The DMC will receive regular updates on the progress of the trial every quarter from the trail management group (TMG). The purpose of these updates is for the DMC to; 1) ensure the quality of data collection 2) ensure that the intervention is being rolled out according to the randomisation plan 3) monitor balance between arms to monitor for potential selection biases and 4) ensure PlGF testing is not overwhelmingly better or worse than no PlGF testing with respect to maternal morbidity with neonatal morbidity. Once 1500 outcomes are available an interim analysis will be conducted and reviewed by the DMC. The interim analysis will report on the co-primary outcomes, follow the same methods as those of the primary analysis, and examine if there is proof beyond reasonable doubt that one particular intervention is definitely indicated or definitely contra-indicated in terms of a net difference of a major endpoint. There will be no formal stopping criteria put in place, but the DMC will be guided by the knowledge that proof beyond reasonable doubt cannot be specified precisely, but a difference of at least three standard deviations in an interim analysis of the primary outcome would be consistent with strong level of evidence. No allowance for this interim analysis has been made in power calculations.
There will be no stopping of the trial for futility as the study will be underpowered to detect small effects".

Discussion: Possible limitations of the study were not described; it is also missing in the abstract.
Discussion now includes potential limitations. Not possible to add to abstract given word count restrictions.

Reviewer: 2
Reviewer Name: Sarah Stock
Institution and Country: University of Edinburgh, UK Please state any competing interests or state 'None declared': None Declared

Please leave your comments for the authors below 1. Overall a considered and clear study protocol for a pragmatic trial 2. It was not clear what the pathway of care is for a woman who does not consent for participation in the trial (presumably usual care)
Yes they will continue to receive usual care. The following sentence has been added to reflect this "Eligible women who are approached but who decline to participate in the trial will also receive usual hospital care, however their data will not be recorded."

3. Is there any risk of contamination through off protocol use of the test in clusters randomised to control - and how will this be assessed/managed?
The following has been added to the manuscript "The Triage© PlGF test platform and consumables necessary to perform testing are only brought to the cluster just at the point of transition to intervention. It is therefore not available at site for use while the site is in the control arm".

4. What evidence base is there for the estimates of the baseline incidence of the primary outcome(s), and the anticipated reduction in this/these with use of PlGF?
The PELICAN study reported a maternal adverse outcome rate in the region of 35%. Manuscript now amended to reflect this "With a sample size of 4000 and using a two-sided type I error rate of 0.025 (to allow for two co-primary outcomes), we determined the power to detect a 7% reduction in maternal morbidity (relative risk reduction of 20%) from 35% to 28% in the intervention i.e. 'active' group (based on a reported rate of adverse maternal outcome in the region of 35% in the PELICAN trial)"

5. I am not qualified to comment on the statistics in detail. 7 clusters seems a small number to detect a big effect, but I will defer to specialist statistical reviewer regarding this. I do note 80% power (with a reduction if the ICC estimates are too conservative), thus a reasonable chance of failing to detect a real effect. Perhaps this could be discussed in discussion.
The following has been added to the discussion section of the study: "There are a number of advantages with the use of stepped wedge design. It allows a phased implementation of the intervention, which is preferable when commencement in all clusters simultaneously would be challenging. As all clusters ultimately receive the intervention, it increases willingness of the clusters to partake in the trial. We acknowledge that seven clusters is a small number of clusters and this is an important limitation of the study. Mostly this is a limitation because it will mean that the findings have questionable generalisability. But, if these clusters are representative then the findings may still be generalizable in part. The other limitation that seven clusters brings about is questionable internal reliability. However, because all of the clusters receive both the intervention and control condition, the clusters serve as their own controls. Not only does this lessen the impact of chance imbalance but it also increases the power of the study (particularly so when the ICC is large, as is the case here). The study does only have in the region of 80% power and should parameters such as the ICC be very different to that which we have assumed, then it is correct that the study might be underpowered. To ensure that this is properly accounted for at the analysis stage, we will report appropriate CIs around all point estimates, so the impact of any impression is properly reported"

Reviewer: 3
Reviewer Name: Professor Lucy Chappell
Institution and Country: King's College London, UK
Please state any competing interests or state 'None declared': I am Co-Chief Investigator on the PARROT-UK study, and provided the PARROT-Ireland co-investigators with a copy of the PARROT-UK protocol, database and trial handbook on their request.

Please leave your comments for the authors below

I acknowledge that this is a study protocol of a trial underway, that has been through external peer-review and ethics committee scrutiny, and therefore there is limited scope to amend at this stage. The following are therefore intended to increase clarity only and have no substantial impact on the trial design.

1. Study design
a. The authors state: 'The PlGF measurement is reported as 12 the absolute value in pg/ml within 30 minutes of sampling.' This would suggest that results reporting beyond 30 mins is a protocol deviation, but I would be surprised if the investigators can achieve this in all sites. Please could the authors clarify in the protocol whether a) this time frame is essential and b) how they would account (statistically) for protocol deviations if the time frame is shown to be longer than this (e.g. for the scenarios where reporting is within 2 hours of sampling, or within 12 hours etc.).
Thank you for highlighting this error. We have changed this to now read "The PlGF measurement is reported as the absolute value in pg/ml within 30 minutes of commencing processing of the sample. All samples taken will be analysed without delay by the researcher after venepuncture has occurred and in accordance with manufacturer's instructions."

b. Please could the authors clarify whether they anticipate any impact of the algorithm for Belfast being different in the control arm and what steps they might need to take to mitigate this (e.g. in their analysis)?
Very insightful comment, have added this to the discussion. "A potential limitation worth noting is the slightly different management algorithm for one cluster, Belfast, in the control arm. The Belfast control arm algorithm is taken directly from the NICE Hypertension in Pregnancy guidelines. All other clusters are using an algorithm taken from the HSE Guidelines for Hypertension in Pregnancy. The two are essentially the same except the HSE algorithm also includes a recommendation for a fetal ultrasound in cases where the participant is <34 weeks gestation. It is not anticipated that the difference in these algorithms should have any bearing on the overall trial results. We will conduct a sensitivity analysis with the Belfast site removed and see if the result remains consistent".

2. Outcomes
a. The authors state 'For maternal morbidity assessment, the fullPIERS score is used with the addition of severe hypertension (Table 3).' The authors should clarify more clearly that they have adapted the fullPIERS composite, in particular by including ICU admission (which is a process outcome that depends on the health service setting, rather than a clinical outcome), and that they are using a different definition of hepatic dysfunction (based on ALT rather than INR) that will result in a much higher proportion of cases identified as having maternal morbidity.
Thank you for this feedback, have amended the manuscript to reflect this adaption of fullPIERS.

Please could the authors clarify these adaptations in the protocol and also confirm that there will be no double counting (i.e. individuals not events will be presented for the composite).
Thank you for the comment. Manuscript amended as follows "The primary aim of the study is to evaluate whether there is a difference in the two composite outcomes before and after exposure to the intervention. There will be no double counting of outcomes, individuals not events will be presented for the composite"

b. The authors say 'For neonatal morbidity assessment, babies are dichotomised into having or not having objectively identified neonatal morbidity by means of a composite neonatal score (Table 4).'
I am aware (from neonatology colleagues and inclusion in my own trials) that a number of these outcomes can be difficult to adjudicate, and/ or have various stagings, where the early stage is often not included in a definition of morbidity but later stage disease would (e.g. IVH and NEC), and/or where definitions have evolved with new international consensus definitions now available (e.g. FGR). Please could the authors clarify a) whether all stages of each disease included will be part of the composite outcome;
Manuscript amended as follows; "For neonatal morbidity assessment, babies are dichotomised into having or not having identified neonatal morbidity by means of a composite neonatal score (Table 4). In order to avoid subjectivity in the diagnosis of morbidity, the majority of components of the neonatal composite score are objective measures; pH < 7.2, positive cultures, admission to NICU. We acknowledge that some subjectivity can arise with staging of disease hence why all stages of each

disease will be captured and will comprise the composite outcome; NEC Stage 1-3, IVH Grade 1-4 and ROP Stage 1-5.

b) which definitions they are using (e.g. for FGR/ customisation of BW centile);
The following has been added as an addendum to Table 4 and Table 6; "Customised birth weight at delivery is calculated using the GROW centile Antenatal detection of Fetal Growth restriction is based on formal ultrasound assessment of fetal biometry using the Hadlock formula.".

c) whether they plan to have an adjudication committee or use local case note review of clinical diagnoses by midwives or doctors only.
The following has been added to clarify this "Neonatal outcomes and morbidity will be captured from local case note review, as documented by the treating neonatologist. In cases where any uncertainty is present, the researcher will discuss the case with the local PI and or the trial clinical fellow and a consensus will be reached".

Depending on the clarification, the authors may wish to modify 'objectively identified' in the phrase above, as many would consider a number of the morbidity measures to have an understandable degree of subjectivity.
Modified to "identified"

However, I am surprised by their inclusion of a single antenatal component: 'Fetal growth restriction identified on antenatal ultrasound', as any cases that are labelled as such but do NOT have any other of the listed perinatal components of the composite might well be a test false positive, that drives clinical intervention, without associated morbidity. It would be useful to see justification of this criterion.
Manuscript amended as follows "Secondary outcomes include each component of the primary outcome reported individually as well as further maternal and neonatal assessments such as mode of delivery and use of antihypertensive agents (Table 5 & 6). Fetal growth restriction, identified on antenatal ultrasound, has been included as a secondary outcome measure of neonatal morbidity. As PlGF correlates well with placental dysfunction, it may be able to differentiate between those babies with pathological growth restriction rather than constitutional growth restriction and hence improve neonatal outcomes".

The authors should clarify that there will be no double counting, particularly as there is a mix of fetal, perinatal and neonatal components.
Thank you for the comment. Manuscript amended as follows "The primary aim of the study is to evaluate whether there is a difference in the two composite outcomes before and after exposure to the intervention. There will be no double counting of outcomes, individuals not events will be presented for the composite"

3. Statistical methods: Test performance statistics:
a. I might have missed it, but please could the authors clarify the source (and add a reference) of their event rate of 35% for their maternal morbidity primary outcome measure.
The observational PELICAN study showed a maternal adverse outcome rate in the region of 35%. Reference now added to reflect this.

b. Please could the authors clarify if they will they analyse test performance statistics in their trial group(s)? It appears that no blood test is being taken in the pre-intervention arm, and therefore test performance statistics would be confounded by treatment paradox. However, it would be useful to ascertain whether the test has similar diagnostic performance in this (Irish health service) setting (which might be an important explanatory aspect of the trial), and therefore it would be helpful if the authors could include comment on this e.g. in their data analysis plan.
As no PlGF testing is being undertaken in the control arm it is not possible for test performance statistics to be performed. This has been added to the discussion as a limitation of the study.

4. SPIRIT checklist: It appears that the authors have completed the SPIRIT checklist for the protocol upload not for the main BMJ Open manuscript. Could the Editors clarify whether they require the elements in the SPIRIT checklist to be in the main manuscript or whether they are happy for the

authors to refer to the attached protocol – I am unclear. If only the BMJ Open manuscript will be published, it would be more useful for the SPIRIT checklist to be completed for the manuscript, as I have not personally checked that all checklist items are present in the manuscript. I note that the CF and PIL (SPIRIT notes pp41-46) do not appear to be attached.
Thanks you for this feedback. I have confirmed this with the Editor and have now updated the Spirit checklist to reflect the submitted manuscript rather than the full trial protocol.

5. Health economic analysis: I have not commented on this (as not my main skillset) but I note with interest that the authors say: 'A systematic literature review will be conducted, the results of which will be used to inform a meta-analysis so as to estimate fetal quality of life outcomes for the estimation of QALYs.' My understanding was that this aspect is particularly lacking in the literature and that health economists struggled with fetal QALYs. This statement implies that deriving fetal QALYs will be feasible – do the authors still think that this will be possible?
Thank you for the comment. The systematic review aims to see what others have used as proxies/ if anything so as help identify values that could be incorporated into the decision analytical model to estimate QALYs for the economic evaluation. Indeed this is a challenging area in economics with not a lot of literature expected but we will combine any we do get a formal way. We have amended the text to reflect this "A systematic literature review will be conducted to identify QOL/utilities (or proxies for same) associated with neonate outcomes which will be incorporated into the decision analytical model to estimate QALYs."

6. It would be useful if the authors could clarify the nature of any links (or not) to the similarly named PARROT UK study, as this is frequently asked.
This is an excellent suggestion. The following paragraph has been added to the discussion "A separate RCT, also entitled "PARROT", has completed recruitment in the United Kingdom since the end of 2017. Although recruiting a similar population of women and using the same PlGF platform, the primary outcome measure for the two RCT's is different, with the UK PARROT trial focusing on time from enrolment to diagnosis/delivery. Both studies are using the same electronic clinical record forms developed by MedSciNet and thus will have a large cross-over of data. The advantage of having these two similar RCT's conducted almost simultaneously is that robust information on the impact of incorporation of PlGF into clinical care will be generated. In addition the potential exists for a collaborative project such as an individual participant data meta-analyses in the future."

7. Minor point: In formal writing, 'PET' should be avoided as an abbreviation for pre-eclampsia. My preference is that the word should be spelled out in full for clarity.
Noted with thanks and manuscript amended

Reviewer: 4
Reviewer Name: Elena Ricci
Institution and Country: Fondazione IRCCS Ca' Granda - Ospedale Maggiore - Policlinico, Milan, Italy
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
I've been invited to review the statistics of this protocol; other aspects of this study are out of my field of expertise, although I think that it deals with an interesting issue and aims to provide a much needed test.
The stepped wedge design is adequate for the study question.
Clusters are clearly defined; randomization is concealed, start dates at 2-3 month intervals; transition immediate from control to intervention arm.
Sample size is comprehensively explained, based on pragmatic consideration and potential for enrollment in the study group.
The SS estimate is robust to internal coefficient correlation variation (0.01 to 0.05).
There is an unclear step in the third paragraph of SS section (p.32): that is why 8,500 women eligible for the study represent 13% of the combined annual delivery rate, since 13% of 44,000 is about 5,720 and 13% of 90,000 is about 11,700. I gather that this percentage is derived from the number of eligible women in 7 centers over the study duration (18 months): it needs to be explicitly stated. Maybe it's clear for an English native speaker, not for me, among other readers. (first question)
Thank you for highlighting this error, the manuscript has been corrected as follows " A recruitment feasibility audit conducted in Cork University Maternity Hospital (CUMH) over the course of a typical week in July 2016 identified 21 women who would be eligible for inclusion in the PARROT Ireland

study. This would equate to almost 1100 women per annum in CUMH, approximately 13% of its annual delivery rate. This is in keeping with the quoted 10% incidence of hypertensive disorders of pregnancy (HDP) in the population (41). It is anticipated that over the 22 month duration of the study across the 7 hospitals approximately 10,486 women will meet the study inclusion criteria (13% of the combined annual delivery rate), and of these 4,000 will be recruited into this trial (approximately 38% of those eligible)"

As usual, the statistical analysis plan will be finalized prior to the commencement of data analysis. Data analysis is not detailed, as stated, but in the corresponding section main adjustment factors are considered and proposed analysis models are appropriate. Multiple testing will be managed using the Bonferroni correction: this could lower the study power, however repeat enrollment is possible, thus increasing the SS. Missing data >5% will be managed using multiple imputation methods.
Comments noted. Even if the primary outcome analyses on maternal morbidity does not show a significant difference in favour of the intervention, the co-primary outcome neonatal morbidity will be analysed for any significant difference as both outcomes are equally important. To adjust for the co-primary endpoints multiple testing, we will use the Bonferroni method; and this has been allowed for in our power calculation. Missing data >5% will be managed using multiple imputation methods.

There's no allowance for loss to follow-up: the authors state that a realistic assumption is no missing data for primary outcome(s). I think that the readers need an explanation of why no enrolled women will be lost to follow-up. Is impossible that a woman delivers her baby in an Obstetric ward other than that where she was enrolled?
(second question) Reported analyses are appropriate both for primary and secondary objectives.

Thank you for the comment. Additional sentences added to further explain "In the PELICAN study only 1% of the cohort were lost to follow up. The risk of incomplete data collection of outcomes in studies such as this is more relevant if women deliver in a different unit to that which they are recruited in to the trial. However, all seven clusters in our trial are large tertiary referral units and patient transfer during pregnancy is rare. We are therefore confident that the likely rate of loss to follow up will be similar and in the order of 1%".

VERSION 2 – REVIEW

REVIEWER	Lucy Chappell King's College London, UK
REVIEW RETURNED	13-Oct-2018

GENERAL COMMENTS	I am happy that the authors have adequately addressed the reviewers' comments, including mine, and do not have further comments for the authors to address.

REVIEWER	Elena Ricci Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
REVIEW RETURNED	09-Oct-2018

GENERAL COMMENTS	My previous questions have been satisfactorily addressed. I highly recommend publication.