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.

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 Paediatrics Open. The paper was subsequently accepted for publication at BMJ Paediatrics Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Changes in survival and neurodevelopmental outcomes of infants born at <25 weeks' gestation: a retrospective observational study in tertiary centers in Japan
AUTHORS	Kono, Yumi; Yonemoto, Naohiro; Nakanishi, Hidehiko; Kusuda, Satoshi; Fujimura, Masanori

VERSION 1 - REVIEW

REVIEWER	Marlow, Neil University College London, Institute for Womens Health Competing interests: None
REVIEW RETURNED	05-Sep-2017

GENERAL COMMENTS	This is a national study conducted retrospectively from a multicentre database in Japan. They are important data and worth publication but I have several concerns over the manuscript as submitted: 1. Title – I think this needs to reflect the biases in the ascertainment and denominator – these are changes in survival and outcome in Tertiary centres in Japan. 2. Abstract – the phrase at the threshold of viability is redundant – they are births <25w. 3. Introduction – the first two sentences are obscure – please clarify. The phrase threshold of viability is not useful and what is "societal agreement"? The English needs some attention. 4. Methods – the principle problem is that of denominator. The only hard point the group appear to have is admission for NIC. They have then excluded low throughput units, but such units may be encouraged to carry out active care after publications such as these, and it is important to know what these outcomes are like as well. UK experience suggests these will have much worse outcomes. Transfers similarly are an important but excluded group. These are whole population data minus the more challenging ones! It may be that the numbers are very small but without reporting them the size of this bias cannot be ascertained. Are there national data that could provide such a denominator for births such as been done for deaths? I accept that common practice is to be active but it needs to be confirmed in the first epoch with more certainty than is described. 5. By excluding the smaller centres (62% of births) how has this biased their data? 6. A further problem is they claim that deaths in the delivery room were included in their definitions of death but data were only collected after 2006 and after the first study period (2003-7). DR deaths could not be included therefore so the denominator for the comparison must surely be admission for NIC. Using just 2 years of

	the first period begs the question why look at the period before 2006 (and it is unclear if units started recording data during 2006 when full recording was achieved). How can we interpret table 3 survival data then? 7. The figure shows a group “follow up without data” I presume they meant not followed up or something similar, 8. How to report multiple imputations is challenging and a matter of style I suspect. Personally, I feel the main body of the paper needs to show clearly the follow up rates – which includes 39% full evaluation and 42% incomplete evaluations. I would promote Supp Table 2 to the main manuscript as this is the table of results. A statement should then indicate any changes following imputation. For example, the rates of developmental delay are higher in the non-KSPD tested patients – how do the authors account for this? 9. Were there differences between infants that had a full assessment and those that did not? Were there systemic differences in the rates of GMFCS grades for example? 10. Was imputation the correct technique? Experience in other studies suggests that although the clinical details suggest missingness is random, there are major biases from social and demographic factors. Furthermore, are the authors suggesting they imputed for each discreet impairment? This seems overkill and all of the data for survivors in table 3 has been manipulated, making it really difficult to know what was found unless Supp Table 2 is in the main text. A single imputation of overall impairment rates would be a better, informative and more conservative approach, and more clearly define the biases in ascertainment. 11. The report on rates of CP is misleading – it is the rate of CP with GMFCS 2 or greater. 12. The use of <70 is interesting as the SD of the KSPD is 13 making -2sd 74 and -3sd 61 – is there a reason for this? 13. Was it not possible to grade the overall impairment as severe (DQ <-3SD; GMFCS 3-5) such a distinction may be better if the data are to be used for defining action. 14. Table 4 is slightly challenging as most of the difference in sensory outcomes is driven by death. The large number of visual impairments I suspect is due to the inclusion of amblyopia as an impairment and not by visual function, which is more important – I would not alter my care strategy based on a risk of amblyopia. Please make these impairments clearer. The authors have removed children with CP GMFCS grade 1, but added in this large group.

REVIEWER	Doyle, Lex Royal Women's Hospital Competing interests; I have no competing interests to declare
REVIEW RETURNED	11-Sep-2017

GENERAL COMMENTS	The study reports the changes in survival and neurodevelopmental impairment rates over a decade (2003-2012) of infants born between 22-24 weeks’ gestation and cared for among the Neonatal Research Network of Japan. The authors conclude that survival rates have increased and impairment rates at 3 years of age have decreased in the second 5-year period compared with first 5-year period. Major comments 1. As survival is a major outcome, and because all other outcomes are expressed linked with survival, to interpret the data correctly it is vital that the denominators reflect the population served by hospitals

	within the NRN. Since the denominator is those cared for within the NRN network, and then further restricted to those units contributing at least 10 cases in both eras, there is no idea how many infants were born 22-24 weeks' gestation within the geographic areas served by the hospitals within the NRN. Have there been changes in referral patterns from maternity hospitals to hospitals within the NRN over time? 2. The authors acknowledge that suboptimal follow-up rate of 62%, which raises serious doubt as to the validity of the rates of the various impairments. They then use multiple imputation to estimate rates of these impairments. However, in Table 3 there is no alteration to the numerators of most of the outcomes reported relative to the numerators in supplementary table 2. Surely the numerators should be higher in Table 3, after multiple imputation? 3. The multiple imputation methods are not described. Moreover, the variables included in the models comprise only events known up to the time of birth. As the aim of the imputation is to provide the best estimate of the outcome at 3 years of age for children not assessed, why are the events occurring after birth, particularly those more likely to be related to long-term outcome, such as major cerebral injury and postnatal steroids, not included in the models? 4. The major results all include death as part of the outcome, particularly in the abstract – the impairment outcomes excluding death should also be emphasised in the abstract and major conclusions, and in Table 4. Minor comments • Who performed the neurodevelopmental assessments at 3 years of age and what was the level of agreement among assessors for the various outcomes described? • Why was age not corrected for prematurity? • Why were children with CP at GMFCS level 1 excluded? These are the most numerous group among preterm children with any CP • Could the rates of bilateral blindness also be reported to be able to compare with other studies? • It is stated that psychologists were blinded to perinatal details – did they know the participants were all extremely preterm, or were they also assessing term-born children? • In tables 1 and 2 more data, such as mean differences or odds ratios and 95% CIs should be provided; not just the p-values. • Have any of the children remained as inpatients from the time of birth until 3 years of age? • In supplemental table 1, the rate of males in the without FU group should be 51%, not 56% • In supplemental table 2, "blocks" are misspelled as "brocks" • The study protocol was approved by the Ethics Review Committees of Tokyo Women's Medical University and Jichi Medical University – did families give informed consent to participate?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1	Comments	Response *Answers to reviewer's questions*	Related pages and line in the revised manuscripts
1	Title – I think this needs to reflect the biases in the ascertainment and denominator – these are changes in survival and outcome in Tertiary centers in Japan.	The title was changed to "Changes in survival and neurodevelopmental outcomes of infants born at <25 weeks' gestation in tertiary centers in Japan".	Title
2	Abstract – the phrase at the threshold of viability is redundant – they are births <25w.	Changed to "born at <25 weeks' gestation"	P2, L2
3	Introduction – the first two sentences are obscure – please clarify. The phrase threshold of viability is not useful and what is "societal agreement"? The English needs some attention.	The first two sentences were changed to "The mortality rates of extremely preterm infants born at a gestational age (GA) of 22 and 23 weeks are high, and those who survive often have neurological and developmental impairments". The sentence on "societal agreement" was removed.	P4, L2-4

4-1	Methods – the principle problem is that of denominator. The only hard point the group appear to have is admission for NIC.	denominator: infants born at 22 weeks 0 days to 24 weeks 6 days between 2003 and 2012 who were born at 52 tertiary centers in Japan and admitted to the NICU of the same centers. We added to the Introduction that this was a study on infants born and cared for at 52 tertiary centers for comparisons between the first and the second periods. In the Methods, the reasons for selecting the 52 centers were added; the numbers of participating centers increased from 38 centers to 103 centers during the study period. In order to compare the outcomes of infants from the same tertiary centers between the two periods, 2003-2007 (period 1) and 2008-2012 (period 2), centers that registered less than 10 infants during period 1 or 2 were excluded from this study. Coverage for the national population was described in the Discussion and Supplemental Table 1. These subjects accounted for 34% in period 1 and 42% in period 2 of all infants born at 22-24 weeks' gestation in the nationwide population.	P4, L16-18 P5, L1-5 P13, L4-6 Supplemental Table 1
4-2	They have then excluded low throughput units, but such units may be encouraged to carry out active care after publications such as these, and it is important to know what these outcomes are like as well. UK experience suggests these will have much worse outcomes.	Same as 4-1 The mortality rates of infants at centers not selected were 163/178 (35.4%) in period 1 and 182/495 (26.9%) in period 2, and were slightly higher. Since the aim of the present study was not to evaluate between-center variations, we added the following to the Discussion: Further studies are needed in order to investigate between-center variations in the outcomes of these extremely preterm infants.	P13, L6-7

4-3	Transfers similarly are an important but excluded group. These are whole population data minus the more challenging ones! It may be that the numbers are very small but without reporting them the size of this bias cannot be ascertained.	In the Methods, we added the reasons for excluding transferred infants and one reference as follows: “Infants transferred to centers after birth were also excluded in order to reduce selection bias if only infants in good condition [11] or those who needed specific treatments, such as PDA ligation, were transferred after birth.” The mortality rates at the NICU discharge of transferred infants were 50/162 (30.9%) in period 1 and 39/169 (23.1%) in period 2, and these were slightly lower than those of non-transferred infants (not added to the text).	P5, L6-8 Reference #11
4-4	I accept that common practice is to be active but it needs to be confirmed in the first epoch with more certainty than is described.	Since deaths in the delivery room were not registered between 2003 and 2005, we added the annual numbers of deaths in the delivery room, deaths in the NICU, and active treatment to Supplement Table 1. There were no significant changes in the proportions of active treatment throughout the study period.	Supplement Table 1 P8, L12-14
5	By excluding the smaller centers (62% of births) how has this biased their data?	answered in 4-2.	
6	A further problem is they claim that deaths in the delivery room were included in their definitions of death but data were only collected after 2006 and after the first study period (2003-7). DR deaths could not be included therefore so the denominator for the comparison must surely be admission for NIC. Using just 2 years of the first period begs the question why look at the period before 2006 (and it is unclear if units started recording data during 2006 when full recording was achieved). How can we interpret table 3 survival data then?	As described in 4-4, the annual numbers of deaths in the delivery room, deaths in the NICU, and active treatment were provided in Supplement Table 1. In the Discussion, we added the following sentences: “Although the number of deaths in the delivery room was small, if deaths in the delivery room were accounted for from 2003, the mortality rate in period 1 may have been slightly higher and improvements in period 2 may have been greater. When excluding infants who died in the delivery room from the study subjects, the proportions of death in the NICU were 471/1462 (32.2%) in period 1 and 417/1794 (23.2%) in period 2.”	Supplement Table 1 P13, L10-15
7	The figure shows a group “follow up without data” I presume they meant not followed up or something	It means not followed or followed without the registration of data. We changed “without data” to “no follow-up data” (Figure 1).	Figure 1

	similar,		
8	How to report multiple imputations is challenging and a matter of style I suspect. Personally, I feel the main body of the paper needs to show clearly the follow up rates – which includes 39% full evaluation and 42% incomplete evaluations. I would promote Supp Table 2 to the main manuscript as this is the table of results. A statement should then indicate any changes following imputation. For example, the rates of developmental delay are higher in the non-KSPD tested patients – how do the authors account for this?	We moved original supplement Table 2 into main text Table 3 as a table of results. The follow-up rates were shown in Table 3 and in the Results as "follow-up data, with assessments at 3 years of age being collected for 1463 infants: 631 (64.1% of survivors) in period 1 and 832 (60.7% of survivors) in period 2, which includes 39% full evaluations and 42% incomplete evaluations." The reason for developmental delays being more common in non-KSPD tested patients was additionally described in the Results as follows: "Among the 175 children not tested by KSPD, 36 (20.6%) had CP and 26 (14.9%) had VI. In contrast, among the 1054 children tested by KSPD, 67 (6.3%) had CP and 59 (5.6%) had VI."	Table 3 P9, L4-5 P9, L15-17
9	Were there differences between infants that had a full assessment and those that did not? Were there systemic differences in the rates of GMFCS grades for example?	Based on 8, we assumed that infants with an incomplete assessment had more severe delays or other handicaps. However, no significant changes were noted in NDI in infants with full evaluations or in those with incomplete evaluations between the two periods. We added this sentence in the Results.	P9, L17-19

10	Was imputation the correct technique? Experience in other studies suggests that although the clinical details suggest missingness is random, there are major biases from social and demographic factors. Furthermore, are the authors suggesting they imputed for each discreet impairment? This seems overkill and all of the data for survivors in table 3 has been manipulated, making it really difficult to know what was found unless Supp Table 2 is in the main text. A single imputation of overall impairment rates would be a better, informative and more conservative approach, and more clearly define the biases in ascertainment.	Since our social and demographic factors were limited, and there were no significant differences in characteristics or morbidities between with and without follow-up data (Supplemental Table 2), we used multiple imputations at random to compare between the two periods. With suggestions by the reviewers, we performed single imputations and added the following to the Methods: "As a sensitivity analysis, we performed single imputations in one scenario in which missing data were imputed as having impairments or another scenario in which missing data were imputed as having no impairments." We added the following to the Discussion as limitations: "Although clinical details suggest that missing data were random, there may be biases from social, economic, and other unknown factors. Although there was no significant difference in prenatal and neonatal variables between with and without follow-up data, based on comparisons of results between Tables 3 and 4, the proportion of each impairment decreased after multiple imputations in both periods. The absolute percentage of impairments needs to be interpreted carefully. Therefore, we performed two single imputations for NDI. Significance of changes in outcomes may differ if all missing data were from infants having impairments, as shown in Supplemental Table 4."	Supplemental Table 2 P8, L1-4 P12, L16- L23 Supplemental Table 4
11	The report on rates of CP is misleading – it is the rate of CP with GMFCS 2 or greater.	Rewritten to "the rate of CP with GMFCS □2".	throughout

12	The use of <70 is interesting as the SD of the KSPD is 13 making -2sd 74 and -3sd 61 – is there a reason for this?	The developmental quotient (DQ) is calculated by dividing developmental age by chronological age and then multiplying the quotient by 100, and it does not have a normal distribution. We provided a more detailed explanation as follows and added one reference to the Methods: “A DQ score of KSPD <70, which represents a 70% achievement of standardized performance for the chronological age, was interpreted as significantly delayed according to the protocol by the Japan Neonatal Follow-up Study Group.[16]”.	P6, L17-20 Reference #16
13	Was it not possible to grade the overall impairment as severe (DQ <-3SD; GMFCS 3-5) such a distinction may be better if the data are to be used for defining action.	Since we did not have a definition to grade the severity of impairments in the NRNJ database, we considered it to be inaccurate if we performed grading.	No change
14	Table 4 is slightly challenging as most of the difference in sensory outcomes is driven by death. The large number of visual impairments I suspect is due to the inclusion of amblyopia as an impairment and not by visual function, which is more important – I would not alter my care strategy based on a risk of amblyopia. Please make these impairments clearer. The authors have removed children with CP GMFCS grade 1, but added in this large group.	We included bilateral amblyopia because fictional vison is impaired as much as lateral blindness. The definition of bilateral amblyopia was added to the Methods and Table 3. The number of infants with blindness was also added to Table 3. Children with GMFCS Level 1 were excluded because they were able to walk at 3 years of age and functional impairments were small. We rewrote “rate of CP” as “rate of CP with GMFCS level 2 or higher”.	P6, L11-13, Table 3 P6, L9
Reviewer: 2			

15	As survival is a major outcome, and because all other outcomes are expressed linked with survival, to interpret the data correctly it is vital that the denominators reflect the population served by hospitals within the NRN. Since the denominator is those cared for within the NRN network, and then further restricted to those units contributing at least 10 cases in both eras, there is no idea how many infants were born 22-24 weeks' gestation within the geographic areas served by the hospitals within the NRN. Have there been changes in referral patterns from maternity hospitals to hospitals within the NRN over time?	We do not have data on how many infants were born at 22-24 weeks' gestation within the geographic areas served by the centers within the NRN. This study is not a regional cohort study. We additionally wrote in the Introduction and Methods that this is a multicenter study and study subjects were from 52 tertiary centers in Japan. Coverage for the national population was described in the Discussion and Supplemental Table 1. These subjects accounted for 34% in period 1 and 42% in period 2 among all infants born at 22-24 weeks' gestation in the nationwide population. We do not have data on referred patterns from maternity hospitals to hospitals within the NRN over time.	P4, L16-18 P5, L10 Supplemental Table 1 P13, L4-7
16	The authors acknowledge that suboptimal follow-up rate of 62%, which raises serious doubt as to the validity of the rates of the various impairments. They then use multiple imputation to estimate rates of these impairments. However, in Table 3 there is no alteration to the numerators of most of the outcomes reported relative to the numerators in supplementary table 2. Surely the numerators should be higher in Table 3, after multiple imputation?	The proportions of CP and visual and hearing impairments were lower than that of developmental delays. This may be one reason for the lack of alterations in the numerators of most of the outcomes reported relative to the numerators after multiple imputations. There may be biases from social, economic, and other unknown factors. Therefore, we performed two single imputations for NDI. As answered in 10 for Reviewer 1, we added this information to the Discussion as limitations and Supplemental Table 4.	Supplemental Table 2 P8, L1-4 P12, L16-L23 Supplemental Table 4

17	The multiple imputation methods are not described. Moreover, the variables included in the models comprise only events known up to the time of birth. As the aim of the imputation is to provide the best estimate of the outcome at 3 years of age for children not assessed, why are the events occurring after birth, particularly those more likely to be related to long-term outcome, such as major cerebral injury and postnatal steroids, not included in the models?	Variables included in the models only comprised events known up to the time of birth and all outcomes. Events occurring after birth in the NICU such as major cerebral injuries and the postnatal administration of steroids may be statistically confounding to prenatal factors. We wanted to clarify that changes in outcomes are related to changes in neonatal care and interventions after birth during these periods; therefore, we only used events known up to the time of birth. We added this information to the Introduction, Methods, and Discussion.	P4, L13-14 P7, 19-21 P12, L9
18	The major results all include death as part of the outcome, particularly in the abstract – the impairment outcomes excluding death should also be emphasized in the abstract and major conclusions, and in Table 4.	The proportion of impairments is related to the proportion of deaths because death is the most unfavorable outcome in these infants. Difference of the mortality rate will affect to the difference of impairments rate in survivors between the two periods. Decreases in death or impairments represent increases in survival without impairments. Therefore, we included death as a main outcome (Table 5). However, as reviewer's comments, outcomes in survivors are clinically important, we showed impairment outcomes in survivors in Tables 3 and 4 and added information to the Results and Abstract.	Table 3 Table 4 P9, L15-19 P2, L17-20
minor-1	Who performed the neurodevelopmental assessments at 3 years of age and what was the level of agreement among assessors for the various outcomes described?	The neurodevelopmental assessment was performed by a trained pediatrician at each center, who was not necessarily blinded to the perinatal details, according to the protocol guidance for the follow-up of VLBW infants. The level of agreement among assessors was measured; however, the protocol was confirmed at the annual meetings of the NRNJ database.	P6, L3-4

minor-2	Why was age not corrected for prematurity?	The protocol for the follow-up of VLBW infants was designed to evaluate prematurity without corrections in the 3-year assessment. We added a reference for this to the Methods. We agree that corrections for prematurity may be better for assessing very extremely preterm infants. However, the purpose of the present study was to describe changes in outcomes between the 2 periods; therefore, we used age without corrections for prematurity.	Reference #16
minor-3	Why were children with CP at GMFCS level 1 excluded? These are the most numerous group among preterm children with any CP	Children with CP at GMFCS level 1 were excluded from the CP group because they were able to walk without devices at 3 years of age, and, thus, were not functionally impaired. We added this to the Methods.	P6, L9
minor-4	Could the rates of bilateral blindness also be reported to be able to compare with other studies?	The numbers and proportions of infants with blindness were added to Table 3.	Table 3
minor-5	It is stated that psychologists were blinded to perinatal details – did they know the participants were all extremely preterm, or were they also assessing term-born children?	They knew the participants were all extremely preterm and they were also assessing term-born children. We added this to the Methods.	P6, L22-23
minor-6	In tables 1 and 2 more data, such as mean differences or odds ratios and 95% CIs should be provided; not just the p-values.	Mean differences and 95% CI were added to Table 1. Since the numerator and denominator are provided for categorical data, we did not add odds ratios.	Table 1
minor-7	Have any of the children remained as inpatients from the time of birth until 3 years of age?	We do not have data for children who remained as inpatients from the time of birth until 3 years of age. Children assessed in this study were all discharged from the NICU and visited outpatient clinics.	No change
mino-8	In supplemental table 1, the rate of males in the without FU group should be 51%, not 56%	Corrected.	Supplemental Table 2
minor-9	In supplemental table 2, “blocks” are misspelled as “brocks”	Corrected.	Table 3

minor-10	The study protocol was approved by the Ethics Review Committees of Tokyo Women's Medical University and Jichi Medical University – did families give informed consent to participate?	Families gave informed consent to participate with written forms or public opt-out in accordance with the Ethics Protocol at each participating center. We added this to the part of appendix, Ethics approval.	P17, L32-P18, L2

VERSION 2 – REVIEW

REVIEWER	Nevitt, Sarah University of Liverpool Competing interests: I have no competing interests
REVIEW RETURNED	05-Dec-2017

GENERAL COMMENTS	I have provided a statistical review of the manuscript “Changes in survival and neurodevelopmental outcomes of infants born at <25 weeks’ gestation in tertiary centers in Japan.” The authors present a comparative analysis of a large retrospective observational database over two time periods and show some important decreases in mortality and morbidity over time. The authors are clearly aware and acknowledge the limitations of retrospective and database analyses (inherent issues with these types of analyses, rather than any fault of the authors), notably missing and incomplete data. The authors carry out detailed assessments and draw a reasonable conclusion that missing data is missing at random. The authors also conduct single and multiple imputations to explore a range of assumptions and demonstrate that results are fairly robust to the missing data, perhaps with the exception of Supplementary Table 4 which is acknowledged by the authors. All conclusions are made clearly in light of the limitations of the data. Overall, I have a couple of very minor comments on wording of this manuscript that the authors may wish to address but otherwise happy to recommend this work for publication Methods: The important consideration here is the change in the database registration methods between the first and the second period. The authors state that “In order to compare the outcomes of infants from the same tertiary centers between the two periods; 2003-2007 (period 1) and 2008-2012 (period 2), centers that registered less than 10 infants during period 1 or 2 were excluded from this study.” I think I understand this statement and I interpret it to mean that any centre that did not recruit 10 infants in BOTH the first AND second period is excluded. That would make sense based on the results presented below but I’m not sure that the statement in the methods is completely clear. It could imply that centres that were not on the database in period 1 could be included for period 2 if more than 10 infants were recruited? Page 7: “These factors were not controllable after birth and were identified as variables associated with outcomes in previous follow-up studies” I don’t understand what ‘controllable after birth’ means in this context – do the authors mean that these are maternal / pregnancy related factors rather than factors related to the infants themselves? Page 9: “The proportion of infants receiving medical treatment or interventions in the NICU significantly increased.” What exactly is this referring to from Table 2?

VERSION 2 – AUTHOR RESPONSE

	Comments	Response *Answers to reviewer's questions*	Pages and lines in the revised text
Reviewer: 1	Methods: The important consideration here is the change in the database registration methods between the first and the second period. The authors state that "In order to compare the outcomes of infants from the same tertiary centers between the two periods; 2003-2007 (period 1) and 2008-2012 (period 2), centers that registered less than 10 infants during period 1 or 2 were excluded from this study." I think I understand this statement and I interpret it to mean that any centre that did not recruit 10 infants in BOTH the first AND second period is excluded. That would make sense based on the results presented below but I'm not sure that the statement in the methods is completely clear. It could imply that centres that were not on the database in period 1 could be included for period 2 if more than 10 infants were recruited?	It means that any center that did not recruit 10 infants in BOTH the first AND second period is excluded. Rewritten to "In order to compare the outcomes of infants from the same tertiary centers between the two periods; 2003-2007 (period 1) and 2008-2012 (period 2), any center that did not recruit 10 infants in both the first and second period were excluded from this study."	Page 5 Line 4-5
2	Page 7: "These factors were not controllable after birth and were identified as variables associated with outcomes in previous follow-up studies" I don't understand what 'controllable after birth' means in this context – do the authors mean that these are maternal / pregnancy related factors rather than factors related to the infants themselves?	We mean that these are maternal / pregnancy related factors and they are not able to change after birth. We rewrote to "These factors were maternal or pregnancy related factors that were not able to change after birth and they were identified as variables associated with outcomes in previous follow-up studies"	Page 7 Line 18-19

3	Page 9: “The proportion of infants receiving medical treatment or interventions in the NICU significantly increased.” What exactly is this referring to from Table 2?	All the medical treatment or interventions listed in Table 2 increased between the first and second periods. We rewrote to “The proportion of infants receiving medical treatment or interventions in the NICU, such as treatments for PDA, late-onset circulatory collapse, chronic lung disease and retinopathy of prematurity, or receiving surfactant, antibiotics, and parental nutrition as listed in Table 2, significantly increased.	Page 9 Line 4-6