PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Increasing use of inhaled nitric oxide in neonatal intensive care units in England: a retrospective population study
AUTHORS	Subhedar, Nimish V Jawad, Sena Oughham , Kayleigh Gale, Chris Battersby, Cheryl

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Dr. anne debeer Institution and Country: Katholieke Universiteit Leuven UZ Leuven, neonatology, Belgium Competing interests: None
REVIEW RETURNED	12-Nov-2020

GENERAL COMMENTS	This article present a nice study on the use of iNO in patients admitted to NICU in the UK over different time epochs. Main question is to why the author divides the patient groups in <29w, 29-33 and>33 w. Internationally and according to WHO definitions <28, 28-32 and >32 would make more sense. Second comment is on the <29w infants where iNO was started after the first week of life, I think it would be worht commenting on possible diagnoses/explanations. The duration of iNO in the different epochs is numerically nearly equal but statistically highly significantly different - please comment. Also the information on duration of iNO in baby's that did not die would be nice to mention.

REVIEWER	Reviewer name: Dr. Michele C. Walsh Institution and Country: Case Western Reserve Univ, United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	29-Oct-2020

GENERAL COMMENTS	Dr. Sedhar and colleagues conducted a retrospective cohort trial to assess temporal changes in inhaled Nitric Oxide (iNO) use in English neonatal units between 2010 and 2015. They compared the rate of use, duration of treatment and postnatal age at treatment in three epochs- 2010-2011, vs 2012-2013, 2014-2015. Further they analyzed these factors between extremely low gestational age infants (EPT, < 29 weeks), moderate preterm infants (MPT 29-33 weeks), and late preterm/term infants (Term, ≥34 weeks. While iNO use has been shown to be beneficial above 34 weeks gestation, there is little evidence of efficacy, and perhaps evidence of harm especially in EPT in the first week of life. There has been concern of “therapeutic creep” with increasing use among infants <34 weeks gestation. Results: The groups were comparable in demographics across the three epochs. The use of iNO in infants ≥ 34 weeks was relatively

	stable across Epoch 1 and Epoch 3 (4.5% vs 5.0%). However, the use in both EPT (4.9 vs 15.9%) and MPT (1.1 vs 4.9%) increased significantly. Fully 31% of those receiving iNO were EPT across the study- an astonishing number for an expensive and unproven therapy. As has been shown with nearly every neonatal therapeutic agent examined, there was wide variation in use between different neonatal units- even after the authors eliminated units that had < 5 cases of iNO use. Use was fairly consistent among Term infants, but much more variation was seen in EPT. The manuscript will be of interest both to neonatologists but also to health policy makers. Only minor revisions are needed. 1. Abstract and Text Conclusion: Both could be stronger than more study is needed. Given expensive, limited evidence of efficacy in EPT and potential for harm, your data suggest that iNO be restricted to clinical trials rather than off label use. Policy makers could decline to pay for it outside of trials, and use may be reduced. 2. Introduction, Methods, Analytic Plan: All are strong. Written clearly. Please clarify in methods and tables what the denominator is for key measures: for example Table 1. Last Row Deaths. Appears to be deaths among those treated with iNO. Please state clearly. As other potential denominators could be all infants in that gestational age class. 3. Human Subject Protections: Reviewed by West of Scotland Ethics Review Board and approved. Say more clearly whether was approved as a QI project with permission given to use anonymized data and a waiver of patient level consent. Or other type of approval. 4. Results: Generally clear. The authors are careful to not over reach. Tables can be improved (see #2 above). 5. Discussion: Page 18: “The reason for increasing use of iNO off-label in preterm infants is not known but is likely to be multifaceted and reflect the absence of other proven ‘rescue’ cardiorespiratory interventions for infants with severe hypoxaemic respiratory failure (such as ECMO) in this population, growing experience in the use of iNO, the absence of evidence of short-term harm, and that off-label use of iNO is fully reimbursed in England”. The Van Meurs Premie iNO study showed increased mortality in the group treated with iNO- therefore there is some evidence of harm that would be difficult to detect at the individual unit level. Thus the statement of absence of evidence of harm should be revised. In addition, specific mention of the increased mortality should be discussed. Additional Reference needed for oligohydramnios and possible efficacy. Suggest add: “Chock VY, Van Meurs KP, Hintz SR, Ehrenkranz RA, Lemons JA, Kendrick DE, Stevenson DK; NICHD Neonatal Research Network. Inhaled nitric oxide for preterm premature rupture of membranes, oligohydramnios, and pulmonary hypoplasia. Am J Perinatol. 2009 Apr;26(4):317-22. doi: 10.1055/s-0028-1104743. Epub 2008 Dec 9. PMID: 19067285; PMCID: PMC2676224.” Also add statement that it is unlikely that an RCT could be feasible in this subset of premature infants due to the enormous sample size required. 6. Conclusion: As in #1 above: a more firm conclusion is warranted.

REVIEWER	Reviewer name: Dr. Andrei Harabor

	Institution and Country: University of Saskatchewan, Pediatrics, Canada Competing interests: None
REVIEW RETURNED	28-Oct-2020

GENERAL COMMENTS	This is overall an excellent and very well written paper. Of note, the Canadian cohort of infants < 29 weeks gestational age described by Soraisham et al is very similar to the one described in this paper and it is not mentioned. The iNO use was high but stable across the 4 years of that cohort, 2010-2013, as opposed to the very large increase in use of iNO in the UK - essentially doubling for each time interval. This is a contrasting finding worth emphasizing. The other finding was that later start was associated with longer duration of therapy, possibly for prevention / treatment of BPD / associated PPHN. This does not seem to be the case in the present cohort. For the later start of iNO - is the starting time a multimodal distribution with altered proportion between peaks ? If not, could that have something to do with a trend of later intubation, initial time spent on CPAP or non-invasive ventilation ? Would the dataset be able to provide a hint ? The discussion seems to emphasize the increased use in the <29 weeks GA but the proportional increase is the same in the 29-33 weeks GA group, together with a decrease in diagnoses most commonly associated with possible iNO effect.

REVIEWER	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: None
REVIEW RETURNED	28-Oct-2020

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. Unfortunately, I have some fairly major concerns. However, these are remediable, so I am asking for a revision rather than recommending rejection. There are two major problems: First, the researchers have the whole population. Therefore, I think any use of inference (or of tests, or of significance) is unnecessary. You have no inference to do. Some statisticians think that you can posit a "super population" from which this is a random sample. This isn't completely unreasonable, but I don't think it really works. Even if you can define such a super-population, there is no reason to think that you have a random sample from it. (E.g. the super population might be "all preterm babies on Earth" or "preterm babies in England over a greater time span" but then your data are clearly not a random sample.) So .... just do description. The second problem is the model. Even though I don't think you want to do statistical tests, you can still model the data you have. But, rather than stratify by gestational age, it would be better to analyze all the data at once and add GA in weeks as a variable and then use a spline of that in the model. (Categorizing a

	continuous variable is nearly always a bad idea, see Harrell *Regression Modeling Strategies). Similarly, year of study should not be categorized but left continuous. The other problem with the model is whether you need to account for the different units in the analysis at the individual level. I think you do, as there could be variation. One way to do this is to use a multilevel model. Sorry to be so critical, but, as I said, I think these problems are remediable. Peter Flom

VERSION 1 – AUTHOR RESPONSE

Responses to reviewers' comments

Reviewer: 1

I confine my remarks to statistical aspects of this paper. Unfortunately, I have some fairly major concerns. However, these are remediable, so I am asking for a revision rather than recommending rejection. There are two major problems:

Comment 1 : First, the researchers have the whole population. Therefore, I think any use of inference (or of tests, or of significance) is unnecessary. You have no inference to do. Some statisticians think that you can posit a "super population" from which this is a random sample. This isn't completely unreasonable, but I don't think it really works. Even if you can define such a super-population, there is no reason to think that you have a random sample from it. (E.g. the super population might be "all preterm babies on Earth" or "preterm babies in England over a greater time span" but then your data are clearly not a random sample.) So .... just do description.

Response: We agree with the reviewer's comments and apologies if this was not clear in the paper, but the intention was always to be a descriptive paper. Hence the aims in the introduction read "We aimed to describe temporal changes in the use of iNO in neonates admitted to neonatal units in England using national data routinely recorded during clinical care and held in the National Neonatal Research database (NNRD). Our objectives were to i) describe the proportion and characteristics of preterm and term infants who receive iNO between 2010 and 2015; ii) determine whether there is variation in iNO use across tertiary level neonatal units, and over time between 2010 and 2015."

The discussion also contains a sentence that reads "This study was not designed to analyse changes in outcomes beyond simple descriptive data."

To add more clarity that this is a descriptive study, under "Statistical analyses" in the "Methods" section, we have changed "We conducted analyses for the primary outcome…." to "We describe the cohort at two levels…".

Comment 2: The second problem is the model. Even though I don't think you want to do statistical tests, you can still model the data you have. But, rather than stratify by gestational age, it would be better to analyze all the data at once and add GA in weeks as a variable and then use a spline of that in the model. (Categorizing a continuous variable is nearly always a bad idea, see Harrell *Regression Modeling Strategies). Similarly, year of study should not be categorized but left continuous.

Response: We agree with the first comment of reviewer 1, which is to just describe the cohort. Therefore, we do not think modelling the data would be necessary. We have presented the results in clinically meaningful categories. The categories were chosen because the license for Nitric oxide is for

infants >=34 weeks; we then split the other 2 categories because other comparable studies have used similar thresholds, albeit all slightly different to each other.

Comment 3: The other problem with the model is whether you need to account for the different units in the analysis at the individual level. I think you do, as there could be variation. One way to do this is to use a multilevel model.

Please see response to comment 2: We did not set out to model any data, but to merely describe the cohort. Allocating nitric oxide use to one single unit is difficult as many babies receive nitric oxide in more than one unit due to transfers.

Reviewer: 2

This is overall an excellent and very well written paper.

Comment 1: Of note, the Canadian cohort of infants < 29 weeks gestational age described by Soraisham et al is very similar to the one described in this paper and it is not mentioned. The iNO use was high but stable across the 4 years of that cohort, 2010-2013, as opposed to the very large increase in use of iNO in the UK - essentially doubling for each time interval. This is a contrasting finding worth emphasizing. The other finding was that later start was associated with longer duration of therapy, possibly for prevention / treatment of BPD / associated PPHN. This does not seem to be the case in the present cohort.

Response: We thank reviewer 2 for the positive comments. Thank you for highlighting this omission of reference. We have added to the third paragraph of the following discussion and reference:

"The Canadian Neonatal network (CNN) found similar rates (1 in 25; 4.2%) of iNO use among infants born <34 weeks between 2010 and 2013. As different gestational age categories were used, direct comparisons cannot be made, but the use of iNO was broadly similar to the recent UK figures; 16.1%, 6.0%, and 1.3%, for babies born 22 to 25, 26 to 29, and 30 to 33 weeks, respectively. However, in contrast to the increasing use in the UK, iNO use was stable across the 4 years in the CNN."

We note the contrasting results, but as we did not set out to investigate or test the association between initiation and duration of iNO use, we not included this.

Comment 2:
For the later start of iNO - is the starting time a multimodal distribution with altered proportion between peaks ?If not, could that have something to do with a trend of later intubation, initial time spent on CPAP or non-invasive ventilation ? Would the dataset be able to provide a hint ?

Response: This was outside the scope of the study and we had not planned to look at this; but we thank the reviewer for their useful suggestion for future work as an extension to this descriptive study.

Comment 3:
The discussion seems to emphasize the increased use in the <29 weeks GA but the proportional increase is the same in the 29-33 weeks GA group, together with a decrease in diagnoses most commonly associated with possible iNO effect.

Response: We have added this to the first paragraph of the results "There was a similar 3-4 fold increase in rates of iNO use for infants born <29 weeks and 29-33 weeks, from 4.9 to 15.9% and 1.1 to 4.8% for infants, respectively."

Reviewer: 3
The manuscript will be of interest both to neonatologists but also to health policy makers. Only minor revisions are needed.

Comment 1. Abstract and Text Conclusion: Both could be stronger than more study is needed. Given expensive, limited evidence of efficacy in EPT and potential for harm, your data suggest that iNO be restricted to clinical trials rather than off label use. Policy makers could decline to pay for it outside of trials, and use may be reduced.

Response: Thank you for this comment. We have amended the abstract conclusion to read "Given the cost of iNO therapy, limited evidence of efficacy in preterm infants, and potential for harm, we suggest that exposure to iNO should be limited, ideally to infants included in research studies (either observational or RCT) or within a protocolised pathway that permits a short trial of iNO to assess acute oxygenation response. Development of consensus guidelines might also help to standardise practice."

Comment 2. Introduction, Methods, Analytic Plan: All are strong. Written clearly. Please clarify in methods and tables what the denominator is for key measures: for example Table 1. Last Row Deaths. Appears to be deaths among those treated with iNO. Please state clearly. As other potential denominators could be all infants in that gestational age class.

Response: Agree we should be clearer on the denominator. For tables 1-3, we have
i) added to the Table title "treated with iNO"
ii) To the footnotes below the table, we have added "The denominator for all proportions is the number of babies treated with iNO unless indicated otherwise'
iii) Added the symbol † to the second row of the table where the denominator is all admissions to neonatal unit admissions requiring ≥ 1 day of intensive care.
This is referenced in the footnote below the table.
Iv) The last row of the table has been amended for clarity and reads "Death among infants who received iNO"

Comment 3. Human Subject Protections: Reviewed by West of Scotland Ethics Review Board and approved. Say more clearly whether was approved as a QI project with permission given to use anonymized data and a waiver of patient level consent. Or other type of approval.

Response: We have added the following to the sentence ". This study using anonymised data was approved by the West of Scotland Research Ethics Committee 5; reference number 16/WS/0228"

Comment 4: Results: Generally clear. The authors are careful to not over reach. Tables can be improved (see #2 above).

Response: Amendments made to tables as describe in response to comment 2

Comment 5: Page 18: "The reason for increasing use of iNO off-label in
preterm infants is not known but is likely to be multifaceted and reflect the absence of other proven 'rescue' cardiorespiratory interventions for infants with severe hypoxaemic respiratory failure (such as ECMO) in this population, growing experience in the use of iNO, the absence of evidence of short-term harm, and that off-label use of iNO is fully reimbursed in England".

The Van Meurs Premie iNO study showed increased mortality in the group treated with iNO- therefore there is some evidence of harm that would be difficult to detect at the individual unit level. Thus the statement of absence of evidence of harm should be revised. In addition, specific mention of the increased mortality should be discussed. Additional Reference needed for oligohydramnios and possible efficacy.

Response: Thank you. We have removed "absence of evidence of short-term harm" to the paragraph in the discussion.

We have added the following to the discussion "Post-hoc analyses from a study which randomised 420 neonates born <34 weeks gestation to placebo or iNO found an apparent increase in mortality and

higher rate of intraventricular haemorrhage in infants with a birth weight ≤1000g". ; followed by the reference Van Meurs study.

This reference for Chock et al has been added to the discussion.

Also add statement that it is unlikely that an RCT could be feasible in this subset of premature infants due to the enormous sample size required.

Response: We have added this to the last paragraph of the discussion
"For an RCT to be adequately powered, international collaboration is necessary to achieve the large sample sizes necessary. Such studies will help better understand the long-term impact, identify infants most likely to benefit and inform international guidance."

Comment 6. Conclusion: As in #1 above: a more firm conclusion is warranted.

Please see response to Reviewer 3 comment 1
We have added this to the conclusion of the abstract and discussion.
read "Given the cost of iNO therapy, limited evidence of efficacy in preterm infants, and potential for harm, we suggest that exposure to iNO should be limited, ideally to infants included in research studies (either observational or RCT) or within a protocolised pathway that permits a short trial of iNO to assess acute oxygenation response. Development of consensus guidelines might also help to standardise practice."

Reviewer: 4

This article present a nice study on the use of iNO in patients admitted to NICU in the UK over different time epochs.

Main question is to why the author divides the patient groups in <29w, 29-33 and>33 w. Internationally and according to WHO definitions <28, 28-32 and >32 would make more sense.

Please see also response to reviewer 1 comment 2. The gestational age categories were chosen because the license for Nitric oxide is for infants >=34 weeks; we then split the other 2 categories because other comparable studies have used similar thresholds, albeit all slightly different to each other.

Second comment is on the <29w infants where iNO was started after the first week of life, I think it would be worth commenting on possible diagnoses/explanations. The duration of iNO in the different epochs is numerically nearly equal but statistically highly significantly different - please comment. Also the information on duration of iNO in baby's that did not die would be nice to mention.

This was outside the scope of the study and we had not planned to look at this; but we thank the reviewer for their useful suggestion for future work as an extension to this descriptive study.

Associate Editor
Comments to the Author: we sincerely thank you for this interesting paper. The different reviewers have carefully commented the paper, and provided complementary remarks. Based on this, the paper needs revision.

Some additional general remarks:
The authors refer to available data in US hospitals. It might be interesting to expand on this, and to mention available data. The same for the following sentences on data from multicenter studies.

Please see response to reviewer 2 comment 1.
We have added the following to the discussion:
"The Canadian Neonatal network (CNN) found similar rates (1 in 25; 4.2%) of iNO use among infants born <34 weeks between 2010 and 2013. As different gestational age categories were used, direct

comparisons cannot be made, but the use of iNO was broadly similar to the recent UK figures; 16.1%, 6.0%, and 1.3%, for babies born 22 to 25, 26 to 29, and 30 to 33 weeks, respectively. However, in contrast to the increasing use in the UK, iNO use was stable across the 4 years in the CNN. (11)."

The introduction may overall contain some more background information.

The use of iNO is highest in the <29 weeks group. I would expect some more reflection on this in the discussion compared to what is currently provided.

Response:
Thank you. A paragraph in the discussion speculates the reason for increasing use in preterms, but we have added the underlined phrases to the discussion.

"The reason for increasing use of iNO off-label, particularly in the most preterm infants is not known but is likely to be multifaceted and reflect the absence of other proven 'rescue' cardiorespiratory interventions for these smaller infants with severe hypoxaemic respiratory failure (such as extracorporeal membrane oxygenation) and full reimbursement of off-label iNO use in England. Furthermore, there is some limited evidence for the use of iNO in specific groups of preterm infants including those born following preterm prolonged rupture of membranes and those with echocardiographic criteria of PPHN physiology, supported by expert opinion and consensus statements (15-18). There is growing experience in the use of iNO and the immediate short-term oxygenation response can be gratifying for clinicians, and may encourage further use. However, whether the short-term benefit in oxygenation is translated into longer-term benefit in preterm infants is unknown and needs further investigation. Moreover, the perception of absence of harm should not be extrapolated from term infants simply because there is a lack of convincing evidence of harm in preterm infants."

The figures are quite small. Can these be presented larger or in a different visual form to improve clarity for the reader?

We have attached a figure with larger fonts to make it clearer visually

Editor in Chief
"Use of iNO increased significantly over time from 3.4% (1,293/37,885) in 2010-2011 to 6.4% (3,112/48,838) in 2014-2015." This statement should also be in the abstract and the % are worth mentioning in What this study adds.

Thank you for the suggestions. We have added the suggested to "What this study adds"

VERSION 2 – REVIEW

REVIEWER	Reviewer name: Dr. anne debeer Institution and Country: Katholieke Universiteit Leuven UZ Leuven, neonatology, Belgium Competing interests: None
REVIEW RETURNED	21-Dec-2020

GENERAL COMMENTS	The authors responded well to the comments of the reviewers and for me the article is acceptable in its current form.

REVIEWER	Reviewer name: Dr. Andrei Harabor Institution and Country: University of Saskatchewan, Pediatrics, Canada Competing interests: None
REVIEW RETURNED	09-Dec-2020

GENERAL COMMENTS	Very last line needs to be deleted, it is a repetition from 5 lines above.

REVIEWER	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting, United States Competing interests: None
REVIEW RETURNED	13-Dec-2020

GENERAL COMMENTS	I am confused. In their response to my earlier comments which suggested not doing any testing or modeling, the authors agreed and said that their intent was to do a descriptive analysis. But the new version still has models and tests.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Very last line needs to be deleted, it is a repetition from 5 lines above.

Apologies, and thank you for spotting this repetition. This has been deleted

Reviewer: 2

I am confused. In their response to my earlier comments which suggested not doing any testing or modeling, the authors agreed and said that their intent was to do a descriptive analysis.

But the new version still has models and tests.

Apologies for this. We have removed all tests of significance and p values from the tables.

Reviewer: 3

The authors responded well to the comments of the reviewers and for me the article is acceptable in its current form

Editor in Chief
Comments to the Author:
The statistician has major concerns re your paper. I asked him for specific advice which is reproduced below "My recommendation would be to remove tests of significance (p values, CIs and so on) and to user a spline (e.g. a restricted cubic spline, but other splines are also fine) of age to examine nonlinearity."

I have highlighted in red the sections that MUST be removed re tests of significance.
We are prepared to let you keep your current gestational age division, as it is in keeping with other papers.
I have also highlighted one sentence in the Discussion to remove -re "the only study to " as this does not comply with our instructions to authors

Thank you very much for your suggestions and clear instructions. We have removed the red sections that have tests of significance, and deleted the sentence in discussion as requested.