PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Do data from child protective services and the police enhance modelling of perinatal risk for paediatric abusive head trauma? A retrospective case-control study
AUTHORS	Kelly, Patrick; Thompron, John; Rungan, Santuri; Ameratunga, Shanthi; Jelleyman, Timothy; Percival, Teuila; Elder, Hinemoa; Mitchell, Edwin

VERSION 1 – REVIEW

REVIEWER	Jacob Andersson Institution of Surgical Sciences, Uppsala University, Sweden
REVIEW RETURNED	07-Jun-2018

GENERAL COMMENTS	Although well written and with some interesting results, this article suffers from methodological errors which cannot be overseen. Considering that the entire study design depends on how AHT is defined and that there is no unequivocal description of the process for diagnosing abuse, there is a high risk of the results being wrong due to false positive AHT-cases. There is no possibility for the reader to assess how a case was considered to be AHT. On page 6, row 38 the classification of AHT is described as: “Briefly, cases were admitted to Starship Children’s Hospital from 1991 to 2010 and met three criteria: (1) Age <2 years (2) Intracranial injury and/or skull fracture (3) AHT diagnosed through a rigorous multi-disciplinary process and reported to statutory authorities (16).” Reference number 16 (Kelly, 2015) describes the process of diagnosing abuse as follows. ”Child protection assessment There is joint assessment by a consultant paediatrician and hospital social worker, meticulous re-taking of the history, detailed physical examination and additional investigations as indicated. Investigations are reviewed with appropriate subspecialists, including radiology side-by-side with paediatric radiologists and neuroradiologists. Findings and differential diagnoses are discussed with the primary team. Further investigations and second opinions are sought where indicated, and all assessments receive weekly multidisciplinary peer review.” As previously stated, it is not possible for the reader to assess how the conclusion of AHT was drawn. A recent and independent review on shaken baby syndrome from the SBU in Sweden concluded there is no evidence to use “the triad” or any of its subcomponents to diagnose abuse and in a later letter discussed the potential dangers of using circular reasoning in future research regarding AHT/SBS. This is a reference which cannot be excluded from a discussion regarding AHT/SBS. The major issue in the AHT/SBS literature today is the presence of

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

	circular reasoning and lack of a proper golden standard for what cases can be considered true cases. Especially since the concept has changed its name from shaken baby syndrome to abusive head trauma which includes a wide arrange of findings with and without obvious signs of trauma. An infant with severe external signs of trauma, skull fracture and fresh subdural hematoma with midline shift cannot be joined in the same group as an infant with bilateral hygroma, increased head circumference and lack of external signs of injury. Was the triad or any of the subcomponents a part of the diagnosis for AHT? What differential diagnoses were discussed? Benign external hydrocephalus? When was a fall regarded as a fall and when was it regarded as a false history? Were there infants which had chronic subdural hematoma or hygroma? These are all questions of importance and which have not been addressed. If there are false positive cases there is also a risk of losing effect in the calculations regarding risk factors for true abusive head trauma. On page 15, row 10, the authors have interesting thoughts on what characteristics to focus on with regard to risk for abuse, neglect or filicide in general. Considering that these risk factors are have been described previously, first and foremost with regard to filicide. “With respect to AHT, we suggest that if the quality and consistency of perinatal health care could be improved, it is the health system which may hold the key to identifying those families most likely to benefit from early intervention. Such improvements could include: routine and universal enquiry during pregnancy for matters of possible relevance such as alcohol and drug abuse, intimate partner violence, unplanned pregnancy and untreated mental illness; routine and universal follow-up of families with missing data or poor engagement with antenatal care; and routine and universal access to evidence-based early intervention programmes when matters of concern are identified by health providers.” Best regards.

REVIEWER	Laura Schummers University of British Columbia, Canada
REVIEW RETURNED	09-Jul-2018

GENERAL COMMENTS	This clearly and well-written manuscript examines the relationship between social data risk of AHT -- a question with potentially important clinical and policy implications. The data set is unique and appropriate for the research question. I would suggest the following revisions to improve this manuscript: 1. In the settings/study population description, please clarify whether the study population was drawn from one hospital (Starship) or 9 hospitals. 2. In the description of the bivariate analyses (page 9, lines 13-16), it is unclear whether the tests used accounted for the matched design. For example, McNemar's test may be preferred to a simple Mantel-Haentzel chi2 test in this context, although not required. Please clarify. 3. I have some concerns regarding the methods used to examine the model performance. While discrimination (measured by AUC) is one important component of risk prediction model performance, risk stratification capacity (extent to which the model is able to divide

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

	patients into groups with clinically distinct risk profiles, such as high risk vs. low risk) is equally important and was not examined for the original model or the expanded model. Adding an evaluation of the risk stratification capacity of the original published model and the model with abuse/criminality data would strengthen the presentation of this model's usefulness considerably. One method the authors may consider is using likelihood ratios: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC478236/. 4. The authors may consider additional methods to assess overall model fit, such as Nagelkerke's r^2, and the extent to which this was improved by the addition of new predictors. 5. Some examination of overfitting of the prediction model should be considered, particularly given the large number of predictors considered and with the relatively small sample size. Specifically, I would suggest presenting optimism-corrected AUC (e.g., after bootstrapping with replacement according to the method described in Steyerberg Chapter 5: Steyerberg EW. Clinical prediction models: A practical approach to development, validation, and updating, vol. 1. New York: Springer; 2009). 6. Although outside the scope of the review for this paper, which examines the increase in predictive ability after including additional variables in the model, I would urge the team to examine risk stratification capacity, and to estimate optimism-corrected measures of discrimination, before the original prediction model is used in clinical practice or policy. 6. It is well documented that new predictors must have very strong associations with the outcome to increase predictive ability substantially (dependent also on prevalence in the population under study). The discussion should include reference and comment on previous literature pertaining to how new predictors can improve prediction model performance (eg Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882–90, among others).

REVIEWER	John M. Leventhal Yale School of Medicine, USA
REVIEW RETURNED	03-Nov-2018

GENERAL COMMENTS	This study uses data from New Zealand and compares cases of abusive head trauma with controls who did not have abusive head trauma and were identified from the same birth hospital. A previous study by the same authors examined perinatal data from birth hospitals comparing cases and controls. The present study aims to determine whether data from child protective services (child welfare) and from police that would be available at the time of birth would improve the predictive model of the occurrence of AHT (abusive head trauma) based on perinatal health data alone. I have three major concerns: 1. The authors do not provide a rationale for why child protective service data or police data obtained prior to the child's birth might be helpful in the predictive model. Maybe it is obvious to the authors, but it would be helpful to explain the rationale to the reader. For example, are there studies that have examined predictive models and used CPS or police data as some of the predictors of child physical abuse or maltreatment? Why might these data be important? Please help the reader understand why

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

	this study is important and what the gap is in the literature that is being filled. 2. Some of the key results of the study are hidden in the supplemental table, which very few readers actually read. Why are these data not presented as a regular table? The supplemental table presents information in too much detail. For example, every variable shows the number of cases and controls with both a yes and a no response. Another example of the for clarity: there are five responses under substantiation, and these should have been indented so that it is clear to the reader that these variables fall under substantiation. One more example: It would help the reader if the authors indicated the number of cases and the number of controls at the top of each column. Maybe some variables like self-harm do not need to be included in the table since the frequency was so very low. Please make the table more accessible to the reader and make it a table in the actual manuscript (article) as opposed to a supplemental table. 3. The main results are presented in table 2, but I am concerned that these are basically the same results that were presented in the original case control study described in reference number 13. Perhaps, the study should be reported as a brief report or a research letter. Additional comments: 4. Page 1, title: is it necessary to have the word “pediatric” before abusive head trauma? 5. Page 2, abstract: very little of the methods related to the case control study is presented in the abstract. It might help the reader to understand more about the selection of cases and controls. In the results it might be clearer to present the univariate odds ratios as opposed to the percentages I the cases and controls. 6. Page 4, strengths and limitations of the study: I am concerned about the first strength. While this is the first case control study as noted by the authors, the first report of this study is reference number 13 in the Journal of Pediatrics. The authors’ statement here should be clarified. 7. Page 5, lines 38 to 40: the sentence is about protective factors for AHT. It is at the end of a paragraph about home visiting. This sentence does not seem to belong here. Please rewrite. 8. Page 6, line 10: please clarify who is “our.” 9. Page 6, lines 54 to 56: Please clarify for the reader that the Ministry for Children relates to all of New Zealand and the same related to the police. For example, in my community police data are kept by towns. Also, the next sentence about decisions on types of data is confusing. Please clarify what is meant here. Does this sentence refer to the researchers or what happened before the researchers of the study were involved? 10. Page 7, line 33: it would be helpful to the reader to provide a brief summary of the data collection, even though that has been described previously. 11. Page 7, line 56: please clarify how far back prior to the baby’s birth data were collected. For example, was there a fixed time prior to the child’s birth or could data have been obtained as far back as the birth of the mother or father? 12. Page 8, line 6: please clarify in the text that a notification (or report of concern) was a report to the child protective services agency.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

	13. Page 9, line 17: this analysis section refers to the conditional logistic regression. Please use a sentence or two to explain this to the reader. Also, why is it necessary to present both the unconditional and conditional results in table 2? If both are necessary, please explain the purpose to the reader. 14. Page 9, line 24 to page 10, line 13: This section describes how the model was built in the actual results, and I think it should be moved to the results section. 15. Page 11, line 45: the text says "notification to child welfare," but Table 2 says "report of concern." Please use the same language in the text and the table so that the reader does not get confused. 15. Page 11, line 47: why is the unconditional odds ratio presented here as opposed to the conditional one? Please clarify for the reader. 16. Page 12, lines 12 to 26: much of this text can be eliminated or shortened since the information is in the table. 17. Page 12, line 54: the authors present the area under the curve. It would help the reader if in the methods, the authors briefly explain this. Also, why was sensitivity and specificity not provided here? 18. Page 15: the authors compare their results to other New Zealand studies on predictive risk modeling. This is done in the concluding paragraph. It would help the reader if this were done in the body of the discussion. More information could be provided about some of the differences between this study and the previous studies; this information would be helpful to the reader. 19. No paragraph was providing about limitations. This needs to be included in the discussion. 20. The STROBE checklist appears to be provided in Table 1. There is no mention in the text about this, and the table is not labeled as Table 1. Unless required by the journal, I think this could be a supplemental table.

VERSION 1 – AUTHOR RESPONSE

REVIEWER: 1 REVIEWER NAME: JACOB ANDERSSON
INSTITUTION AND COUNTRY: INSTITUTION OF SURGICAL SCIENCES, UPPSALA UNIVERSITY, SWEDEN
PLEASE STATE ANY COMPETING INTERESTS OR STATE 'NONE DECLARED': NONE

ALTHOUGH WELL WRITTEN AND WITH SOME INTERESTING RESULTS, THIS ARTICLE SUFFERS FROM METHODOLOGICAL ERRORS WHICH CANNOT BE OVERSEEN. CONSIDERING THAT THE ENTIRE STUDY DESIGN DEPENDS ON HOW AHT IS DEFINED AND THAT THERE IS NO UNEQUIVOCAL DESCRIPTION OF THE PROCESS FOR DIAGNOSING ABUSE, THERE IS A HIGH RISK OF THE RESULTS BEING WRONG DUE TO FALSE POSITIVE AHT-CASES. THERE IS NO POSSIBILITY FOR THE READER TO ASSESS HOW A CASE WAS CONSIDERED TO BE AHT.

ON PAGE 6, ROW 38 THE CLASSIFICATION OF AHT IS DESCRIBED AS: "BRIEFLY, CASES WERE ADMITTED TO STARSHIP CHILDREN'S HOSPITAL FROM 1991 TO 2010 AND MET THREE CRITERIA: (1) AGE <2 YEARS (2) INTRACRANIAL INJURY AND/OR SKULL FRACTURE (3) AHT DIAGNOSED THROUGH A RIGOROUS MULTI-DISCIPLINARY PROCESS AND REPORTED TO STATUTORY AUTHORITIES (16)." REFERENCE NUMBER 16 (KELLY, 2015)

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

DESCRIBES THE PROCESS OF DIAGNOSING ABUSE AS FOLLOWS. ”CHILD PROTECTION ASSESSMENT. THERE IS JOINT ASSESSMENT BY A CONSULTANT PAEDIATRICIAN AND HOSPITAL SOCIAL WORKER, METICULOUS RE-TAKING OF THE HISTORY, DETAILED PHYSICAL EXAMINATION AND ADDITIONAL INVESTIGATIONS AS INDICATED. INVESTIGATIONS ARE REVIEWED WITH APPROPRIATE SUBSPECIALISTS, INCLUDING RADIOLOGY SIDE-BY-SIDE WITH PAEDIATRIC RADIOLOGISTS AND NEURORADIOLOGISTS. FINDINGS AND DIFFERENTIAL DIAGNOSES ARE DISCUSSED WITH THE PRIMARY TEAM. FURTHER INVESTIGATIONS AND SECOND OPINIONS ARE SOUGHT WHERE INDICATED, AND ALL ASSESSMENTS RECEIVE WEEKLY MULTIDISCIPLINARY PEER REVIEW.” AS PREVIOUSLY STATED, IT IS NOT POSSIBLE FOR THE READER TO ASSESS HOW THE CONCLUSION OF AHT WAS DRAWN.

A RECENT AND INDEPENDENT REVIEW ON SHAKEN BABY SYNDROME FROM THE SBU IN SWEDEN CONCLUDED THERE IS NO EVIDENCE TO USE “THE TRIAD” OR ANY OF ITS SUBCOMPONENTS TO DIAGNOSE ABUSE AND IN A LATER LETTER DISCUSSED THE POTENTIAL DANGERS OF USING CIRCULAR REASONING IN FUTURE RESEARCH REGARDING AHT/SBS. THIS IS A REFERENCE WHICH CANNOT BE EXCLUDED FROM A DISCUSSION REGARDING AHT/SBS. THE MAJOR ISSUE IN THE AHT/SBS LITERATURE TODAY IS THE PRESENCE OF CIRCULAR REASONING AND LACK OF A PROPER GOLDEN STANDARD FOR WHAT CASES CAN BE CONSIDERED TRUE CASES. ESPECIALLY SINCE THE CONCEPT HAS CHANGED ITS NAME FROM SHAKEN BABY SYNDROME TO ABUSIVE HEAD TRAUMA WHICH INCLUDES A WIDE ARRANGE OF FINDINGS WITH AND WITHOUT OBVIOUS SIGNS OF TRAUMA. AN INFANT WITH SEVERE EXTERNAL SIGNS OF TRAUMA, SKULL FRACTURE AND FRESH SUBDURAL HEMATOMA WITH MIDLINE SHIFT CANNOT BE JOINED IN THE SAME GROUP AS AN INFANT WITH BILATERAL HYGROMA, INCREASED HEAD CIRCUMFERENCE AND LACK OF EXTERNAL SIGNS OF INJURY.

WAS THE TRIAD OR ANY OF THE SUBCOMPONENTS A PART OF THE DIAGNOSIS FOR AHT? WHAT DIFFERENTIAL DIAGNOSES WERE DISCUSSED? BENIGN EXTERNAL HYDROCEPHALUS? WHEN WAS A FALL REGARDED AS A FALL AND WHEN WAS IT REGARDED AS A FALSE HISTORY? WERE THERE INFANTS WHICH HAD CHRONIC SUBDURAL HEMATOMA OR HYGROMA? THESE ARE ALL QUESTIONS OF IMPORTANCE AND WHICH HAVE NOT BEEN ADDRESSED.

IF THERE ARE FALSE POSITIVE CASES THERE IS ALSO A RISK OF LOSING EFFECT IN THE CALCULATIONS REGARDING RISK FACTORS FOR TRUE ABUSIVE HEAD TRAUMA.

ON PAGE 15, ROW 10, THE AUTHORS HAVE INTERESTING THOUGHTS ON WHAT CHARACTERISTICS TO FOCUS ON WITH REGARD TO RISK FOR ABUSE, NEGLECT OR FILICIDE IN GENERAL. CONSIDERING THAT THESE RISK FACTORS ARE HAVE BEEN DESCRIBED PREVIOUSLY, FIRST AND FOREMOST WITH REGARD TO FILICIDE. “WITH RESPECT TO AHT, WE SUGGEST THAT IF THE QUALITY AND CONSISTENCY OF PERINATAL HEALTH CARE COULD BE IMPROVED, IT IS THE HEALTH SYSTEM WHICH MAY HOLD THE KEY TO IDENTIFYING THOSE FAMILIES MOST LIKELY TO BENEFIT FROM EARLY INTERVENTION. SUCH IMPROVEMENTS COULD INCLUDE: ROUTINE AND UNIVERSAL ENQUIRY DURING PREGNANCY FOR MATTERS OF POSSIBLE RELEVANCE SUCH AS ALCOHOL AND DRUG ABUSE, INTIMATE PARTNER VIOLENCE, UNPLANNED PREGNANCY AND UNTREATED MENTAL ILLNESS; ROUTINE AND UNIVERSAL FOLLOW-UP OF FAMILIES WITH MISSING DATA OR POOR ENGAGEMENT WITH ANTENATAL CARE; AND ROUTINE AND UNIVERSAL ACCESS TO EVIDENCE-BASED EARLY INTERVENTION PROGRAMMES WHEN MATTERS OF CONCERN ARE IDENTIFIED BY HEALTH PROVIDERS.”

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Our rebuttal:
This reviewer's published expertise in the field of child abuse is limited to two scientific articles, both published within the last 12 months in Acta Paediatrica and both representing an extreme and highly controversial view in the field of abusive head trauma. The first was published in November 2017 ("National study shows that abusive head trauma mortality in Sweden was at least 10 times lower than in other Western countries") (1) and the second in February 2018 ("It is important not to assume an aetiology for the triad before the outcomes of diagnostic investigations") (2). Both articles challenge the currently-accepted international diagnostic criteria for abusive head trauma. We should point out that the so-called "triad" is a straw man. It is not a diagnostic criterion for abusive head trauma and does not appear in any of our papers on the subject of abusive head trauma.

Dr Andersson states that in our paper "there is a high risk of the results being wrong due to false positive AHT-cases". He provides no scientific data whatsoever to support that assertion. He appears to derive this view from what he later describes as "a recent and independent review on shaken baby syndrome from the SBU in Sweden" (3). We will discuss this review in more detail shortly - suffice it to say that it is widely regarded as junk science by multiple paediatric specialties on both sides of the Atlantic.

Dr Andersson states that in our paper "there is no unequivocal description of the process for diagnosing abuse... There is no possibility for the reader to assess how a case was considered to be AHT." This is nonsense. He quotes one of our previous papers, which describes in great detail a 20-year series of cases with abusive head trauma. He concludes that paper provides inadequate information about our diagnostic process. In fact, the paper in question was rigorously peer-reviewed and published in Archives of Disease in Childhood in 2015.(4) The case-control study which you at first chose to reject based on Dr Andersson's opinion, is the second phase of a case-control study which was also rigorously peer-reviewed and published in the Journal of Pediatrics in 2017. (5) That paper in the Journal of Pediatrics used exactly the same set of cases of AHT as our current paper, diagnosed in the manner described in Archives in 2015.

Our 2015 study in Archives has been cited 12 times. Two of the most recent citations are of direct relevance to Dr Andersson's critique.

Firstly, it was cited most recently in August in an article from Cardiff University, using the diagnostic process in our study as an example of "a thorough clinical and forensic investigation". (6)

Secondly, it was cited in May in a "Consensus statement on abusive head trauma in infants and young children".(7) This was published in the journal Pediatric Radiology, authored jointly by the Society for Pediatric Radiology (SPR), the European Society of Paediatric Radiology (ESPR), the American Society of Pediatric Neuroradiology (ASPNR), the American Academy of Pediatrics (AAP), the European Society of Neuroradiology (ESNR), the American Professional Society on the Abuse of Children (APSAC), the Swedish Paediatric Society, the Norwegian Pediatric Association and the Japanese Pediatric Society. This paper states that "The diagnosis of AHT is a medical diagnosis made by a multidisciplinary team of pediatricians and pediatric subspecialty physicians, social workers and other professionals based on consideration of all the facts and evidence", which Is exactly the methodology we followed in our study. This consensus statement cites with approval some of the key data from our 2015 paper from Archives. In contrast, the same article describes the SBU report (regarded with approval by Dr Andersson) as "flawed by "(1) improper search and systemic review questions, (2) improper criteria for assessing bias and (3) inequitable application of quality of study assessment standards."

Dr Andersson asserts that the SBU report "is a reference which cannot be excluded from a discussion regarding AHT/SBS". In fact, it can and should be excluded, because it is junk science. The SBU

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

report is so bad that it triggered the creation of the consensus statement I have just cited. That statement represents the current international scientific consensus on abusive head trauma. The SBU report - and, clearly, Dr Andersson - do not.

We refer the Editors to the recent critique published by the Royal College of Paediatrics and Child Health in Archives of Disease in Childhood, on behalf of the Child Protection Standing Committee, Royal College of Paediatrics and Child Health. This paper concludes that "Due to the critical methodological flaws, we recommend that [the SBU] review is withdrawn from publication for the sake of the unbiased protection of children who may have suffered from AHT".(8)

If you chose to reject our paper on the basis of the SBU review, then you would be rejecting good science on the basis of junk science

References:
1. Andersson J, Thiblin I. National study shows that abusive head trauma mortality in Sweden was at least 10 times lower than in other Western countries. Acta Paediatr. 2018 Mar;107(3):477-483.
2. Andersson J, Thiblin I. It is important not to assume an aetiology for the triad before the outcomes of diagnostic investigations. Acta Paediatr. 2018 Aug;107(8):1308-1309.
3. SBU. Traumatic shaking: The role of the triad in medical investigations of suspected traumatic shaking: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU), 2016.
4. Kelly P, John S, Vincent AL, Reed P. Abusive head trauma and accidental head injury: a 20 year comparative study of referrals to a hospital child protection team. Arch Dis Child. 2015;100(12):1123-30
5. Kelly P, Thompson JMD, Koh J, Ameratunga S, Jelleyman T, Percival TM, Elder H, Mitchell EA. Perinatal risk and protective factors for pediatric abusive head trauma: a multicenter case-control study. J Pediatr. 2017;187:240-246.
6. Cowley LE, Maguire S, Farewell DM, Quinn-Scoggins HD, Flynn MO, Kemp AM. Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. Child Abuse Negl. 2018 Aug;82:178-191.
7. Choudhary AK, Servaes S, Slovis TL, Palusci VJ, Hedlund GL, Narang SK, Moreno JA, Dias MS, Christian CW, Nelson MD Jr, Silvera VM, Palasis S, Raissaki M, Rossi A, Offiah AC. Consensus statement on abusive head trauma in infants and young children. Pediatr Radiol. 2018 Aug;48(8):1048-1065.
8. Debelle GD, Maguire S, Watts P, Nieto Hernandez R, Kemp AM; Child Protection Standing Committee, Royal College of Paediatrics and Child Health. Abusive head trauma and the triad: a critique on behalf of RCPCH of 'Traumatic shaking: the role of the triad in medical investigations of suspected traumatic shaking'. Arch Dis Child. 2018 Jun;103(6):606-610

REVIEWER: 2 REVIEWER NAME: LAURA SCHUMMERS
INSTITUTION AND COUNTRY: UNIVERSITY OF BRITISH COLUMBIA, CANADA
PLEASE STATE ANY COMPETING INTERESTS OR STATE ‘NONE DECLARED’: NONE DECLARED.

1. THIS CLEARLY AND WELL-WRITTEN MANUSCRIPT EXAMINES THE RELATIONSHIP BETWEEN SOCIAL DATA RISK OF AHT -- A QUESTION WITH POTENTIALLY IMPORTANT CLINICAL AND POLICY IMPLICATIONS. THE DATA SET IS UNIQUE AND APPROPRIATE FOR THE RESEARCH QUESTION.
We are grateful to the Reviewer for these positive remarks about the style of our manuscript, the question we set out to study and the appropriateness of our data set

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

2. I WOULD SUGGEST THE FOLLOWING REVISIONS TO IMPROVE THIS MANUSCRIPT: IN THE SETTINGS/STUDY POPULATION DESCRIPTION, PLEASE CLARIFY WHETHER THE STUDY POPULATION WAS DRAWN FROM ONE HOSPITAL (STARSHIP) OR 9 HOSPITALS.
We agree this was unclear in our manuscript. Although the cases were admitted to Starship and diagnosed in Starship, Starship is not a maternity hospital and this was a study of perinatal records. Cases were therefore included in the study if we could access their perinatal records, and for reasons of resource constraint, only nine of the sixteen maternity hospitals where cases were born could be enrolled in the study. However, those nine maternity hospitals accounted for 92% of the cases of AHT identified at Starship. The reasons for this inclusion criterion (and the numbers excluded by using that criterion) are covered in detail in the original article, and we hope the Reviewer will agree that it is not necessary to repeat them here.

To clarify this in the text, we have deleted the phrase "in nine maternity hospitals" from the first sentence and added a new sentence as one of the inclusion criteria: "(4) Born in one of nine participating maternity hospitals in the North Island of New Zealand." We have also rewritten the next sentence to read: "The study population included these cases and four controls for each case randomly selected from babies born on the same day in the same maternity hospital."

3. IN THE DESCRIPTION OF THE BIVARIATE ANALYSES (PAGE 9, LINES 13-16), IT IS UNCLEAR WHETHER THE TESTS USED ACCOUNTED FOR THE MATCHED DESIGN.
We agree. We have re-written the first part of the second paragraph under "Statistical Analysis" in the Methods section, as follows: "Data were first tested for difference in frequency between cases and controls in simple cross-tabulations and logistic regressions, not accounting for the matched design. The χ^2 test was used for categorical data and t-tests for continuous data."

4. FOR EXAMPLE, MCNEMAR'S TEST MAY BE PREFERRED TO A SIMPLE MANTEL-HAENTZEL CHI2 TEST IN THIS CONTEXT, ALTHOUGH NOT REQUIRED. PLEASE CLARIFY.
We take the reviewer's point. However, as shown in Table 1 (previously a supplementary table), the relationships are mostly highly significant. We have now looked at these relationships with McNemar's test, and the results are essentially the same. As one example, the only variable from Table 1 included in the multivariable model is notification to child protective services. For this variable, the unconditional univariable OR (7.24, 95%CI 4.70-11.14) and the conditional univariable OR (8.59, 95%CI 5.24-14.09) were very similar in magnitude. Rather than presenting the results of McNemar's test as a separate analysis, we have now provided both sets of OR in Table 1 (conditional and unconditional) so the readers can see the similarity for themselves.

5. I HAVE SOME CONCERNS REGARDING THE METHODS USED TO EXAMINE THE MODEL PERFORMANCE. WHILE DISCRIMINATION (MEASURED BY AUC) IS ONE IMPORTANT COMPONENT OF RISK PREDICTION MODEL PERFORMANCE, RISK STRATIFICATION CAPACITY (EXTENT TO WHICH THE MODEL IS ABLE TO DIVIDE PATIENTS INTO GROUPS WITH CLINICALLY DISTINCT RISK PROFILES, SUCH AS HIGH RISK VS. LOW RISK) IS EQUALLY IMPORTANT AND WAS NOT EXAMINED FOR THE ORIGINAL MODEL OR THE EXPANDED MODEL. ADDING AN EVALUATION OF THE RISK STRATIFICATION CAPACITY OF THE ORIGINAL PUBLISHED MODEL AND THE MODEL WITH ABUSE/CRIMINALITY DATA WOULD STRENGTHEN THE PRESENTATION OF THIS MODEL'S USEFULNESS CONSIDERABLY. ONE METHOD THE AUTHORS MAY CONSIDER IS USING LIKELIHOOD RATIOS: https://scanmail.trustwave.com/?c=7264&d=k-7g24g95hz5pRUi4asYe-zzLKRK-Ax78LSgK63WGQ&u=https%3a%2f%2fwww%2encbi%2enlm%2enih%2egov%2fpmc%2farticles%2fPMC478236%2f
We agree with the reviewer that discrimination is only one measure of model performance.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

The reviewer’s question made it apparent that we had not stated clearly enough (in this manuscript) that our model should NOT be used to divide patients into distinct risk profiles. It is not fit for that purpose. In the first published paper describing this model, we wrote: “Despite the high AUC and the fact that our model has significant predictive value, we do not suggest that our study provides an adequate framework (e.g., appropriate sensitivity and specificity) for predicting risk in individual families. It must be emphasized that more research is required to replicate our findings and clarify their significance. We cannot, for example, exclude the possibility of residual confounding from measured or unmeasured factors.”

In terms of risk stratification capacity, we were not entirely clear what the reviewer was asking for. After consultation with several statisticians, we concluded that the reviewer wished us to identify whether any of the individual variables in our model had a likelihood ratio that would justify stratification of the population in the use of our model. Accordingly, we assessed the likelihood ratios for each of the variables in the model. The largest positive likelihood ratio was 4.43, and the smallest negative likelihood ratio was 0.56. Therefore, we do not believe that there is any variable that would be useful for stratification purposes.

We have now made it very clear in the discussion that our model is not suitable for dividing patients into groups with clinically distinct risk profiles (see changes in the text set out below). We have also added the pseudo R-squared statistic (Nagelkerke’s R2) as requested by the reviewer in her next comment. We feel that the further addition of the likelihood ratios to the manuscript is unnecessary and is likely to be unhelpful to the average reader.

The changes made in response to this reviewer’s points 6 and 9 are as follows: “Finally, although our model can discriminate between two populations (as shown by the AUC), it explains only one third of the variation between those populations (as shown by the pseudo R squared statistic). Statistical associations between AHT and a variable do not necessarily mean that the variable will usefully identify those individuals more or less likely to experience AHT. Factors associated with very high ORs may still turn out to be unhelpful as individual-level predictors.”

In specific response to point 5, we have also added the following concluding paragraph:
“Finally, while our study demonstrates that health data have the potential to be useful in predicting the risk of AHT, our current model is not good enough to guide clinical practice or policy. Neither the model, nor any of the variables it contains, can be used to guide clinical interventions in specific families. More work is needed to replicate our findings, investigate other potentially relevant variables and examine possible confounders which may underlie or explain some of the variables in our model.”

6. THE AUTHORS MAY CONSIDER ADDITIONAL METHODS TO ASSESS OVERALL MODEL FIT, SUCH AS NAGELKERKE'S R^2, AND THE EXTENT TO WHICH THIS WAS IMPROVED BY THE ADDITION OF NEW PREDICTORS.
We agree, and again we thank the reviewer for this suggestion. We have assessed Nagelkerke’s R2 to our data and have made the following changes to the text of the manuscript.

In the Methods section, last paragraph under the heading of “Statistical Analysis”, we have added: “The proportion of the total variability of the outcome (AHT) that could be accounted for by our model was assessed using a pseudo R-squared statistic (Nagelkerke’s R2).”
In the Results section, last paragraph, we have added the following: “Similarly, Nagelkerke’s R2 remained much the same. For the previous model it was 33.1% (for both conditional and unconditional regression), and for the new model it was 34.5% (conditional) or 35.5% (unconditional).”

We have also noted this in the strengths and limitations paragraph in the discussion, as set out above in our response to point 5.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

7. SOME EXAMINATION OF OVERFITTING OF THE PREDICTION MODEL SHOULD BE CONSIDERED, PARTICULARLY GIVEN THE LARGE NUMBER OF PREDICTORS CONSIDERED AND WITH THE RELATIVELY SMALL SAMPLE SIZE. SPECIFICALLY, I WOULD SUGGEST PRESENTING OPTIMISM-CORRECTED AUC (E.G., AFTER BOOTSTRAPPING WITH REPLACEMENT ACCORDING TO THE METHOD DESCRIBED IN STEYERBERG CHAPTER 5: STEYERBERG EW. CLINICAL PREDICTION MODELS: A PRACTICAL APPROACH TO DEVELOPMENT, VALIDATION, AND UPDATING, VOL. 1. NEW YORK: SPRINGER; 2009).
Thank you for this suggestion, which we have considered seriously. We take the Reviewer's point that we have 10 predictors for a sample size of 142 cases and 550 controls. However, our "sample" is effectively almost the entire population (92%) of cases seen in Starship, the only tertiary referral centre in New Zealand for AHT. Bootstrapping would be unlikely significantly to change our results and would be complex given that each case was matched to up to 4 controls by date and hospital of birth. In addition, this reviewer's other suggestions (Nagelkerke's R2 and more discussion of the limitations of OR) have clarified our presentation of the limitations of our data.

8. ALTHOUGH OUTSIDE THE SCOPE OF THE REVIEW FOR THIS PAPER, WHICH EXAMINES THE INCREASE IN PREDICTIVE ABILITY AFTER INCLUDING ADDITIONAL VARIABLES IN THE MODEL, I WOULD URGE THE TEAM TO EXAMINE RISK STRATIFICATION CAPACITY, AND TO ESTIMATE OPTIMISM-CORRECTED MEASURES OF DISCRIMINATION, BEFORE THE ORIGINAL PREDICTION MODEL IS USED IN CLINICAL PRACTICE OR POLICY.
We agree. However, we do not believe our current model (in either version) should be used in clinical practice or policy, and we hope we have now made this clear in the changes in the text which we have described above.

9. IT IS WELL DOCUMENTED THAT NEW PREDICTORS MUST HAVE VERY STRONG ASSOCIATIONS WITH THE OUTCOME TO INCREASE PREDICTIVE ABILITY SUBSTANTIALLY (DEPENDENT ALSO ON PREVALENCE IN THE POPULATION UNDER STUDY). THE DISCUSSION SHOULD INCLUDE REFERENCE AND COMMENT ON PREVIOUS LITERATURE PERTAINING TO HOW NEW PREDICTORS CAN IMPROVE PREDICTION MODEL PERFORMANCE (EG PEPE MS, JANES H, LONGTON G, LEISENRING W, NEWCOMB P. LIMITATIONS OF THE ODDS RATIO IN GAUGING THE PERFORMANCE OF A DIAGNOSTIC, PROGNOSTIC, OR SCREENING MARKER. AM J EPIDEMIOL. 2004;159(9):882–90, AMONG OTHERS).
We agree, and we thank the reviewer for drawing our attention to this article, which has an excellent discussion of the issues. We have now added consideration of these issues in two places in the discussion section of our manuscript, both referencing Pepe's 2004 article.

Firstly, as already noted above in point 5, under strengths and limitations, as below:
"Statistical associations between AHT and a variable do not necessarily mean that the variable will usefully identify those individuals more or less likely to experience AHT. Factors associated with very high ORs may still turn out to be unhelpful as individual-level predictors."

Secondly, with specific reference to addition of a new predictor, with two new sentences at the end of paragraph 3 of the discussion, which discusses notification as an indicator of risk:
"However, although the OR was high by traditional epidemiologic standards, it is not surprising that it added little to the predictive value of our model. Because of the limitations mentioned above, a variable with an apparently strong independent association with the outcome (estimated by OR) will often not contribute meaningfully to predictive accuracy."

REVIEWER: 3 REVIEWER NAME: JOHN M. LEVENTHAL

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

INSTITUTION AND COUNTRY: YALE SCHOOL OF MEDICINE, USA
PLEASE STATE ANY COMPETING INTERESTS OR STATE 'NONE DECLARED': NONE DECLARED

10. THIS STUDY USES DATA FROM NEW ZEALAND AND COMPARES CASES OF ABUSIVE HEAD TRAUMA WITH CONTROLS WHO DID NOT HAVE ABUSIVE HEAD TRAUMA AND WERE IDENTIFIED FROM THE SAME BIRTH HOSPITAL. A PREVIOUS STUDY BY THE SAME AUTHORS EXAMINED PERINATAL DATA FROM BIRTH HOSPITALS COMPARING CASES AND CONTROLS. THE PRESENT STUDY AIMS TO DETERMINE WHETHER DATA FROM CHILD PROTECTIVE SERVICES (CHILD WELFARE) AND FROM POLICE THAT WOULD BE AVAILABLE AT THE TIME OF BIRTH WOULD IMPROVE THE PREDICTIVE MODEL OF THE OCCURRENCE OF AHT (ABUSIVE HEAD TRAUMA) BASED ON PERINATAL HEALTH DATA ALONE.
We thank the reviewer for this concise summary of our study methodology

11. I HAVE THREE MAJOR CONCERNS: THE AUTHORS DO NOT PROVIDE A RATIONALE FOR WHY CHILD PROTECTIVE SERVICE DATA OR POLICE DATA OBTAINED PRIOR TO THE CHILD'S BIRTH MIGHT BE HELPFUL IN THE PREDICTIVE MODEL. MAYBE IT IS OBVIOUS TO THE AUTHORS, BUT IT WOULD BE HELPFUL TO EXPLAIN THE RATIONALE TO THE READER. FOR EXAMPLE, ARE THERE STUDIES THAT HAVE EXAMINED PREDICTIVE MODELS AND USED CPS OR POLICE DATA AS SOME OF THE PREDICTORS OF CHILD PHYSICAL ABUSE OR MALTREATMENT? WHY MIGHT THESE DATA BE IMPORTANT? PLEASE HELP THE READER UNDERSTAND WHY THIS STUDY IS IMPORTANT AND WHAT THE GAP IS IN THE LITERATURE THAT IS BEING FILLED.
We agree. The most appropriate place for this explanation is in the introduction, so we have largely rewritten the introduction to provide it. This entailed rewriting the second paragraph, deleting most of the third paragraph (essentially replacing it with a new third paragraph), adding a new fourth paragraph which explains some of the key literature regarding the role of CPS or police data in risk assessment or predictive models and deleting most of the second sentence of the old fourth paragraph (now the fifth paragraph). We also added a sentence at the end of the last paragraph to further explain why the data we sought was restricted to information which might be obtainable at or around the time of birth. Our specific changes are highlighted in red on the mark-up copy, but the amended paragraphs are included here in toto for the benefit of the Reviewer.

"It would clearly be best, if possible, to take steps to prevent AHT before it occurs. Because AHT often occurs in response to crying, current prevention strategies focus on teaching all new parents about the dangers of shaking and how to cope with a crying baby.

However, it is likely that age and crying are not the only risk factors for AHT. Studies identify a variety of risk factors for other forms of child abuse, some of which have also been identified in cohort studies of AHT. It seems reasonable to suggest that there may be circumstances in which the risk of AHT is increased. If those could be identified, there may be benefit in interventions targeted at those circumstances and/or at specific families where such circumstances exist.

Targeted interventions are common in prevention strategies for other forms of child abuse. One example is regular home visits in early childhood ("home visiting"), where families qualify for visits after a risk assessment often including criteria such as a history of child abuse, intimate partner violence (IPV), substance abuse or criminal justice involvement. Recently, New Zealand economists used public benefit and child protection records for 57,986 children to develop a general predictive risk model for child abuse. The outcome variable was defined as a "substantiated report of maltreatment by the age of 5 years." The authors suggested this model could be used to target home visiting at those most likely to benefit. Predictor variables included (among many others) "child protection service reports for other children", "substantiated physical or sexual abuse before age 16

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

years”, “partner has criminal record”, “police family violence reports” and “youth justice referrals for partner before age 16 years.”

In a recent multi-centre case-control study, we used variables from routinely collected perinatal health records in an attempt to construct a model that could predict the risk of AHT. However, those records contained little or no information on the possible risk factors outlined above.

The purpose of this study was, therefore, to obtain data from sources outside the health system which respond to child abuse and adult criminality, to investigate whether there is a relationship between those data and the risk of AHT and to determine whether incorporating such data would improve the ability of primary healthcare providers to assess risk in the perinatal period. Because AHT has a median age at diagnosis of 5 months, it seemed appropriate to focus our investigation on information which might be obtainable at or before the time of birth.”

12. SOME OF THE KEY RESULTS OF THE STUDY ARE HIDDEN IN THE SUPPLEMENTAL TABLE, WHICH VERY FEW READERS ACTUALLY READ. WHY ARE THESE DATA NOT PRESENTED AS A REGULAR TABLE?
We agree. The Reviewer makes an excellent point. We now include these data as a regular Table. This has enabled us to shorten considerably some of the text of the results section, where we had previously repeated some of the data from the Supplementary Table for readers’ convenience.

13. THE SUPPLEMENTAL TABLE PRESENTS INFORMATION IN TOO MUCH DETAIL. FOR EXAMPLE, EVERY VARIABLE SHOWS THE NUMBER OF CASES AND CONTROLS WITH BOTH A YES AND A NO RESPONSE.
We agree. We have deleted all the “No” rows.

14. ANOTHER EXAMPLE OF THE FOR CLARITY: THERE ARE FIVE RESPONSES UNDER SUBSTANTIATION, AND THESE SHOULD HAVE BEEN INDENTED SO THAT IT IS CLEAR TO THE READER THAT THESE VARIABLES FALL UNDER SUBSTANTIATION.
We agree. We have indented these 5 variables

15. ONE MORE EXAMPLE: IT WOULD HELP THE READER IF THE AUTHORS INDICATED THE NUMBER OF CASES AND THE NUMBER OF CONTROLS AT THE TOP OF EACH COLUMN.
We agree. We have added the number of cases and the number of controls at the top of each column

16. MAYBE SOME VARIABLES LIKE SELF-HARM DO NOT NEED TO BE INCLUDED IN THE TABLE SINCE THE FREQUENCY WAS SO VERY LOW.
We agree. We have deleted this variable and one other variable with a similarly low frequency. This necessitated other changes. Firstly, another sentence in the second paragraph of the section on Statistical Analysis: “Variables with very low frequency in the dataset were removed.” Secondly, in the Results section (first line under “Data from child protective services”), “Sixteen” changes to “Fourteen.”

17. PLEASE MAKE THE TABLE MORE ACCESSIBLE TO THE READER AND MAKE IT A TABLE IN THE ACTUAL MANUSCRIPT (ARTICLE) AS OPPOSED TO A SUPPLEMENTAL TABLE.
We agree. The Reviewer’s earlier suggestions have made the Table much less cumbersome and we have now included it in the actual manuscript as Table 1.

18. THE MAIN RESULTS ARE PRESENTED IN TABLE 2, BUT I AM CONCERNED THAT THESE ARE BASICALLY THE SAME RESULTS THAT WERE PRESENTED IN THE ORIGINAL CASE CONTROL STUDY DESCRIBED IN REFERENCE NUMBER 13. PERHAPS, THE STUDY SHOULD BE REPORTED AS A BRIEF REPORT OR A RESEARCH LETTER.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

We seriously considered reporting this study as a short report. However, despite our best efforts to be concise, even the first version of this manuscript required 2790 words. The original study was a complex case-control study. Although the findings of this further study are essentially negative (additional data obtained from the statutory authorities made very little difference to our ability to predict risk), we again used an innovative approach and gathered and analysed a large amount of novel data from child protective services and the police. A study of AHT using such data has not been done before. The significance or lack of significance of our negative findings cannot be properly appreciated without fully presenting the rationale for the further study, the scope of the data obtained and the limitations of those data and our data analysis. The complexity of the issues is made evident by the extensive revisions requested by Reviewers 2 and 3, which are all reasonable requests, but have resulted in further expansion of the manuscript to 3222 words. In summary, we do not believe it is possible adequately to present this study as a short report.

19. PAGE 1, TITLE: IS IT NECESSARY TO HAVE THE WORD “PEDIATRIC” BEFORE ABUSIVE HEAD TRAUMA?
We agree that in a specifically paediatric journal, adding the adjective paediatric may be unnecessary - although we note that the phrase “pediatric abusive head trauma” appears in the title of our case-control study published in the Journal of Pediatrics. However, we are not sure if the same applies in a journal such as the BMJ Open, which is directed to a wider audience. At present we have left the adjective in, but we would be happy to remove it at the editors’ request.

20. PAGE 2, ABSTRACT: VERY LITTLE OF THE METHODS RELATED TO THE CASE CONTROL STUDY IS PRESENTED IN THE ABSTRACT. IT MIGHT HELP THE READER TO UNDERSTAND MORE ABOUT THE SELECTION OF CASES AND CONTROLS.
We agree. Under “Design” we have added that this is a retrospective case control study “of child protective service and police records”. We have expanded the description of the selection of cases and controls under the heading “Participants”. Previously, this was “142 cases and 550 controls matched by date and hospital of birth”. It now reads: “142 consecutive cases of abusive head trauma admitted to a tertiary children’s hospital from 1991 to 2010 and born in one of the nine participating maternity hospitals. 550 controls matched by the date and hospital of birth”

21. IN THE RESULTS IT MIGHT BE CLEARER TO PRESENT THE UNIVARIATE ODDS RATIOS AS OPPOSED TO THE PERCENTAGES I THE CASES AND CONTROLS.
We agree. We have replaced the percentages for five variables with the univariable odds ratios. These changes are shown in red in the Marked-Up copy. Because we expanded the description in the abstract of our selection criteria (in response to the reviewer’s earlier request), we have had to delete the data for two variables from the abstract (partner violence offence and alcohol offence) and shorten the text slightly in order to remain within the word limit. The shortened text is shown in red in the Marked-Up copy.

22. PAGE 4, STRENGTHS AND LIMITATIONS OF THE STUDY: I AM CONCERNED ABOUT THE FIRST STRENGTH. WHILE THIS IS THE FIRST CASE CONTROL STUDY AS NOTED BY THE AUTHORS, THE FIRST REPORT OF THIS STUDY IS REFERENCE NUMBER 13 IN THE JOURNAL OF PEDIATRICS. THE AUTHORS’ STATEMENT HERE SHOULD BE CLARIFIED.
We agree. To avoid any misunderstanding, we have deleted this first bullet-point and replaced it by the following: “This study is a case-control study examining risk factors for paediatric abusive head trauma, using data collected well before the outcome of interest took place.” This is a methodological strength of our approach, which we have now applied to new data sources, but does not imply that this is the first time we have published using this approach. We have strengthened the next bullet point by adding “This study examined data from multiple sources…”

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

23. PAGE 5, LINES 38 TO 40: THE SENTENCE IS ABOUT PROTECTIVE FACTORS FOR AHT. IT IS AT THE END OF A PARAGRAPH ABOUT HOME VISITING. THIS SENTENCE DOES NOT SEEM TO BELONG HERE. PLEASE REWRITE.
We agree. We have deleted this sentence and most of the preceding paragraph in the process of rewriting the introduction in response to this Reviewer's other suggestions (point 16).

24. PAGE 6, LINE 10: PLEASE CLARIFY WHO IS "OUR."
We agree. We have replaced "our ability" with "the ability of primary healthcare providers". We have done the same in the third-to-last paragraph of the discussion (previously the last paragraph) where the same phrase occurred.

25. PAGE 6, LINES 54 TO 56: PLEASE CLARIFY FOR THE READER THAT THE MINISTRY FOR CHILDREN RELATES TO ALL OF NEW ZEALAND AND THE SAME RELATED TO THE POLICE. FOR EXAMPLE, IN MY COMMUNITY POLICE DATA ARE KEPT BY TOWNS.
We agree. We have added the sentence: "Each agency serves the entire country and records their data in a national electronic database."

26. ALSO, THE NEXT SENTENCE ABOUT DECISIONS ON TYPES OF DATA IS CONFUSING. PLEASE CLARIFY WHAT IS MEANT HERE. DOES THIS SENTENCE REFER TO THE RESEARCHERS OR WHAT HAPPENED BEFORE THE RESEARCHERS OF THE STUDY WERE INVOLVED?
We agree. We have expanded this as follows: "Because these databases are not normally accessed by health professionals, we were unsure what data they could provide. Early in study design, investigators conferred with experienced statutory social workers and police officers to ensure we understood how their data were collected and structured. Decisions were then made as to which variables could be extracted from each database."

27. PAGE 7, LINE 33: IT WOULD BE HELPFUL TO THE READER TO PROVIDE A BRIEF SUMMARY OF THE DATA COLLECTION, EVEN THOUGH THAT HAS BEEN DESCRIBED PREVIOUSLY.
We agree. The original sentence read: "Full details of perinatal health data collection in 2011 and 2012 are provided elsewhere." We have now expanded the sentence as follows: "In 2011 and 2012, health data were collected retrospectively by study investigators directly from maternal and child perinatal records". Reference is provided to the original study in the Journal of Pediatrics

28. PAGE 7, LINE 56: PLEASE CLARIFY HOW FAR BACK PRIOR TO THE BABY'S BIRTH DATA WERE COLLECTED. FOR EXAMPLE, WAS THERE A FIXED TIME PRIOR TO THE CHILD'S BIRTH OR COULD DATA HAVE BEEN OBTAINED AS FAR BACK AS THE BIRTH OF THE MOTHER OR FATHER?
We agree. There was no fixed time prior to the child's birth. We have added an additional paragraph which applies to both databases and is therefore located further down the page than the line mentioned by the Reviewer: "For both databases, data were searched for as far back in time as they could be found. It therefore included data dating back to the birth of the parents, if electronically recorded."

29. PAGE 8, LINE 6: PLEASE CLARIFY IN THE TEXT THAT A NOTIFICATION (OR REPORT OF CONCERN) WAS A REPORT TO THE CHILD PROTECTIVE SERVICES AGENCY.
We agree. We have added the phrase "to the statutory child protective services agency". Because the use of this phrase might cause confusion with the term "child welfare services" previously used elsewhere in the manuscript, we have been through the entire manuscript and replaced the phrase "child welfare services" with "child protective services", wherever it occurred (including in the Title, the Tables and the footnotes to the Tables).

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

30. PAGE 9, LINE 17: THIS ANALYSIS SECTION REFERS TO THE CONDITIONAL LOGISTIC REGRESSION. PLEASE USE A SENTENCE OR TWO TO EXPLAIN THIS TO THE READER.
We agree. We have added the following sentence: “This is a logistic regression in which each case was matched with their specified controls by date and hospital of birth.”

31. ALSO, WHY IS IT NECESSARY TO PRESENT BOTH THE UNCONDITIONAL AND CONDITIONAL RESULTS IN TABLE 2? IF BOTH ARE NECESSARY, PLEASE EXPLAIN THE PURPOSE TO THE READER.
The decision to present both sets of results reflects our response to current debate about which type of analysis (matched or unmatched) is most appropriate in matched case-control studies. The reference provided (Pearce, BMJ, 2016) goes into this question in considerable detail. We had attempted to summarise this in one sentence in the last paragraph of the Methods section (“To test validity and statistical precision the final model was analysed with both conditional and unconditional logistic regression”). It seems clear that this sentence was too concise to be helpful.

We have therefore deleted that sentence and considerably expanded the last paragraph of the Methods section. Everything in that last paragraph is new, except for parts of the last three sentences which introduce the concept of the AUC. The last three sentences now also explain the concept of the AUC, as requested by this reviewer (point 41). The paragraph now reads as follows:

“The case control design enabled us to control for potential confounders such as age and community characteristics. However, matching in the design can introduce confounding in the analysis. We analysed the data using both matched (conditional) and unmatched (unconditional) logistic regression. Results consistent across both methods are more likely to be robust. Both are provided in the tables so readers can judge for themselves. Also, it is useful to describe the performance of a predictive model by the area under the Receiver Operator Characteristic Curve (ROC). The ROC plots sensitivity against specificity across the entire distribution of the two populations (cases and controls). The area under the curve (AUC) is one measure of how well the model distinguishes between the two populations. However, a ROC can only be determined from the results of unconditional logistic regression.”

32. PAGE 9, LINE 24 TO PAGE 10, LINE 13: THIS SECTION DESCRIBES HOW THE MODEL WAS BUILT IN THE ACTUAL RESULTS, AND I THINK IT SHOULD BE MOVED TO THE RESULTS SECTION.
We agree. We did include some results data in this section (the numbers of the variables in each step of the process, and the paragraph describing the nine variables from our previous study). We have now deleted all those numbers and moved the paragraph describing the nine perinatal variables into the Results section. However, with those changes made, what remains is three short paragraphs describing the statistical method (multivariable analysis of two groups, combining groups and finally adding the variables which survived this elimination process into our original model), and those three paragraphs should therefore remain in the Methods section.

We have extensively rewritten the Results section to make allowance for these changes. The second sentence of the Results section has been deleted. There is a new first section entitled “Data from health records” which contains the paragraph moved from the Methods section. The section entitled “Data from the police” has been moved up, one sentence deleted and another added to give the number (one) and name of the variable which remained significant after group multivariable analysis (the number was previously in the Methods section). The section now entitled “Data from child protective services” has been reorganised to follow more clearly the sequence of the analysis outlined in the Methods section (including the numbers). We have added a new section entitled “Combined analysis of child protective services and police data” and moved into this section the result of the

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

combined analysis (including the numbers). This was previously included in the section on child protective services data

33. PAGE 11, LINE 45: THE TEXT SAYS “NOTIFICATION TO CHILD WELFARE,” BUT TABLE 2 SAYS “REPORT OF CONCERN.” PLEASE USE THE SAME LANGUAGE IN THE TEXT AND THE TABLE SO THAT THE READER DOES NOT GET CONFUSED.
We agree and add that the same problem occurred in what is now Table 1. We have replaced Report of Concern with notification wherever it occurred, and (as noted above) have also replaced all mention of “child welfare” with “child protective services” (or in one footnote, youth justice). To avoid confusion, we have also deleted the phrase “further action required”, which is the terminology used by the statutory child protective services in New Zealand for an investigation. This appeared once in each Table and once in the text, so we have replaced it with “investigation.”

34. PAGE 11, LINE 47: WHY IS THE UNCONDITIONAL ODDS RATIO PRESENTED HERE AS OPPOSED TO THE CONDITIONAL ONE? PLEASE CLARIFY FOR THE READER.
We agree. We originally chose to present the unconditional OR for simplicity, and because these were generally of lower magnitude and had greater statistical precision (narrower CI). However, in the light of our expanded explanation in the Methods section (see point 36), this simplification is unnecessary and goes against the principle of allowing the reader to judge for themselves. We have therefore deleted the OR from the text entirely and referred the reader to Table 2.

35. PAGE 12, LINES 12 TO 26: MUCH OF THIS TEXT CAN BE ELIMINATED OR SHORTENED SINCE THE INFORMATION IS IN THE TABLE.
We agree, now that Table 1 is in the manuscript and no longer a supplementary online Table. We have eliminated all the data from the text (the whole paragraph after the first sentence).

36. PAGE 12, LINE 54: THE AUTHORS PRESENT THE AREA UNDER THE CURVE. IT WOULD HELP THE READER IF IN THE METHODS, THE AUTHORS BRIEFLY EXPLAIN THIS.
We agree. We have now done this in the last paragraph of the Methods section, see the new paragraph set out in full in point 31 above.

37. ALSO, WHY WAS SENSITIVITY AND SPECIFICITY NOT PROVIDED HERE?
We did not provide these, because we did not mean to suggest that our model was good enough to be used as a tool to guide clinical interventions in specific families with specific combinations of variables. As we wrote in our previous paper: “Despite the high AUC and the fact that our model has significant predictive value, we do not suggest that our study provides an adequate framework (e.g., appropriate sensitivity and specificity) for predicting risk in individual families.”

We remain of this view.

We have now clearly set out the limitations of our model in the discussion, in two places.

Firstly, under limitations: “Finally, although our model can discriminate between two populations (as shown by the AUC), it explains only one third of the variation between those populations (as shown by the pseudo R squared statistic). Statistical associations between AHT and a variable do not necessarily mean that the variable will identify individuals more or less likely to experience AHT. Associations with extremely high OR may still turn out to be meaningless in individual-level prediction.”

Secondly, in our last paragraph: “Finally, while our study demonstrates that health data have the potential to be useful in predicting the risk of AHT, our current model is not good enough to guide clinical practice or policy. Neither the model, nor any of the variables it contains, can be used to guide

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

clinical interventions in specific families. More work is needed to replicate our findings, investigate other potentially relevant variables and examine possible confounders which may underlie or explain some of the variables in our model."

Please see also our response to Reviewer 2 point 5, which addresses a similar issue.

38. PAGE 15: THE AUTHORS COMPARE THEIR RESULTS TO OTHER NEW ZEALAND STUDIES ON PREDICTIVE RISK MODELING. THIS IS DONE IN THE CONCLUDING PARAGRAPH. IT WOULD HELP THE READER IF THIS WERE DONE IN THE BODY OF THE DISCUSSION. MORE INFORMATION COULD BE PROVIDED ABOUT SOME OF THE DIFFERENCES BETWEEN THIS STUDY AND THE PREVIOUS STUDIES; THIS INFORMATION WOULD BE HELPFUL TO THE READER.
We agree. Firstly, in response to the reviewer's suggestion, we have now provided a more comprehensive introduction to the New Zealand study (Vaithianathan et al) in a new fourth paragraph of the introduction (see point 16, above). Secondly, as the reviewer has requested, in the discussion we have moved the comparison to the Vaithianathan study out of the concluding paragraph. We agree that it properly belongs in the body of the discussion, probably in that section which the BMJ Open "Instructions to Authors" recommends should focus on "strengths and weaknesses in relation to other studies, discussing important differences in results".

The "Instructions to Authors" recommend that section should consist of only one paragraph. However, we accept the reviewer's point that more information about the differences between this study and that study would be helpful to the reader, and as a result we have felt obliged to add a second paragraph to the existing section, enlarging the discussion to seven paragraphs in total.

That new paragraph reads as follows. The first sentence is transferred from the conclusion, the rest is new information: "Our findings stand in contrast to the research mentioned in our introduction, which argued for the value of risk modelling for child abuse using data from the public benefit system and child protective services. Despite access to large amounts of data, that research had serious limitations. These included the assumption that "substantiation" is a valid outcome variable; the risk of bias inherent in the exclusion of families outside the public benefit system; the potential for breach of privacy and stigmatization without evidence for benefit; and the possibility of unintended consequences if their model was used to allocate interventions by influencing or overriding frontline clinicians. In addition, their final model included 132 separate variables, many of which did not differ significantly between cases and controls. Our study used a much more tightly-defined outcome variable, excluded no sector of the population, was more parsimonious (achieving a higher AUC with fewer variables) and ended by reinforcing the value of data already routinely collected by health professionals. The interpretation of such data by health professionals to guide health interventions would involve no breach of privacy."

39. NO PARAGRAPH WAS PROVIDING ABOUT LIMITATIONS. THIS NEEDS TO BE INCLUDED IN THE DISCUSSION.
We agree with the reviewer that the limitations of our study should be openly discussed, but we are constrained in our approach to this by the "Instructions to Authors" of the BMJ Open. These specify that "We also recommend, but do not insist, that the discussion section is no longer than five paragraphs and follows this overall structure (you do not need to use these as subheadings): (1) a statement of the principal findings; (2) strengths and weaknesses of the study; (3) strengths and weaknesses in relation to other studies, discussing important differences in results; (4) the meaning of the study: possible explanations and implications for clinicians and policymakers; and (5) unanswered questions and future research." We have followed this approach.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

We have now expanded and made more explicit our discussion of our limitations in the second paragraph, while retaining the sequence recommended by the BMJ Open, as follows:
Previously, the second and third sentences read as follows: “The principal weakness is that the study was retrospective. It is also important to note that data from child welfare and the police may not be sensitive indicators of risk.”
We have replaced the word “weakness” with “limitation” and added discussion of additional limitations. The second to fourth sentences of the second paragraph of the discussion now read as follows:
“There are several limitations. First, it was retrospective, so we had no control over the quality of the data. Second, data from child protective services and police may not be sensitive indicators of risk.”
After explaining why those might not be sensitive indicators of risk, we have added a further three sentences concerning the limitations of our predictive model, as requested by the second reviewer:
“Finally, although our model can discriminate between two populations (as suggested by the AUC), it explains only one third of the variation between those populations (as suggested by the pseudo R squared statistic). Statistical associations between AHT and a variable do not necessarily mean that the variable will identify individuals more or less likely to experience AHT. Factors associated with very high ORs may still turn out to be unhelpful as individual-level predictors.”

After reflecting on the sequence of discussion recommended in the “Instructions to Authors”, we have also swapped two paragraphs in the discussion. The last paragraph was focused on “the meaning of the study: possible explanations and implications for clinicians and policymakers” and the second-to last paragraph was focused on “unanswered questions and future research”. We have now swapped them over, so the sequence sits better with the Instructions to Authors, and added a new final paragraph which again emphasises the limitations of our study:
“Finally, while our study demonstrates that health data have the potential to be useful in predicting the risk of AHT, our current model is not good enough to guide clinical practice or policy. Neither the model, nor any of the variables it contains, can be used to guide clinical interventions in specific families. More work is needed to replicate our findings, investigate other potentially relevant variables and examine possible confounders which may underlie or explain some of the variables in our model.”

40. THE STROBE CHECKLIST APPEARS TO BE PROVIDED IN TABLE 1. THERE IS NO MENTION IN THE TEXT ABOUT THIS, AND THE TABLE IS NOT LABELED AS TABLE 1. UNLESS REQUIRED BY THE JOURNAL, I THINK THIS COULD BE A SUPPLEMENTAL TABLE.
We agree. We uploaded the STROBE checklist as requested by the journal and noted that it appeared in the pdf proof, but we do not believe it needs to be part of the paper or even a supplemental table (unless required by the journal). We have uploaded it again because the page numbers in the revised manuscript have changed, but we have made no other changes to it

VERSION 2 – REVIEW

REVIEWER	Laura Schummers University of British Columbia, Canada
REVIEW RETURNED	04-Jan-2019

GENERAL COMMENTS	The authors have a done a great job of substantially revising this manuscript to thoughtfully take the many reviewer comments into account. Specifically, the revised discussion section appropriately describes the limitations of using this prediction model to guide clinical practice at an individual level. The careful discussion of these limitations is laudable.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

	I have one methodological note to add, though suggest that the authors and editors do not take this as a call for additional analyses for this paper at this time. Stepwise variable selection processes (forward/backward) is not the ideal variable selection strategy, as this can introduce bias in the parameter estimates and model performance. For example, see Steyerberg's discussion bias introduced by stepwise variable selection strategies such as stepwise selection in Chapter 5 (Steyerberg EW. Clinical prediction models: A practical approach to development, validation, and updating. New York: Springer; 2009.) A full model approach (including all variables that were likely predictors a priori) is the preferred approach. However, given that the take-home message of this paper is that the model should not be used to predict on an individual level, this change to the methodology is not necessary. Also, given the authors' substantial revisions already undertaken in the revision process, I am not suggesting this change at this point, but making this suggestion for future prediction modelling work. I applaud the authors for their careful work on this paper and recommend publication in its current state.

REVIEWER	John M. Leventhal Yale Medical School, USA
REVIEW RETURNED	13-Jan-2019

GENERAL COMMENTS	This manuscript is much improved. I have a few suggestions: 1. Table1. It would help the reader if the text said that the main results are in Table 1 as opposed to beginning with birth certificate data. Also, it would be helpful to the reader to have the variables described in the text in the same order as in the table. In T able 1, child protective services data are before police data but in the text they are in the reverse order. Please correct this. The first percentage under Cases is missing related to father on the birth certificate. 2. Page 14, line 19: "counting the data" is not clear. Please re-write: perhaps "counting the number of events" or something like that might be clearer. 3. Page 14, line 31: what does "group multivariable analysis" mean? Is this described in the methods? Please clarify in the text either here or in the methods or both to help the reader. 4. Table 2: the percentages in this table are to one decimal but in Table 1 the percentages are whole numbers. Why the difference? Please make both tables the same. 5. Page 16, line 46: please clarify what is meant by "previous model." 6. Page 18, line 52: in addition to reference 12, the same authors published an article in Pediatrics in 2018 related to risk modeling that is also relevant and perhaps deserves a few sentences in this paragraph as well. 7. Page 19, line 33: I am confused about the words "good reason." Why is there a good reason to believe that your results are generalizable to other countries? Please clarify briefly in the text. 8. Page 20, line 17: please clarify in the text what is meant by "it." 9. Page 20, line 40: it would be helpful to have a brief concluding paragraph with the take-home message

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

VERSION 2 – AUTHOR RESPONSE

REVIEWER: 2 REVIEWER NAME: LAURA SCHUMMERS

INSTITUTION AND COUNTRY: UNIVERSITY OF BRITISH COLUMBIA, CANADA

1. THE AUTHORS HAVE A DONE A GREAT JOB OF SUBSTANTIALLY REVISING THIS MANUSCRIPT TO THOUGHTFULLY TAKE THE MANY REVIEWER COMMENTS INTO ACCOUNT. SPECIFICALLY, THE REVISED DISCUSSION SECTION APPROPRIATELY DESCRIBES THE LIMITATIONS OF USING THIS PREDICTION MODEL TO GUIDE CLINICAL PRACTICE AT AN INDIVIDUAL LEVEL. THE CAREFUL DISCUSSION OF THESE LIMITATIONS IS LAUDABLE.

We thank the Reviewer for these positive remarks about our revised manuscript.

2. I HAVE ONE METHODOLOGICAL NOTE TO ADD, THOUGH SUGGEST THAT THE AUTHORS AND EDITORS DO NOT TAKE THIS AS A CALL FOR ADDITIONAL ANALYSES FOR THIS PAPER AT THIS TIME. STEPWISE VARIABLE SELECTION PROCESSES (FORWARD/BACKWARD) IS NOT THE IDEAL VARIABLE SELECTION STRATEGY, AS THIS CAN INTRODUCE BIAS IN THE PARAMETER ESTIMATES AND MODEL PERFORMANCE. FOR EXAMPLE, SEE STEYERBERG'S DISCUSSION BIAS INTRODUCED BY STEPWISE VARIABLE SELECTION STRATEGIES SUCH AS STEPWISE SELECTION IN CHAPTER 5 (STEYERBERG EW. CLINICAL PREDICTION MODELS: A PRACTICAL APPROACH TO DEVELOPMENT, VALIDATION, AND UPDATING. NEW YORK: SPRINGER; 2009.) A FULL MODEL APPROACH (INCLUDING ALL VARIABLES THAT WERE LIKELY PREDICTORS A PRIORI) IS THE PREFERRED APPROACH. HOWEVER, GIVEN THAT THE TAKE-HOME MESSAGE OF THIS PAPER IS THAT THE MODEL SHOULD NOT BE USED TO PREDICT ON AN INDIVIDUAL LEVEL, THIS CHANGE TO THE METHODOLOGY IS NOT NECESSARY. ALSO, GIVEN THE AUTHORS' SUBSTANTIAL REVISIONS ALREADY UNDERTAKEN IN THE REVISION PROCESS, I AM NOT SUGGESTING THIS CHANGE AT THIS POINT, BUT MAKING THIS SUGGESTION FOR FUTURE PREDICTION MODELLING WORK.

We thank the Reviewer for her thoughtful suggestion and will take this suggestion into consideration for any future prediction modelling work. We have not made any changes to the current manuscript, as the Reviewer is not suggesting this to be necessary.

3. I APPLAUD THE AUTHORS FOR THEIR CAREFUL WORK ON THIS PAPER AND RECOMMEND PUBLICATION IN ITS CURRENT STATE.

We thank the Reviewer for her positive review of this revised manuscript, and for her earlier review which helped us to produce a better manuscript.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

REVIEWER: 3 REVIEWER NAME: JOHN M. LEVENTHAL

INSTITUTION AND COUNTRY: YALE SCHOOL OF MEDICINE, USA

4. THIS MANUSCRIPT IS MUCH IMPROVED.

We thank the reviewer for this positive assessment, and for his previous and current suggestions which have helped us to produce a better manuscript.

5. TABLE1. IT WOULD HELP THE READER IF THE TEXT SAID THAT THE MAIN RESULTS ARE IN TABLE 1 AS OPPOSED TO BEGINNING WITH BIRTH CERTIFICATE DATA.

We agree. We have added the following sentence as the third sentence in the results section: “The results for variables from the Birth Certificate, child protective services and the police included in logistic regression are presented in Table 1.”

6. ALSO, IT WOULD BE HELPFUL TO THE READER TO HAVE THE VARIABLES DESCRIBED IN THE TEXT IN THE SAME ORDER AS IN THE TABLE. IN TABLE 1, CHILD PROTECTIVE SERVICES DATA ARE BEFORE POLICE DATA BUT IN THE TEXT THEY ARE IN THE REVERSE ORDER. PLEASE CORRECT THIS.

We agree. We have corrected the sequence in the text to follow the sequence in the Table, which follows the sequence in which data were collected (as already described in the Methods section).

7. THE FIRST PERCENTAGE UNDER CASES IS MISSING RELATED TO FATHER ON THE BIRTH CERTIFICATE.

We agree and thank the reviewer for spotting this. It was in the Table but had disappeared from sight because the column was slightly too narrow. We have corrected this.

8. PAGE 14, LINE 19: “COUNTING THE DATA” IS NOT CLEAR. PLEASE RE-WRITE: PERHAPS “COUNTING THE NUMBER OF EVENTS” OR SOMETHING LIKE THAT MIGHT BE CLEARER.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

We agree. We have changed “counting the data” to “counting the number of events” in two places: both on line 19 (data from child protective services) and line 8 (data from the police). These two sections have been transposed in response to suggestion (6), so the line numbers have changed.

9. PAGE 14, LINE 31: WHAT DOES “GROUP MULTIVARIABLE ANALYSIS” MEAN? IS THIS DESCRIBED IN THE METHODS? PLEASE CLARIFY IN THE TEXT EITHER HERE OR IN THE METHODS OR BOTH TO HELP THE READER.

We agree. This is described in the Methods already (the third paragraph under statistical analysis), but we have now clarified this phrase in the two places it occurs in the text of the Results section. Firstly, in the location identified by the Reviewer (p14, line 31), we have replaced the phrase “in group multivariable analysis” with the phrase “In multivariable analysis of these 14 variables from child protective services…” Secondly, in the section concerning police data (p14, line 10), we have replaced the phrase “in group multivariable analysis” with the phrase: “In multivariable analysis of these 5 variables from the police…” These two sections have been transposed in response to suggestion (6), so the line numbers have changed.

10. TABLE 2: THE PERCENTAGES IN THIS TABLE ARE TO ONE DECIMAL BUT IN TABLE 1 THE PERCENTAGES ARE WHOLE NUMBERS. WHY THE DIFFERENCE? PLEASE MAKE BOTH TABLES THE SAME.

We agree. We have made the percentages in both Tables the same (to one decimal)

11. PAGE 16, LINE 46: PLEASE CLARIFY WHAT IS MEANT BY “PREVIOUS MODEL.”

We agree. We have clarified this by expanding “The AUC of the previous model” to “The AUC of the model derived from health data alone”

12. PAGE 18, LINE 52: IN ADDITION TO REFERENCE 12, THE SAME AUTHORS PUBLISHED AN ARTICLE IN PEDIATRICS IN 2018 RELATED TO RISK MODELING THAT IS ALSO RELEVANT AND PERHAPS DESERVES A FEW SENTENCES IN THIS PARAGRAPH AS WELL.

We were aware of that article (Vaithianathan R, Rouland B, Putnam-Horstein E. Injury and Mortality Among Children Identified as at High Risk of Maltreatment. Pediatrics. 2018 Feb;141(2).doi: 10.1542/peds.2017-2882) and considered referring to it. The authors took the criteria which they developed to predict substantiated child maltreatment and showed that children predicted (by those criteria) to be at high risk of child maltreatment are also at high risk of hospitalisation and death for other reasons. However, although that is an interesting finding, our study provides no new data which would qualify us to comment on it. We would be able to comment on that article if we also had studied

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

whether our model (derived from health data alone) similarly predicted a higher risk of hospitalisation and death for reasons other than AHT, but we did not study that question.

13. PAGE 19, LINE 33: I AM CONFUSED ABOUT THE WORDS “GOOD REASON.” WHY IS THERE A GOOD REASON TO BELIEVE THAT YOUR RESULTS ARE GENERALIZABLE TO OTHER COUNTRIES? PLEASE CLARIFY BRIEFLY IN THE TEXT.

We agree. We have removed the phrase “good reason” and added an additional phrase (“many variables…AHT”) and the word “internationally”, along with appropriate references. Our reasoning is now clarified briefly in the text as follows: “We suggest that our findings will be generalizable to other countries. Our method was robust, many variables in our model are consistent with other literature on risk factors for AHT, and the limitations of police and child protective services data (described above) are well-recognised internationally.”

14. PAGE 20, LINE 17: PLEASE CLARIFY IN THE TEXT WHAT IS MEANT BY “IT.”

We agree. We have deleted “It” as the first word of the sentence and clarified the sentence by replacing that pronoun with a phrase referencing the previous sentence: “The absence of such a relationship.” The sentence now reads: “The absence of such a relationship may suggest that these families at risk are overlooked by both health professionals and statutory authorities.”

15. PAGE 20, LINE 40: IT WOULD BE HELPFUL TO HAVE A BRIEF CONCLUDING PARAGRAPH WITH THE TAKE-HOME MESSAGE.

We agree. Our take-home message is that adding data from child protective services and the police did not improve the predictive accuracy of data from perinatal records, and therefore that “if the quality and consistency of perinatal health care could be improved, it is the health system which may hold the key to identifying families most likely to benefit from early intervention.” This take-home message was already included in the third-to-last paragraph. We had placed it there in an attempt to follow the recommendations of the BMJ Open for the sequence of the discussion.

However, we accept the reviewer’s point that a concluding paragraph with the take-home message would be helpful to many readers.

In our view, the most efficient way to do this is to make the third-to-last paragraph the final paragraph. We have now achieved this as follows. Firstly, by moving what was previously the last paragraph (emphasising that “our current model is not good enough to guide clinical practice or policy”) into the paragraph on limitations. It fits very smoothly there, and we lose none of the clarified text concerning the limitations of our methods which was praised by Reviewer Two. Secondly, by moving what was

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

the third-to-last paragraph to the end of the manuscript, we respond to the current comment by having a more effective ‘concluding paragraph.’

The concluding paragraph now reads: “This study tested the hypothesis that combining information from perinatal health records with information from child protective services and the police would enhance the ability of primary healthcare providers to predict (and therefore possibly prevent) AHT. We found little evidence to support that hypothesis. We suggest that our findings will be generalizable to other countries. Our method was robust, many variables in our model are consistent with other literature on risk factors for AHT, and the limitations of police and child protective services data (described above) are well-recognised internationally. With respect to AHT, we suggest that if the quality and consistency of perinatal health care could be improved, it is the health system which may hold the key to identifying families most likely to benefit from early intervention. Such improvements could include: routine and universal enquiry during pregnancy for matters of possible relevance such as alcohol and drug abuse, IPV, unplanned pregnancy and untreated mental illness; routine and universal follow-up of families with missing data or poor engagement with antenatal care; and routine and universal access to evidence-based early intervention programmes when matters of concern are identified by health providers.”

This paragraph is not in itself “brief”, but it has the advantage that (except for the modifications made in response to point 13) it adds no additional text to the manuscript. Although it is not quite consistent with the recommendations of the BMJ Open for the sequence of the discussion, we hope that the editors will agree that this is the best solution in the circumstances.

BMJ Open: first published as 10.1136/bmjopen-2018-024199 on 1 March 2019. Downloaded from http://bmjopen.bmj.com/ on April 25, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.