PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Transcutaneous Bilirubinometry during and after Phototherapy in Preterm Infants, Prospective Observational Study
AUTHORS	Raba, Ali; O'Sullivan, Anne; Miletin, Jan

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Peter Flom Institution and Country: Peter Flom Consulting USA Competing interests: None
REVIEW RETURNED	17-Mar-2020

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. Unfortunately, there were some fairly major problems. The authors did t-tests to see if pairs of results were significantly different. But what would be more appropriate is test of equivalence, where the authors pick a difference that is small enough to not matter. This can be done in most major stat software. They can Google TOST to find some info. Also, they should do regression using "difference" as the dependent variable and various things as independent variables (age, weight, etc. from the bottom of p. 6). This would allow you to use a formula to get one variable from the other More specific comments: p. 2 line 40 The term "agreement limits" needs a definition. p. 6 line 47-49 Data or variables cannot be parametric or non-parametric, only models can be. Do you mean "normal"? p. 6 line 59-60 Please operationalize all variables. Also, you can't use a t-test with age or BW and the results for age and BW don't seem to be in the paper. The figures should be separated and should be much larger. p. 7 line 31 These are pretty clearly not normal as that would mean some values would be below 0.

REVIEWER	Reviewer name: Peter H Dijk Institution and Country: Beatrix Childrens Hospital, University Medical Center Groningen, University of Groningen, The Netherlands Competing interests: None
REVIEW RETURNED	30-Mar-2020

GENERAL COMMENTS	In this prospective observational study the authors aim to examine

	the accuracy of TcB measurement in a convenience sample of 196 preterm infants. TcB was measured covered and uncovered an compared to TSB values, in 299 paired samples during phototherapy and in 137 paired measurements after phototherapy. Accuracy is reported as correlations and B-A- agreement plots. They found significant correlations and agreements, but with a large range, especially the uncovered TcB measurements during phototherapy. The authors conclude that as result of the wide disagreement between TcB and TSB during PT TcB measurements cannot be recommended while TcB should be used post-PT. Although not new, these data are important to be presented. The study has been performed well and the results are presented clearly. A few details of the methodology as lacking. But that is not my major issue. I think, much more valuable information is in this research but not presented yet. And that information is essential to support their conclusion that TcB can be used after PT, but not during in preterm infants. In my opinion correlations and agreement plots are not statistically enough to draw such a conclusion without predefined limits. This research would gain far more value if calculations and statistics would be added that show in a quantitative manner how TcB measurements affects guidance in clinical decisions as starting, continuing and stopping PT in preterm infants. This means that statistical measures as sensitivity and specificity and PPV and NPV etc of TcB measurements related to TSB and their PT thresholds should be added to this research. Because this was a prospective study this information should be available. Recently a few other papers about the value of TcB measurements in preterms during and after PT have been published, that are not yet included in this manuscript (e.g. Hulzebos et al in PLOS 2019, Cucuy in the J.Matern Fetal Neonatal Med 2018, Bhargava in J.Perinat Med in 2019 and Arman in J of Perinatol in 2020). In some of these, but also other studies, the authors propose different rules how TcB values could be used as safe as possible in clinical guidance: e.g. adding 50 umol/l and/or using a 70% value of the TSB-PT-thresholds. This research would gain far more value by adding this statistical information and adding more recent evidence, and evaluating the rules that are proposed by those authors. Minor remarks. I do not understand the reason for the 3rd and 4th affiliations of the authors because this research was performed in Dublin only – and not in Praque (as far as can be read in the manuscript)? I suppose that the sample size is a convenience sample? Because there is not a clear hypothesis or sample size calculation in the manuscript. In the method section it was stated that PTH was commenced based on TSB levels according to the hospital guidelines… It would be valuable to show these in additional information or to give an reference. This information is essential for the readers is the authors do decided to add the statistical measures as sensitivity and specificity etc. In the methods section it is explained that the covered TcB measurement was on the upper outer quadrant of the buttock. In my experience it is not that easy to do a TcB on soft tissue. Therefor the manufacturers of TcB meters recommend to measure it on a bony part of the body. Could you please explain? The JM devices can be set how repeated measurements are handled. In this research three measurements were done – but is was not explained whether the highest, lowest or mean or median value was used for further analyses.

	TSB levels were measured using direct spectrophotometry – by what device? In 196 infants 328 simultaneous measurements where performed during phototherapy, but only in 137 after PT – why is this not at least 196? In my experience preterm infants have much more bilirubin measurements done and often need a second course of PT. Could you please explain. I think table 2 and 3 is not adding much information, because it is in the text too. The figures are very small to read. The discussion section should be updated with recent studies on TcB measurements during and after PT in preterm infants. And as I explained before – I think the final conclusions cannot be drawn without quantifying the potential clinical consequences of the TcB measurements on starting and stopping PT (sensitivity and specificity etc statistics related to rules that have been proposed by other researchers).

REVIEWER	Reviewer name: Christian Hulzebos Institution and Country: University Medical Center Groningen, The Neterlands Competing interests: None
REVIEW RETURNED	31-Mar-2020

GENERAL COMMENTS	In this manuscript Raba and colleagues describe the accuracy of transcutaneous bilirubin (TcB) measurements during (n=328) and after (n=137) phototherapy (PT) in 196 preterm infants at uncovered and covered skin. Correlations and Bland-Altman data are presented. TcB underestimates total serum bilirubin (TSB) obtained at covered and uncovered sites during phototherapy, and this underestimation persists after discontinuation of PT. This is an interesting manuscript on a topic that has clinical relevance. However, the manuscript raises some questions that need to be addressed. And the clinical relevance should be stresses much more. That is an important issue – please expand on this matter. General remarks There are numerous data of TcB measurements during PT (see PubMed). It is well known that PT bleaches the skin. As such TcB measurements on PT-exposed skin areas are not recommended, but there are few studies that state that TcB measurements provide accurate TSB estimations under PT (using a different cut-off rule). After PT there is inaccuracy as well up to a certain time point. This and what this manuscript adds should be mentioned in the Introduction and Discussion. The authors should be able to compare different TcB cut-off methods when to determine a TSB. In general TSB measurement is indicated when TcB+50 µmol/L exceeds the PT threshold. But a different rule may change their conclusion. Specific remarks The title is not very attractive. Abstract Please check grammar and punctations. It is not clear at which site TCB is measured under the nappy (abdominal or at the ileal bone?). Please use PT, not PTH, as abbreviation of phototherapy. The conclusion contains not much new information. Please provide clinical relevance of the data. Introduction

	The statement that aggressive PT needs to be avoided needs to be clarified including clinical data of published RCTs on that matter in ELBW preterm infants. TcB measurements are not recommended in the first 24 hours. This should be added in ll 42-47. Please provide an explanation why PT affects the relationship between TcB and TSB. Methods Why were infants with co-morbidity excluded? Please provide local PT guidelines. Please provide (measured) irradiance of PT devices. What was the agreement between the two TcB devices? Please provide background of the decision of the attending neonatologist to measure TSB. Please provide background of why it is important to define EOS. Ll 22: Our main outcome was to examine should be changed in "main aim…" or "primary outcome was the correlation….". The next sentence on secondary outcomes should be changes as well. Please provide rationale of the effect of clinical parameters on the relationship between TcB and TSB. Ll. How were non-parametric data tested? Please explain why the t-test was chosen to study influence of clinical parameters on TcB-TSB relationship. This seems not appropriate. Results Please provide time points of TcB and TSB measurements after discontinuation of PT. Figures seem appropriate although too small to read and compare. I suggest to remove correlations and only show B-A plots in one Figure (with an a, b, and c panel). Tables. I did not see 328 measurements, but 309 as maximum number. Please explain. Tables 2 and 3 should be in one Table. Table 4 and 5 should be changed after having applied other statistics. The effect of clinical parameters such as RDS seems not statistically appropriate tested. Discussion The Discussion is limited to few studies on the same topic, whereas there are lots of published data regarding the relationship between TcB and TSB. The authors speculate on the explanation of the underestimation of TSB by TcB. They mention immaturity of the skin and absence of subcutaneous fat. If true, then the relationship between TcB and TSB in term infants undergoing PT better? But what is the point of action of PT – intra or extravascular bilirubin? Please check grammar. For example in ll 29 on page 9: "were more.." – "were higher". P 10-ll10: results should not be repeated. Data should be discussed instead. Conclusion is clear about the correlation – but that's is not new. Clinical relevance, ie the authors' recommendation, should be stressed much earlier in the manuscript. However, there are also studies reporting that TcB could be used even under PT – maybe when using other cut-off rules. The authors mention one study, but there are other studies available. Regarding the post PT measurements: please provide after how many hours (6? 12?) TcB measurements are accurate again.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

We would like to thank Reviewer 1 for his comments and suggestions.

1. I confine my remarks to statistical aspects of this paper. Unfortunately, there were some fairly major problems.

The authors did t-tests to see if pairs of results were significantly different. But what would be more appropriate is test of equivalence, where the authors pick a difference that is small enough to not matter. This can be done in most major stat software. They can Google TOST to find some info.

Also, they should do regression using "difference" as the dependent variable and various things as independent variables (age, weight, etc. from the bottom of p. 6). This would allow you to use a formula to get one variable from the other

Response: We would respectfully disagree with the reviewer on the first point. We used paired-samples t-test (with the null hypothesis that the mean difference between paired observations is zero) as an alternative to the test of equivalence suggested and we would believe this would be commonly employed in this situation (NCSS.com – Chapter 519 - Paired T-Tests for Equivalence). Paired t-test is justified in our opinion for comparison of two different methods of measurement for the same biological parameter (Shier 2004, Mathematic Learning Support Centre, http://www.statstutor.ac.uk/ resources/ uploaded/ paired-t-test.pdf) . We believe that paired samples t-test is a justified alternative to two one-sided test (TOST) (Mara CA, Cribbie RA. Paired-Samples Tests of Equivalence. Communications in Statistics – Simulation and Computation, 41:1928-1943, 2012). We have amended our statement in the Methods to: 'Paired-samples t-test was used …' to avoid any confusion. However after some debate when the differences between pairs were not normally distributed, we have now employed Wilcoxon signed-rank test for two sample comparisons (McDonald JH, Handbook for Biological Statistics, 3rd Edition, http://www.biostathandbook.com/ index.html). We amended our Methods accordingly (please see Reviewer 3, Response 19)

We would like to thank the reviewer for the regression suggestion, we have decided to take this part of our work out of the manuscript.

More specific comments:

2. p. 2 line 40 The term "agreement limits" needs a definition.

Response: We would like to thank the reviewer for this helpful suggestion. We added a definition of agreement limits to the Methods part of manuscript – 'The agreement limits are demonstrated as a 95% confidence interval (95% CI = mean ± 1.96 standard deviations), where the ideal agreement difference between measurements is zero.'

3. p. 6 line 47-49 Data or variables cannot be parametric or non-parametric, only models can be. Do you mean "normal"?

Response: We have amended our statement in the Methods part as suggested by the reviewer: 'Mean and standard deviation (SD) were used for normally distributed data, while non normal distribution data was summarised using median and interquartile range (IQR).'

4. p. 6 line 59-60 Please operationalize all variables. Also, you can't use a t-test with age or BW and the results for age and BW don't seem to be in the paper.

Response: We agree with the reviewer. We have decided to remove this part of our work from the manuscript as it was not adding any new information and we have to admit it was of very limited relevance. This statement was taken from the Methods part of the Manuscript – 'We used t-test to study the influence of gestational age, birth weight, sepsis, RDS, PDA and ABO incompatibility on the difference between TSB and TCB during and post-phototherapy.' We have also removed the statement from the results – 'We determined that gestational age, birth weight, sepsis, RDS, PDA and ABO incompatibility had no influence on the mean difference between the TSB and TCB (TCBU and TCBC) readings during the PT (Table 4). We found also that the difference between the TSB and TCB measurements after the phototherapy were not affected by gestational age, birth weight, PDA, ABO incompatibility. However, infants with a diagnosis of RDS had a statistically significant reduction in the difference between TCB and TSB, in comparison to those infants without RDS, after cessation the PT (Table 5).' As the result of this, Table 4 and Table 5 have also been removed.

5. The figures should be separated and should be much larger.

Response: We would believe that the reviewer did not have our original Figures at hand, but rather the conversion to PDF format (done automatically at submission by the software and that would inevitably change the look of the Figures – most likely done at lower DPI and scaled to fit the A4 page – portrait orientation). Our raw Figures (all three of them) are 297 mm x 210 mm large (landscape A4 format) with 600 or 300 DPI (Figure 3) resolution and 7016 x 4961 (Figure 3 - 3508 x 2480) pixels dimension. We would believe that these characteristics are very appropriate for quality and reproduction purposes. However, we have now removed any compression from the JPG files to enhance the quality and we have made Figure 3 600 DPI 7016 x 4961 pixels to be consistent.

6. p. 7 line 31 These are pretty clearly not normal as that would mean some values would be below 0.

Response: We would like to thank the reviewer for this comment. We are aware that some values (TSB measurements during and post phototherapy) are not normally distributed. However, we decided to present these values as mean ±SD for two reasons. Firstly, some of these values are not normally distributed (TSB during and after phototherapy) and others (TCBC and TCBU during and post phototherapy) are normally distributed, therefore it would be easier to use mean ±SD throughout the study to compare the results instead of using mean and median. Secondly, this reporting gave us opportunity to compare our results to previous studies (for example Fonseca et al, Journal of Perinatology 2012, Grabenhenrich et, Pediatrics 2014, Hulzebos et al in PLOS 2019) that have also used mean ±SD for non normal distribution of TSB and TCB values. However, we added medians and interquartile ranges to newly created Table 2 (merged Table 2 and 3) for more clarity, in addition to mean ±SD.

Reviewer: 2

We would like to thank Reviewer 2 for his kind comments and suggestions.

In this prospective observational study the authors aim to examine the accuracy of TcB measurement in a convenience sample of 196 preterm infants. TcB was measured covered and uncovered an compared to TSB values, in 299 paired samples during phototherapy and in 137 paired measurements after phototherapy. Accuracy is reported as correlations and B-A- agreement plots. They found significant correlations and agreements, but with a large range, especially the uncovered TcB measurements during phototherapy. The authors conclude that as result of the wide disagreement between TcB and TSB during PT TcB measurements cannot be recommended while TcB should be used post-PT.

1. Although not new, these data are important to be presented. The study has been performed well and the results are presented clearly. A few details of the methodology as lacking. But that is not my major issue. I think, much more valuable information is in this research but not presented yet. And that information is essential to support their conclusion that TcB can be used after PT, but not during in preterm infants. In my opinion correlations and agreement plots are not statistically enough to draw such a conclusion without predefined limits. This research would gain far more value if calculations and statistics would be added that show in a quantitative manner how TcB measurements affects guidance in clinical decisions as starting, continuing and stopping PT in preterm infants. This means that statistical measures as sensitivity and specificity and PPV and NPV etc of TcB measurements related to TSB and their PT thresholds should be added to this research. Because this was a prospective study this information should be available.

Response: We would like to thank the reviewer for his comments and suggestions. Although we might agree to some extent that adding screening/predictive measures (sensitivity, specificity, PPV and NPV of TCB measurements) could be theoretically helpful, we believe that these measures would not be practical for our dataset and work for couple of reasons. Firstly, there are wide ranges of PT thresholds (cut off levels), according to the gestation (weeks) and age (time at TCB measurements). We have used National Institute for Health and Care Excellence (NICE) phototherapy guidelines (newly added Appendix 1) for infants below 32 weeks of gestation and adapted National Health Service (NHS), Glasgow, UK phototherapy charts for infants ≥32 weeks of gestation (newly added Appendix 2) in the study. Because of the increasing slope of the PT thresholds during the first 72 to 120 hours after birth and different chart for every week of gestation from 23 to 31 weeks of gestation, it would be impossible to define exact cut-off points for phototherapy. Furthermore, most of our measurements (up to 70%) happened within the first 72 hours of life. Therefore, to get accurate measures (Sensitivity, Specificity, PPV and NPV), it should include different cut off values for each gestation and only for infants after 72/120 hours of age (similarly Hulzebos et et al., PLoS One, 2019 defining only thresholds after 48 hours of age and thus avoiding increasing slope in Dutch charts). To do these calculations for our cohort, it would require a very large sample of infants in each gestational category (by week) after 72/120 hours of age and as such would not be feasible. Secondly, the aim of our study was to investigate the correlation and agreement between TSB and TCB (during and after PT), and we would rather try to avoid any unplanned post hoc analysis.

2. Recently a few other papers about the value of TcB measurements in preterms during and after PT have been published, that are not yet included in this manuscript (e.g. Hulzebos et al in PLOS 2019, Cucuy in the J.Matern Fetal Neonatal Med 2018, Bhargava et al in J.Perinat Med in 2019 and Arman in J of Perinatol in 2020).

Response: We would like to thank Reviewer 2 for this suggestion. We have already discussed the study conducted by Cucuy et al. We also updated our manuscript and we included suggested studies in our discussion (Hulzebos et al in PLOS 2019 , Arman in J of Perinatol in 2020).

3. In some of these, but also other studies, the authors propose different rules how TcB values could be used as safe as possible in clinical guidance: e.g. adding 50 umol/l and/or using a 70% value of the TSB-PT-thresholds. This research would gain far more value by adding this statistical information and adding more recent evidence, and evaluating the rules that are proposed by those authors.

Response: As outlined in our Response 1 (Reviewer 2), we don't feel this would be feasible for our dataset.

Minor remarks.

4. I do not understand the reason for the 3rd and 4th affiliations of the authors because this research was performed in Dublin only – and not in Prague (as far as can be read in the manuscript)?

Response: The study was performed in Coombe Women and Infants University Hospital, Dublin, Ireland as outlined in Methods part of the manuscript. However authors has more than one affiliation (including academic and/or hospital affiliations in Prague, Czech Republic). We would believe this is not unusual and in fact very important to include for clarity, transparency and potential conflict of interest issues.

5. I suppose that the sample size is a convenience sample? Because there is not a clear hypothesis or sample size calculation in the manuscript.

Answer: We would like to thank reviewer for his question. We defined our outcomes and we have used convenience sampling. We anticipated that our sample size would be large enough to produce significant results, in comparison to previous studies. We have added short statement at the end of Methods: 'We have used a convenience sample for the study with planned one year enrolment.'

6. In the method section it was stated that PTH was commenced based on TSB levels according to the hospital guidelines… It would be valuable to show these in additional information or to give an reference. This information is essential for the readers is the authors do decided to add the statistical measures as sensitivity and specificity etc.

Response: We would like to thank the reviewer for this valuable observation and we added statements regarding phototherapy guidelines in the Methods part of the manuscript and we have also included phototherapy charts as Appendix 1 and 2: 'The NICE treatment charts were used for preterm infants below 32 weeks of gestation (Appendix 1). In infants ≥32 weeks of gestation a chart adapted from the National Health Service (NHS), Glasgow, UK was used (Appendix 2).'

7. In the methods section it is explained that the covered TcB measurement was on the upper outer quadrant of the buttock. In my experience it is not that easy to do a TcB on soft tissue. Therefor the manufacturers of TcB meters recommend to measure it on a bony part of the body. Could you please explain?

Response: TCB was measured from the bony part of upper outer quadrant of the buttock. We amended our statement in the Methods part of the manuscript: 'The process was repeated over the covered area, the bony part of the upper outer quadrant of the buttock (covered by the nappy).'

8. The JM devices can be set how repeated measurements are handled. In this research three measurements were done – but is was not explained whether the highest, lowest or mean or median value was used for further analyses.

Response: The devices automatically calculated the average of the three measured values and this number then appeared on the display. We have amended our statement in the Methods accordingly - 'The device was placed over an uncovered area (sternum) and pressed gently against the skin three times to obtain one reading (the average of the three measured values).'

9. TSB levels were measured using direct spectrophotometry – by what device

Response: Analyser added to the Methods part of the manuscript: '(Abbott Architect C8000, Abbott, USA)'

10. In 196 infants 328 simultaneous measurements where performed during phototherapy, but only in 137 after PT – why is this not at least 196? In my experience preterm infants have much more bilirubin measurements done and often need a second course of PT. Could you please explain.

Response: We would like to thank the reviewer for this comment. We sought to obtain paired TSB and TCB measurement post-phototherapy from every baby. Unfortunately, we included only 142 (there is a typo in the original manuscript, we corrected the same) paired measurements, as a number

of measurements were excluded because: 1) the time between TSB and TCB was more than predefined. 2) a few babies had unpaired measurements (TSB without TCB or TCB without TSB).

11. I think table 2 and 3 is not adding much information, because it is in the text too.

Response: We agree with the Reviewer 2. However, in response to Reviewer 1 and Reviewer 3, we merged the two Tables into one Table (now Table 2) and added more information (median (IQR)) not presented in the text. We have also rounded the results to whole numbers (as the TSB and TCB measurements are without decimal points)

12. The figures are very small to read.

Response: Please see Reviewer 1 Response 5.

13. The discussion section should be updated with recent studies on TcB measurements during and after PT in preterm infants.

Response: We have updated our discussion accordingly, including but not limited to papers suggested by Reviewer 2 (Point 2)

14. And as I explained before – I think the final conclusions cannot be drawn without quantifying the potential clinical consequences of the TcB measurements on starting and stopping PT (sensitivity and specificity etc statistics related to rules that have been proposed by other researchers).

Response: We would respectfully disagree with the reviewer. We have designed the study to answer the question of correlation and agreement of TSB and TCB during and after phototherapy and we would believe the agreement could be a major deciding factor for clinicians to use the non-invasive method rather than the blood sample (when testing any new method and comparing it to gold standard). Increasing slope at the start of life (see Response 1, Reviewer 2) and different charts used in different jurisdiction might play a huge role in any predictive model and it would be difficult to compare these models. Furthermore, although it might be interesting to apply the rule from different researchers (for example Hulzebos et al, PloS One 2019) to clarify if the rule prepared for the specific national phototherapy charts would work for the different charts with different thresholds for phototherapy, this should not be, in our opinion, seen as a major message from our work. Furthermore, as outlined in Response 1, we believe that this calculation would not be feasible for our dataset.

Reviewer: 3

We would like to thank Reviewer 1 for his comments and suggestions.

In this manuscript Raba and colleagues describe the accuracy of transcutaneous bilirubin (TcB) measurements during (n=328) and after (n=137) phototherapy (PT) in 196 preterm infants at uncovered and covered skin. Correlations and Bland-Altman data are presented. TcB underestimates total serum bilirubin (TSB) obtained at covered and uncovered sites during phototherapy, and this underestimation persists after discontinuation of PT.

1. This is an interesting manuscript on a topic that has clinical relevance. However, the manuscript raises some questions that need to be addressed. And the clinical relevance should be stressed much more. That is an important issue – please expand on this matter.

Response: We would like to thank the Reviewer 3 for the positive review of our study and we rephrased ‘What the study adds’ part of the manuscript to stress the clinical relevance. We have also amended our abstract, discussion and conclusions to reflect the major findings from our study.

General remarks

2. There are numerous data of TcB measurements during PT (see PubMed). It is well known that PT bleaches the skin. As such TcB measurements on PT-exposed skin areas are not recommended, but there are few studies that state that TcB measurements provide accurate TSB estimations under PT (using a different cut-off rule). After PT there is inaccuracy as well up to a certain time point. This and what this manuscript adds should be mentioned in the Introduction and Discussion. The authors should be able to compare different TcB cut-off methods when to determine a TSB. In general TSB measurement is indicated when TcB+50 µmol/L exceeds the PT threshold. But a different rule may change their conclusion.

Response: We would like to thank the reviewer for this comment. Please see our Response 1 and 14, Reviewer 2, in relation to different cut-off rules. We amended our Results, Discussion and What This Study Adds parts of the manuscript to reflect better post phototherapy measurements with detailed description of 'rebound' TCB/TSB pairs done at eight hours post phototherapy and at 12 hours post phototherapy. We have discussed our findings in the light of recent publications in relation to TCB use during and after phototherapy in preterm infants.

The new post phototherapy data are also included in Table 3.

Specific remarks

4. The title is not very attractive.

Response: As per comment from Editor-in-Chief, we would respectfully leave the title of the study unchanged.

Abstract

5. Please check grammar and punctations. It is not clear at which site TCB is measured under the nappy (abdominal or at the ileal bone?).

Answer: We would like to thank you for you observation. We corrected the grammar and punctations accordingly.

TCB was measured from the bony part of upper outer quadrant of the buttock. We amended our statement in the Abstract 'TCB was measured from an exposed area of skin (the sternum (TCBU)) and from the covered area of skin under the nappy (the bony part of the upper outer quadrant of the buttock (TCBC)) within an hour of obtaining Total Serum Bilirubin (TSB)'. We also amended our statement in the Methods - 'The process was repeated over the covered area, the bony part of upper outer quadrant of the buttock (covered by the nappy).'

6. Please use PT, not PTH, as abbreviation of phototherapy.

Response: We have changed the abbreviation throughout the manuscript as suggested.

7. The conclusion contains not much new information. Please provide clinical relevance of the data.

Response: We believe that the use of transcutaneous bilirubinometry in preterm infants is not supported by NICE guidelines at the moment despite increasing evidence of usefulness of the device. We still believe that our main finding is that despite very good correlation between TSB and TCB there is too wide, clinical relevant disagreement between the non-invasive method and gold standard during phototherapy and as such it would not be recommended and safe to use the same (and as outlined in Response 1 and 14 to Reviewer 2 it would not be feasible to apply suggested 'cut-off rules' to our dataset for various reasons). And as concluded, we believe that it would be recommended to perform 'rebound' post phototherapy measurement by TCB to avoid serum sampling (due to very

good agreement, correlation and minimal mean difference, as soon as 12 hours after cessation of phototherapy). We have added new information about post phototherapy timings and amended our conclusions accordingly.

Introduction

8. The statement that aggressive PT needs to be avoided needs to be clarified including clinical data of published RCTs on that matter in ELBW preterm infants.

Response: We would believe the Reviewer would like us to clarify term 'aggressive PT'. We have used the term "aggressive PT" based on the previous studies coining this term (for example - an RCT entitled 'Aggressive vs. conservative phototherapy for infants with extremely low birth weight. N Engl J Med. By Morris et al'). However as suggested we have deleted the statement in the introduction part – 'While appropriate PTH based on the above criteria is safe, aggressive PTH should be ideally avoided as it could reduce the antioxidant effects of moderate bilirubin levels leading to oxidative injury at cell membranes.' and we have newly added: 'Evidence is conflicting regarding the best therapeutic approach to hyperbilirubinemia, especially in extremely low birth weight (ELBW) infants. A randomised clinical trial (RCT) performed by the Neonatal Research Network found no significant difference in the rate of death or neurodevelopmental impairment at 18 to 22 months corrected age in ELBW infants who received aggressive PT versus those who received conservative PT. However, aggressive PT was associated with a reduction in the rate of neurodevelopmental impairment alone.(5) However, the post hoc analysis showed that in the smallest and sickest subgroup (mechanically ventilated infants with birth weight less than 750g), aggressive PT may increase mortality while reducing neurodevelopmental impairment..'

9. TcB measurements are not recommended in the first 24 hours. This should be added in ll 42-47.

Response: We would respectfully disagree with the reviewer that this is universally accepted. As per American Academy of Pediatrics (AAP) guidelines 'TCB and/or TSB measurements should be performed on all infants who are jaundiced in the first 24 hours of life.' Therefore, it can be performed within 24 hours of age according to AAP guidelines. We agree that NICE guidelines do not recommend TCB in the first 24 hours of life. Because of this controversy, we felt not to mention this in our manuscript as it would not be relevant to the aim of our study which is investigating the accuracy of TCB during and after phototherapy. However, we added the statement in our introduction 'However, TCB measurements are not recommended in the first 24 hours of life or in preterm infants below 35 weeks of gestation according to the National Institute for Clinical Excellence (NICE) guidelines. (https://www.nice.org.uk/guidance/cg98)'

10. Please provide an explanation why PT affects the relationship between TcB and TSB.

Response: We have added an explanation why phototherapy influences the relationship between TCB and TSB. We have amended our statement in the Introduction – '…, some studies reported that PT blanches the skin thereby affecting the correlation between TCB and TSB during and after phototherapy.'

Methods

11. Why were infants with co-morbidity excluded?

Response: We have excluded infants with major congenital abnormalities as presented in the Methods, not infants with co-morbidities, in line with previous studies (Fonseca et al in Journal of Perinatology 2012, Alsaedi et al in IJOP 2017, Luca et al in Journal of Perinatology 2017, Zecca et al in EHD 2009 Hulzebos et al in PLOS 2019) to make our study consistent and comparable with the previous work. We would also believe this would be a standard for most studies in preterm infants.

12. Please provide local PT guidelines.

Response: We agree with the reviewer that this would be very important and we have included the guidelines, please see Response 6, Reviewer 2.

13. Please provide (measured) irradiance of PT devices.

Response: We would like to thank you for this comment. We added irradiance of PT device used in the study. We added the statement to the Methods: '(overhead PT microlight units deliver ≥ 10 µW/cm2/nm and Halogen spotlights which can deliver 20-25uW/cm2/nm).'

14. What was the agreement between the two TcB devices?

Response: We did not measure the agreement between JM 103 and JM 105, as both devices have identical hardware, software and measuring probe. The only differences between two devices are in dimensions, weight and storage ability.

15. Please provide background of the decision of the attending neonatologist to measure TSB.

Response: The decision regarding the frequency of TSB measurements was taken during the ward rounds which were led by the neonatologists and was not related to TCB measurements.

16. Please provide background of why it is important to define EOS.

Response: We defined EOS as it was presented in our baseline population characteristics.

17. Ll 22: Our main outcome was to examine should be changed in "main aim…" or "primary outcome was the correlation….". The next sentence on secondary outcomes should be changes as well.

Response: We would like to thank the reviewer for this observation. We have amended our statement accordingly in the Methods - 'Our primary outcome was the correlation and agreement between TCB (TCBU and TCBC) and TSB during and after PT.' We have decided to take the secondary outcomes of our work out of the manuscript (please also see Response 4, Reviewer 1)

18. Please provide rationale of the effect of clinical parameters on the relationship between TcB and TSB.

Response: We have decided to take this part of our work out of the manuscript (please also see Response 4, Reviewer 1)

19. Ll. How were non-parametric data tested?

Answer: We would like to thank the reviewer for his comment. We tested TCB/TSB samples by paired-samples t-test (see Response 1, Reviewer 1) as we believed that the differences between the pairs were not severely non-normally distributed (McDonald JH, Handbook for Biological Statistics, 3rd Edition, http://www.biostathandbook.com/index.html). However as we did complete review of our statistics, we newly tested the pair differences for normality and when not normally distributed, we have used Wilcoxon signed-rank test. For non-dependent data we have used unpaired t-test or Mann-Whitney U test as appropriate. We have amended our Methods part of the manuscript accordingly: 'Paired-samples t-test was used to compare TCB and TSB paired measurements, when the differences between pairs were not normally distributed, we used the Wilcoxon signed-rank test for two sample comparisons. For non-dependent variables we have used unpaired t-test or Mann-Whitney U test as appropriate.' As we have re-done all statistics with different software (StatsDirect v 3.2.10, StatsDirect Ltd, UK) there were some non-significant changes to our decimal points and confidence intervals throughout the manuscript. We have decided to report TSB and TCB in the whole numbers to reflect that these values were not measured with any decimal points.

20. Please explain why the t-test was chosen to study influence of clinical parameters on TcB-TSB relationship. This seems not appropriate.

Response: We have decided to take this part of our work out of the manuscript (please also see Response 4, Reviewer 1)

Results

21. Please provide time points of TcB and TSB measurements after discontinuation of PT.

Answer: We would like to thank the reviewer for the suggestion. We have added median time (IQR) for TCB/TSB pairs post phototherapy. To make it more clinically relevant, we then created two main categories, TCB/TSB pairs at eight hours post phototherapy and 12 hours post phototherapy, showing that the correlation and mean difference is improving and clinically very acceptable at 12 hours of age. We reflected this in our Results, Discussion and Conclusions accordingly. We have also included new Table 3 presenting our findings.

22. Figures seem appropriate although too small to read and compare. I suggest to remove correlations and only show B-A plots in one Figure (with an a, b, and c panel).

Response: Please see Response 5, Reviewer 1. We have respectfully kept the Figures unchanged.

23 Tables. I did not see 328 measurements, but 309 as maximum number. Please explain.

Response: Thank you for this observation. 328 is the total number of paired TSB and TCBU and/or TCBC.

24. Tables 2 and 3 should be in one Table. Table 4 and 5 should be changed after having applied other statistics.

Response: We have merged the two tables into one table (Table 2) and added more information to Table 2 (median (IQR)) that is not included in the text. We have removed Table 4 and 5.

25. The effect of clinical parameters such as RDS seems not statistically appropriate tested.

Response: We have decided to take this part of our work out of the manuscript (please also see Response 4, Reviewer 1)

Discussion

26. The Discussion is limited to few studies on the same topic, whereas there are lots of published data regarding the relationship between TcB and TSB.

Response: We have updated our discussion using more recent references. (Please see also Response 2, Reviewer 2)

27. The authors speculate on the explanation of the underestimation of TSB by TcB. They mention immaturity of the skin and absence of subcutaneous fat. If true, then the relationship between TcB and TSB in term infants undergoing PT better? But what is the point of action of PT – intra or extravascular bilirubin?

Answer: Based on two studies included in our discussion (Ozkan et al., Acta Paediatr, 2003 and Kanti et al., Skin Pharmacol Physiol, 2014), we speculate that the immaturity of the skin and the absence of subcutaneous fat in preterm infants leads to rapid clearance of extravascular bilirubin levels from the skin following initiation of PT. We have amended our statement – 'Immaturity of the skin and the absence of subcutaneous fat in preterm infants which leads to rapid clearance of extravascular bilirubin levels from the skin, following initiation of PT, may be an explanation'. Obviously the

reviewer's question in relation what is the pathway of action of PT is much more complicated (as PT working on numerous levels, intravascular and extravascular actions) and we believe this would be beyond the scope of our discussion.

28. Please check grammar. For example in ll 29 on page 9: "were more.." – "were higher".

Answer: We would like to thank the reviewer for his observation. We amended our statement into "…, which were higher than the mean gestational age…". Our manuscript was checked for the grammar by all authors, including a native English speaker within the team and then by an independent native speaker with a research background. This is valid for our first submission and also for our revision.

29. P 10-ll10: results should not be repeated. Data should be discussed instead.

Response: We have amended our discussion accordingly, trying to avoid result repetition.

30. Conclusion is clear about the correlation – but that's is not new. Clinical relevance, ie the authors' recommendation, should be stressed much earlier in the manuscript.

However, there are also studies reporting that TcB could be used even under PT – maybe when using other cut-off rules. The authors mention one study, but there are other studies available.

Response: We have now stressed clinical relevance in the Abstract and Conclusions of our paper. We have added more studies to our discussion, including studies using cut-off rules for estimating TSB when TCB done under PT.

31. Regarding the post PT measurements: please provide after how many hours (6? 12?) TcB measurements are accurate again.

Response: Please see Response 21, Reviewer 3

VERSION 2 – REVIEW

REVIEWER	Reviewer name: Peter Flom Institution and Country: Peter Flom Consulting USA Competing interests: None
REVIEW RETURNED	07-May-2020

GENERAL COMMENTS	The authors have responded to my review. The one area of disagreement is around TOST vs. paired t-tests. The NCSS document that they cited is rather confusing. It appears that NCSS used the paired t test procedure to do both regular paired tests (where the null is that there is no difference and you want to find one) and TOST (where the null is that the difference is larger than delta and you want to find it smaller). Thus, one part of the NCSS document is "Paired t-test for TOST" which requires specifying equivalence margins and an equivalence bound. (To add tot he confusion, when I looked for this document, I got a different edition of the user's manual, where this same chanpter is now chapter 202). The key thing is that the null has to be changed, and the usual paired t-test has the wrong null for the needs of this paper. If the authors did do the TOST test, they should add the equivalence margin that they used and why they chose that margin. If they did a regular paired t-test, they should do TOST.

REVIEWER	Reviewer name: Christian Hulzebos Institution and Country: University Medical Center Groningen The Netherlands Competing interests: None
REVIEW RETURNED	19-May-2020

GENERAL COMMENTS	In this revision of their manuscript Raba and colleagues have satisfactorily addressed most major comments. There are some minor issues. These should be addressed as they distract the reader from the content of this manuscript. Minor comments. Abbreviations. Please remove RDS Introduction. Page 5 line 12. Please remove “Despite the NICE guideline”. Methods. Measurements were performed at the bony part of the upper outer quadrant of the buttock (covered by the nappy). So actually at the back of the infant after they were turned when lying on their back? Please clarify. Please remove “Co-morbidities were recorded during the study period” (page 7, line 20), and remove EOS and antibiotics treatment as well (because you have removed the secondary outcomes from the manuscript). Please write independent instead of non-dependent variables (page 8 , line 56). Discussion. Why would “Immaturity of the skin and the absence of subcutaneous fat in preterm infants” lead to rapid clearance of extravascular bilirubin? This is merely a thought/ hypothesis. And in fact the authors reject this hypothesis by mentioning the data of De Luca. Please adapt. Page 11. The authors state that the strengths of their study are “that it is a large prospective observational study that enrolled not only healthy preterm infants, but also sick and ventilated premature infants.” But not so many infants were sick (5 had an EOS, but these data are not helpful, and should be removed) and RDS and ventilation data are lacking completely. Please remove this as a strength. Page 12. “…limited literature…” (line 4) seems not right. Please remove limited. Page 12. ‘rebound’ measurements should be ‘TSB rebound’ measurements. Page 14. What is already known. Few studies reported that TCB measurement from covered skin during PT could provide more accurate approximations of Total Serum Bilirubin level in term infants. But what does ‘more accurate’ mean? Compared with TSB? With TCB after PT? Please clarify. Tables. Please reread Table 1.The total number of infants is lacking. Remove EOS and antibiotics – it is redundant information. Appendix 1 and 2 should not be included (mind copyright issues), but the authors should refer to the website instead.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

We would like to thank Reviewer 1 for his comment.

1. The authors have responded to my review. The one area of disagreement is around TOST vs. paired t-tests. The NCSS document that they cited is rather confusing. It appears that NCSS used the paired t test procedure to do both regular paired tests (where the null is that there is no difference and you want to find one) and TOST (where the null is that the difference is larger than delta and you want to find it smaller).

Thus, one part of the NCSS document is "Paired t-test for TOST" which requires specifying equivalence margins and an equivalence bound. (To add tot he confusion, when I looked for this document, I got a different edition of the user's manual, where this same chanpter is now chapter 202).

The key thing is that the null has to be changed, and the usual paired t-test has the wrong null for the needs of this paper.

If the authors did do the TOST test, they should add the equivalence margin that they used and why they chose that margin. If they did a regular paired t-test, they should do TOST.

Response: Although we respect the Reviewer opinion, we believe that our statistical approach is valid and commonly employed in a agreement analysis situation as supported by evidence offered in our previous response (for example - Shier 2004, Mathematic Learning Support Centre, http://www.statstutor.ac.uk/ resources/ uploaded/ paired-t-test.pdf). In the case of testing two measures that are likely to be correlated, the t-test for paired sample is the most appropriate (Zaki et al. Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. Plos One, 7(5), 2012). We are not using the paired t-test in isolation, but with correlation and most importantly Bland Altman agreement plot (Armitage et al. Statistical Methods in Medical Research, 3rd Edition, Blackwell 1994; Altman. Practical Statistics for Medial Research, Chapman and Hall, 1991). Furthermore as outlined in our previous response, we have now used paired t-test for normally distributed data and Wilcoxon signed-rank test (McDonald JH, Handbook for Biological Statistics, 3rd Edition, http://www.biostathandbook.com/ index.html) for not normally distributed data.

Despite the Reviewer 1 suggestion that the TOST analysis can be done in most major stat software, the two statistical software packages used by us (originally SPSS, IBM, USA and subsequently StatsDirect, StatsDirect Ltd, UK) do not support TOST in their portfolio. We strongly believe that StatsDirect is considered excellent statistical tool for medical research (Freemantle N., StatsDirect—Statistical Software for Medical Research in the 21st Century, BMJ 12/2000, Volume 321, Issue 7275). Paired-t test has intimate relation with agreement analysis in the StatsDirect and would support our claim.

We are aware that ‘absence of evidence is not evidence of absence’ and that ‘non-significant’ paired t-test does not mean equal means. However that is not our case as we were able to reject the null hypothesis.

Another hurdle for the TOST analysis would be an assumption of the normal distribution (and as correctly pointed out by the Reviewer 1 previously some of our measurements are not normally distributed).

Nevertheless, we have deployed 3rd statistical software (R 4.0.0, The R Foundation) and performed paired TOST on all three measurements in question with raw equivalence margin of ±10 umol/l (chosen by us as clinically relevant).

As expected the null hypothesis was not rejected for the three measurements (there are not equivalent, see Table).

Table

Two One-Sided Test (TOST) Equivalence testing – Transcutaneous bilirubinometry (TCB) versus Total Serum Bilirubin (TSB) – raw equivalence bounds -10 and 10

	Mean Difference (µmol/l)	TOST 90 % CI	p
During phototherapy TCBC vs. TSB	25.0	20.9 – 29.2	1.0
During phototherapy TCBU vs. TSB	47.6	43.3 – 52.0	1.0
Post phototherapy TCB vs TSB	10.2	5.9 – 14.5	0.53

TSB, total serum bilirubin; TCBC, Transcutaneous bilirubinometry readings from covered skin; TCBU, Transcutaneous bilirubinometry readings from exposed skin; PT, phototherapy; TCB, Transcutaneous bilirubinometry

For the reasons discussed by us in this and previous response, we would prefer not to include this in our manuscript as we believe that our analysis of the data at the moment is in line with statistical standards published in the literature by many experts. However if the publication of the TOST results would be the Editor’s choice then we would include these results as an additional information to Table 2.

Reviewer: 2

We would like to thank Reviewer 2 for his additional suggestions.

In this revision of their manuscript Raba and colleagues have satisfactorily addressed most major comments. There are some minor issues. These should be addressed as they distract the reader from the content of this manuscript.

Minor comments.

1. Abbreviations. Please remove RDS

Response: We have removed RDS from the abbreviations part of manuscript as suggested

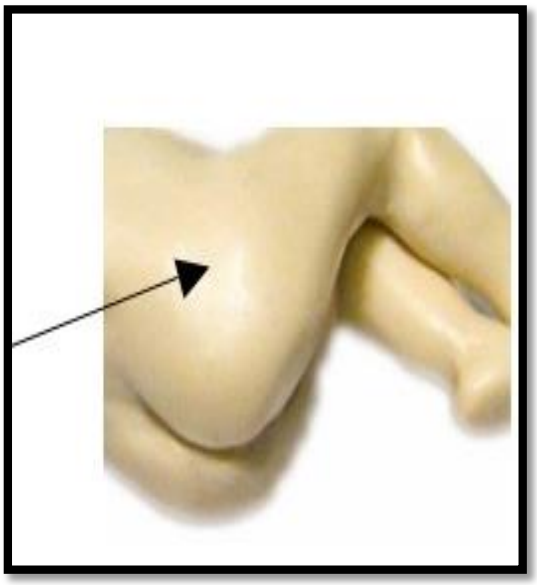
2. Introduction. Page 5 line 12. Please remove “Despite the NICE guideline”.

Response: We have removed ‘~~Despite the NICE guidelines~~’ from the Introduction as suggested

Methods.

3. Measurements were performed at the bony part of the upper outer quadrant of the buttock (covered by the nappy). So actually at the back of the infant after they were turned when lying on their back? Please clarify.

Response: TCB measurements were taken from the bony part of the gluteal area (iliac crest).



4. Please remove “Co-morbidities were recorded during the study period” (page 7, line 20), and remove EOS and antibiotics treatment as well (because you have removed the secondary outcomes from the manuscript).

Response: We have removed statement ‘~~Co-morbidities were recorded during the study period.~~’ from the Methods part of the manuscript. We also removed statement in relation to Early Onset Sepsis (EOS) as suggested - ‘~~Confirmed early onset sepsis (EOS) was defined according to National Institute for Health and Care Excellence (NICE) guidelines (19) as a positive blood culture bacterial infection within the first three days of life.~~’

5. Please write independent instead of non-dependent variables (page 8 , line 56).

Response: We would like to thank the reviewer for this observation. The term has been corrected accordingly.

Discussion.

6. Why would “Immaturity of the skin and the absence of subcutaneous fat in preterm infants” lead to rapid clearance of extravascular bilirubin? This is merely a thought/ hypothesis. And in fact the authors reject this hypothesis by mentioning the data of De Luca. Please adapt.

Response: We would like to thank the reviewer for his comment. We agree that this is our thought/speculation and we adapted the statement accordingly – ‘We speculate, that immaturity of the skin and the absence of subcutaneous fat in preterm infants may lead to rapid clearance of extravascular bilirubin levels from the skin following initiation of PT (9, 24).’

7. Page 11. The authors state that the strengths of their study are “that it is a large prospective observational study that enrolled not only healthy preterm infants, but also sick and ventilated premature infants.” But not so many infants were sick (5 had an EOS, but these data are not helpful, and should be removed) and RDS and ventilation data are lacking completely. Please remove this as a strength.

Response: We have omitted statement in relation to morbidity of the infants in the study as a strength. We reworded this part of the manuscript to – ‘The strength of our study is that it is a large prospective observational study with substantial number of paired TCB-TSB measurements in comparison to previous studies. We have also provided recent data for the agreement between TCB and TSB which is more helpful in clinical practice than correlation coefficient.’

8. Page 12. “…limited literature…” (line 4) seems not right. Please remove limited.

Response: We have removed ~~‘limited’~~ as suggested.

9. Page 12. ‘rebound’ measurements should be ‘TSB rebound’ measurements.

Response: We have amended the statement accordingly.

10. Page 14. What is already known. Few studies reported that TCB measurement from covered skin during PT could provide more accurate approximations of Total Serum Bilirubin level in term infants. But what does ‘more accurate’ mean? Compared with TSB? With TCB after PT? Please clarify.

Response: We would like to thank the Reviewer for this observation. We agree that the statement is slightly confusing. We meant more accurate than the measurement from the uncovered skin. However we changed the statement to – ‘A few studies reported that TCB measurement from covered skin during PT could provide more accurate approximations of Total Serum Bilirubin level in term

infants' that logically follows previous statement ('The accuracy of TCB measurement during phototherapy (PT) is still controversial in term and preterm infants').

11. Tables. Please reread Table 1.The total number of infants is lacking.

Response: We have added the total number to Table 1 (n=196).

12. Remove EOS and antibiotics – it is redundant information.

Response: We have deleted EOS and Antibiotics from Table 1.

13. Appendix 1 and 2 should not be included (mind copyright issues), but the authors should refer to the website instead.

Response: We have removed Appendix 1 and included reference to the website (NICE Guideline phototherapy charts) as suggested. The phototherapy chart for the use in infants ≥32 gestation is unique chart from our hospital (adapted/modified from the NHS, Glasgow at the time). There is no weblink for this chart and we would believe the inclusion as an appendix should not pose any copyright issues as this is now unique chart from our hospital.