PEER REVIEW HISTORY

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

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Direct Bilirubin Levels Observed in Prolonged Neonatal Jaundice: A Retrospective Cohort Study
AUTHORS	Hodgson, Joshua; van Someren, Vivienne; Smith, Colette; Goyale, Atul

VERSION 1 - REVIEW

REVIEWER	Reviewer 1
REVIEW RETURNED	23-Mar-2017

GENERAL COMMENTS	Hodgson et al. present single-center data reflecting a policy of checking bilirubin levels on all infants with persistent jaundice. There are many issues with this study that need further attention: 1. In the title and throughout the text, the term "in health" is used. However, this is a very confusing term. The subjects for this study were identified based on an abnormal clinical sign (jaundice), whereas other infants (presumably 92% of them) were excluded because they did not have the same clinical sign. Hence, if the 420 neonates studied could not be "in health." 2. In the discussion, 20-30% of breastfed infants are said to have persistent jaundice. If 80% of the community's neonates are breastfed, 16-24% should have jaundice. Yet the midwives only identified 8% of neonates. This suggests that midwives are missing some children with slight jaundice and biased towards identifying infants with more severe jaundice (the conclusion is supported by the relatively high total bilirubin values reported in the study). This is important, because in neonatal liver disease the jaundice may be slight in the first months of life, because it is driven by conjugated fractions rather than unconjugated fractions. 3. "Direct bilirubin" varies from hospital to hospital, depending on the way the reaction is run. Hence, using a single direct bilirubin cut-off for many laboratories is not appropriate. For example, a direct bilirubin of 17 umol/L in one hospital may have more conjugated bilirubin than a direct bilirubin of 25 umol/L in another laboratory. This is the major problem with direct bilirubin measurements, and it complicates the authors' goal of finding a universal cut-off. 4. When reference other studies, it is unclear whether direct or conjugated bilirubin is measured. For example, if NICE Guidance uses conjugated bilirubin, then it doesn't make sense for the authors to use 25 numol/L to guide their practice with direct bilirubin (as the

	authors state, the two are not interchangeable). In the discussion of the Davis et al. paper, the 70,000 infants did not have direct bilirubin levels measured as stated. Rather, most had conjugated bilirubin levels measured. 5. The authors interchange mean and percentiles. Means and standard deviations should go together, and medians and percentiles should go together. The figures combine mean with percentiles (but talk about standard deviations in the legends). 6. The authors' conclusion that the NICE Guidance should be followed is not supported by their own data. They do not report any patient with liver disease. Hence, they do not have information either supporting or refuting the utility of the NICE recommendations.

- Permission to publish reviewer 2's review was not received.

REVIEWER	Jancelewicz, Tim University of Tennessee Health Science Center, Pediatric Surgery Competing Interests: I have no competing interests.
REVIEW RETURNED	11-Apr-2017

GENERAL COMMENTS	This paper provides an important analysis of normal bilirubin levels in early infancy, such that practitioners may have a reference standard to use in order to decide which jaundiced patients should undergo further workup. Unfortunately there are a number of structural & stylistic problems that need to be remedied prior to acceptance for publication. Specific points: Background: -What is NICE? -Background is generally well-written; best section of the paper. Methods: -Where is IRB approval / ethical approval? -Statistical analysis? Software used? -This section is underdeveloped. Results: -Page 9, ll. 35 through Page 10, ll.27: This editorializing, while sound, belongs in the discussion, not the results section. Stick to the facts for this information, i.e. state "None of the neonates we investigated further were found to have significant liver disease" or simply present the data. -Page 10: Statistical analysis: these techniques need to be described in the methods section.

	-Define "CI" in the methods -Table 1: define units. Isn't this duplicate information from the figures? Probably not necessary. And difficult/tedious to extract meaningful information from the table. Perhaps combine with the figures in some way such that key information is conveyed without such a complex table. -Figures: need legends. Discussion: -"we believe": you tend to use this phrase. Delete it and say that you do or do not see nearly all jaundiced patients. This phrase and the first part of the discussion is too conversational in tone -Poorly constructed discussion; the beginning of the discussion should summarize the major conclusion of the study -What is Maisels? Must define for the reader Conclusion: -should be "These data provide" (line 20)

VERSION 1 – AUTHOR RESPONSE

Key

Black font = reviewers' comments

Red font = response

Reviewer: 1
Comments to the Author
Hodgson et al. present single-center data reflecting a policy of checking bilirubin levels on all infants with persistent jaundice. There are many issues with this study that need further attention:

1. In the title and throughout the text, the term "in health" is used. However, this is a very confusing term. The subjects for this study were identified based on an abnormal clinical sign (jaundice), whereas other infants (presumably 92% of them) were excluded because they did not have the same clinical sign. Hence, if the 420 neonates studied could not be "in health."

As we explain in the paper, we are confident our population describes accurately cases of breastfed/'physiological' prolonged neonatal jaundice – known to be by far the most common cause(s). Nonetheless we appreciate the point and so have amended the terminology from 'health' to 'disease-free' and in addition on most occasions that this term is used it is qualified by 'neonates with prolonged jaundice'.

2. In the discussion, 20-30% of breastfed infants are said to have persistent jaundice. If 80% of the community's neonates are breastfed, 16-24% should have jaundice. Yet the midwives only identified 8% of neonates. This suggests that midwives are missing some children with slight jaundice and biased towards identifying infants with more severe jaundice (the conclusion is supported by the relatively high total bilirubin values reported in the study). This is important, because in neonatal liver

disease the jaundice may be slight in the first months of life, because it is driven by conjugated fractions rather than unconjugated fractions.

This is explicable and we have amended the discussion to assist in the reader's understanding of this. Most important to emphasise is that our data describe the clinically-relevant population that clinicians will be exposed to and have to make decisions regarding in their day-to-day practice. Nonetheless we have addressed the reviewer's specific points below.

Maisel's et al. found that 20-30% of breastfed infants had visible jaundice at 3-4 weeks and 30-40% of these (i.e. 7-10% of all breastfed neonates) had total bilirubin >90µmol/L. This level is widely regarded as significant and whilst we acknowledge that this is based on little evidence it is nonetheless the level commonly quoted in published studies. Therefore as ~80% of our population initiate breastfeeding we would expect 6-8% of all neonates to be significantly jaundiced at 3-4 weeks. If anything this is less than the 8% referred suggesting that we are seeing the majority of significantly jaundiced neonates.

The second point is that 80% of our Mothers *initiate* breastfeeding, but we know that our population is no different to any other in terms of how rapidly this drops off such that the prevalence of breastfeeding is just 50% by 6 weeks (exclusive breastfeeding 25%) (*https://www.unicef.org.uk/babyfriendly/what-is-baby-friendly/breastfeeding-in-the-uk/breastfeeding-rates-in-the-uk/*). Thus the 6-8% estimate above is a maximum, further decreasing the likelihood that the midwives are missing cases of 'slight' jaundice.

We do not feel that the total bilirubin values we report are 'relatively high'. Our 50th centile is around 150µmol/L and our 5th centiles are all <90µmol/L from ~2 weeks of age. These are very reasonable values in a population referred for visible jaundice – the population that clinicians assess day-to-day.

Lastly, we accept that jaundice may be slight in the first months of life in neonatal/congenital liver disease however we deny the relevance of this. Indeed any disease causing slight or slow-onset jaundice may well not present to a prolonged neonatal jaundice clinic, being more commonly manifested and hence diagnosed in infancy/childhood. Again we must reinforce that we have collected data on the very population that is referred to prolonged neonatal jaundice clinics in practice such that the many clinicians practising in these clinics can make decisions that are better informed. It is simply reassuring to know that our midwives are referring the majority of significantly jaundiced neonates.

3. "Direct bilirubin" varies from hospital to hospital, depending on the way the reaction is run. Hence, using a single direct bilirubin cut-off for many laboratories is not appropriate. For example, a direct bilirubin of 17 umol/L in one hospital may have more conjugated bilirubin than a direct bilirubin of 25 umol/L in another laboratory. This is the major problem with direct bilirubin measurements, and it complicates the authors' goal of finding a universal cut-off.

We acknowledge that, as with any test, there will be a degree of inter-laboratory variation in results and this may be somewhat greater with direct bilirubin than other tests, but we are very sceptical that such wide variation as eluded to here is commonplace with modern analysers. Certainly after spending significant time searching we could find no convincing evidence to that end. We would therefore be interested in seeing the reviewer's source.

Our laboratory utilises modern, accurate and reliable technology. An accurately measured direct bilirubin should only minimally, and fairly consistently, overestimate conjugated bilirubin and so, whilst we accept not perfect, we do not think it unreasonable to use direct bilirubin to inform thresholds (conjugated or direct) for further investigation.

Lastly direct bilirubin (as opposed to conjugated bilirubin) measurement is still widely used and whilst it may be ideal for each laboratory to collect sufficient data to enable them to calculate their own thresholds for further investigation this is unrealistic. Therefore the best approach at this time is to inform national guidance which, even at that scale, remains poorly evidenced. It is that which we have gone some way to achieving.

4. When reference other studies, it is unclear whether direct or conjugated bilirubin is measured. For example, if NICE Guidance uses conjugated bilirubin, then it doesn't make sense for the authors to use 25 numol/L to guide their practice with direct bilirubin (as the authors state, the two are not interchangeable). In the discussion of the Davis et al. paper, the 70,000 infants did not have direct bilirubin levels measured as stated. Rather, most had conjugated bilirubin levels measured.

On every occasion that a study is referenced we clearly state whether direct or conjugated bilirubin is measured. We are grateful for the reviewer having spotted our error with regards to the Davis et al. paper and have duly corrected this. In fact all referenced studies employ direct bilirubin with the exception of Davis et al. (conjugated) and Maisels et al. (transcutaneous).

We fully accept that direct and conjugated bilirubin are not interchangeable. The former provides just a small overestimation of the latter and so, perhaps unfortunately, in practice they are used by clinicians near-interchangeably. This is in fact indicated by the fact that NICE have chosen to use conjugated in their guideline when the majority of the referenced evidence is based on direct bilirubin measurements (we suspect this choice is simply because clinicians have a better understanding of, and familiarity with, the definition of conjugated bilirubin over direct bilirubin).

It would be ideal if all centres used either direct or conjugated bilirubin universally, but this is not the case and so the most appropriate option in informing guidance is to pool evidence sources utilising both, whilst being aware of the limitations that provides.

5. The authors interchange mean and percentiles. Means and standard deviations should go together, and medians and percentiles should go together. The figures combine mean with percentiles (but talk about standard deviations in the legends).

The reviewer has misunderstood here, we suspect they may be confusing percentiles for interquartile ranges that go with medians. In fact it is perfectly reasonable, and indeed often helpful, to use percentiles with means and standard deviations (from which the percentiles are calculated from).

6. The authors' conclusion that the NICE Guidance should be followed is not supported by their own data. They do not report any patient with liver disease. Hence, they do not have information either supporting or refuting the utility of the NICE recommendations.

We are very open that we do not report a patient with liver disease and that that does indeed restrict the extent to which we can support the NICE guidance. Nonetheless we have endeavoured to discuss our own data alongside existing data reporting cases of liver disease. Whilst this is far from statistically ideal, our data suggests that following NICE guidance would result in investigation of ~5% of infants referred to a prolonged neonatal service (which is reasonable) with very low chance of missing a significant case of liver disease as the direct bilirubin levels reported in these cases are many-fold greater than the NICE threshold for investigation.

We accept that not reporting any cases of liver disease is a limitation, but nonetheless we feel that practising clinicians, and indeed NICE, would wish to include our findings in their practice/recommendations (which are currently very poorly evidenced).

Importantly we must also emphasise that determining a threshold for investigation was neither our sole aim nor sole conclusion.

Reviewer: 3
Comments to the Author

This paper provides an important analysis of normal bilirubin levels in early infancy, such that practitioners may have a reference standard to use in order to decide which jaundiced patients should undergo further workup.

Unfortunately there are a number of structural & stylistic problems that need to be remedied prior to acceptance for publication.

Specific points:

Background:

-What is NICE?

This has been defined.

-Background is generally well-written; best section of the paper.

Acknowledged with thanks.

Methods:

-Where is IRB approval / ethical approval?

This is a retrospective study with no intervention. No patients are identifiable and confidentiality has been respected at all times.

-Statistical analysis? Software used?

This is now included.

-This section is underdeveloped.

We have restructured and developed our 'Methods' section.

Results:

-Page 9, ll. 35 through Page 10, ll.27: This editorializing, while sound, belongs in the discussion, not the results section. Stick to the facts for this information, i.e. state "None of the neonates we investigated further were found to have significant liver disease" or simply present the data.

This has been duly moved to the discussion.

-Page 10: Statistical analysis: these techniques need to be described in the methods section.

This has been restructured and now is found in the 'Methods' section.

-Define "CI" in the methods

This acronym has been expanded. If the reviewers and/or editor feel further explanation of 'confidence interval' is of benefit then we can make this addition.

-Table 1: define units. Isn't this duplicate information from the figures? Probably not necessary. And difficult/tedious to extract meaningful information from the table. Perhaps combine with the figures in some way such that key information is conveyed without such a complex table.

As discussed in response to Reviewer 2, one option would be to add n-numbers to the figure (at the risk of busying it) and make Table 1 supplementary. We welcome your opinion regarding this.

-Figures: need legends.

Duly added.

Discussion:

-"we believe": you tend to use this phrase. Delete it and say that you do or do not see nearly all jaundiced patients. This phrase and the first part of the discussion is too conversational in tone

Acknowledged and duly amended.

-Poorly constructed discussion; the beginning of the discussion should summarize the major conclusion of the study

We have accordingly restructured the discussion. We welcome further feedback.

-What is Maisels? Must define for the reader

Duly defined.

Conclusion:

-should be "These data provide" (line 20)

Duly amended.

VERSION 2 – REVIEW

REVIEWER	Karakukcu, Cigdem Saglık Bilimleri University, Education And Research Hospital, Department of Biochemistry, Kayseri, Turkey Competing interests: -Toxicology, TDM, Monitoring of Illicit Drugs, -Hematology -General Biochemistry

REVIEW RETURNED	26-Oct-2017

GENERAL COMMENTS	This is retrospective cohort study or the data collected from 420 neonates (501 blood samples) across an age range of 11-63 days, approximately in two years period to determine the changes in direct bilirubin levels in prolonged jaundice. - Because of the poorly evidenced protocol tresholds for both total and direct bilirubin levels, authors declare that they aimed to establish the spread of direct bilirubin levels and concordantly inform national guidance for the investigation of prolonged neonatal jaundice. -Authors should mention about the Ethical approval fort he study design. -In results the authors declare that they have found no significant liver disease in none of the neonates. However they should proof their conclusion how they decided the normality of liver functions. Is there any added biochemical assays or radiographic findins for that? If there they should includethis in material and methods also. -There are some misspelling and reference writing rules should be obeyed (example superscript before dot). -Manuscript can be accepted after minor revisions.

REVIEWER	Finazzi, Stefano IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, Italy Competing interests: None
REVIEW RETURNED	31-Oct-2017

GENERAL COMMENTS	The manuscript entitled "Prolonged Neonatal Jaundice: Direct Bilirubin Levels Observed in Physiological Jaundice" reports on data of bilirubin fractions in a sample of 420 neonates with prolonged neonatal jaundice. It also discusses the appropriateness of bilirubin thresholds currently used to target further investigations for the assessment for congenital liver diseases. I have a few remarks regarding the presentation and the discussion of the results of this works. a) In the Method section the Authors say "Further follow-up was arranged based on these results". What is the criterion adopted to perform the follow-up? Does this generate any bias on the results of the analysis? Do patients with more than one bilirubin sample have higher bilirubin values than patients with a single sample? Does this bias affect the estimate of the trends of the values of bilirubin fractions? In fact, at a population level the decrease rate of total bilirubin is about 10 times larger than the decrease rate of the direct fraction, whilst at individual level (computed on patients with more than one sample) the mean change of total bilirubin is twice as large as the change of the direct fraction. b) In addition to histogram plot analysis normal distribution of variables should be addressed by some statistical test. c) The results reported in the abstract (mean values) are not presented in the Results section. d) A dispersion parameter around the mean values of bilirubin fractions should be reported.

	e) In both the abstract and the discussion section it is claimed that total bilirubin decreases with age (from 173.4 mu mol / L to 143.2 mu mol / L), but the direct fraction alters little (from 14.3 mu mol / L to 15.2 mu mol / L). However linear regression analysis shows that the rate of change of total and direct bilirubin are -3.72 mu mol / L / day, and -0.39 mu mol / L / day, respectively. Normalizing to the initial values of both variables, one obtains a 2.1% relative decrease per day for the total bilirubin fraction and a 2.8% relative decrease per day for the direct bilirubin fraction. It seems to me that direct bilirubin decreases as well as total bilirubin. Accordingly, the ratio between direct and total bilirubin does not change. Similarly, at the individual level the mean change in total bilirubin was -10 mu mol / L, in direct bilirubin was -4.5 mu mol / L, whereas their ratio does not change significantly. Can the Authors comment on this? Why do the abstract and the discussion session seem to contradict the result section? f) Results - The individual level: direct-total bilirubin ratio is a dimensionless number. g) Can the Authors report the confidence interval of the 95th percentile of bilirubin values?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
This is retrospective cohort study or the data collected from 420 neonates (501 blood samples) across an age range of 11-63 days, approximately in two years period to determine the changes in direct bilirubin levels in prolonged jaundice.

- Because of the poorly evidenced protocol tresholds for both total and direct bilirubin levels, authors declare that they aimed to establish the spread of direct bilirubin levels and concordantly inform national guidance for the investigation of prolonged neonatal jaundice.

-Authors should mention about the Ethical approval for the study design.

Please see information declared in relevant section of online submission form.

-In results the authors declare that they have found no significant liver disease in none of the neonates. However they should proof their conclusion how they decided the normality of liver functions. Is there any added biochemical assays or radiographic findins for that? If there they should includethis in material and methods also.

This is a well highlighted omission which we have duly added to the relevant sections of the text.

-There are some misspelling and reference writing rules should be obeyed (example superscript before dot).

Many thanks for highlighting. Duly addressed. Most superscript citations now succeed punctuation marks where the citation refers to all text in that sentence preceding said punctuation mark. However where the superscript citation refers only to a phrase within the sentence it precedes the punctuation. If the reviewers and/or editor disagree with this methodology we are of course open to alteration.

-Manuscript can be accepted after minor revisions.

Noted with thanks. Above revisions made.

Reviewer: 2

Comments to the Author
The manuscript entitled "Prolonged Neonatal Jaundice: Direct Bilirubin Levels Observed in Physiological Jaundice" reports on data of bilirubin fractions in a sample of 420 neonates with prolonged neonatal jaundice. It also discusses the appropriateness of bilirubin thresholds currently used to target further investigations for the assessment for congenital liver diseases.

I have a few remarks regarding the presentation and the discussion of the results of this works.

a) In the Method section the Authors say "Further follow-up was arranged based on these results". What is the criterion adopted to perform the follow-up? Does this generate any bias on the results of the analysis? Do patients with more than one bilirubin sample have higher bilirubin values than patients with a single sample? Does this bias affect the estimate of the trends of the values of bilirubin fractions? In fact, at a population level the decrease rate of total bilirubin is about 10 times larger than the decrease rate of the direct fraction, whilst at individual level (computed on patients with more than one sample) the mean change of total bilirubin is twice as large as the change of the direct fraction.

These are good points the reviewer was right to raise and we have accordingly addressed them.

The criteria for arranging further follow-up were decided by the responsible clinician on a case-by-case basis. This was due to a lack of confidence in local and national guidance – the very trigger for initiating this work. Whilst we acknowledge that from a statistical point of view this is not optimal, in fact it is the case that most clinicians across the UK (and beyond) practise in this way due to a recognised lack of evidence informing the guidance. Therefore the lack of evidence leads to guidance that is treated with a lack of confidence and in turn makes the generation of very statistically rigorous evidence near-impossible. This is an unfortunate cycle that can only be broken by the gradual accumulation of evidence of progressive quality – a process that our work seeks to contribute to.

The reviewer is right to highlight that there may be bias amongst the patients with more than one bilirubin sample versus those with a single sample. Indeed patients with more than one split bilirubin sample have, on average, a 27% higher total bilirubin ($p = 0.0002$), 6% higher direct bilirubin ($p = 0.0370$) and 30% higher direct-total bilirubin ratio ($p = 0.0067$). Thus there is a degree of bias in the analysis at the individual level and we accept this as a limitation however we are not directly comparing the two populations. In addition this bias is of interest in highlighting that practitioners continue to place the most weight on a raised direct-total bilirubin ratio and the least on direct bilirubin which our data suggest may be inappropriate. We have added these results and discussion to the paper.

b) In addition to histogram plot analysis normal distribution of variables should be addressed by some statistical test.

For each age group and biomarker a number of checks were performed to assess normality. Firstly, a number of visual approaches were undertaken to assess deviations from normality: histograms and Q-Q plots were drawn, and the mean and medians in each sample were compared. Our checks verified that the normal assumption was met. In addition, we calculated non-parametric estimates of the percentiles (in Table 1), and estimates using a normal distribution assumption for Figure 1 – thus reviewers can determine that there are relatively minor differences between these estimates. In our original submission we did not use statistical tests for normality as, when considering large sample sizes, the tests have high power and so relatively minor deviations from normality can lead to rejection of the null hypothesis. However we have now done so using the Kolmogorov-Smirnov test for each age group and bilirubin marker. Each had $p>0.05$ with the exception of 18-24 days (all 3

markers) and 25-31 and 32-42 days (direct-total bilirubin ratio only). Note that these are the groups with the largest sample size.

c) The results reported in the abstract (mean values) are not presented in the Results section.

Duly corrected. In fact these are not the key result values and so have been removed from the abstract.

d) A dispersion parameter around the mean values of bilirubin fractions should be reported.

Duly added.

e) In both the abstract and the discussion section it is claimed that total bilirubin decreases with age (from 173.4 mu mol / L to 143.2 mu mol / L), but the direct fraction alters little (from 14.3 mu mol / L to 15.2 mu mol / L).
However linear regression analysis shows that the rate of change of total and direct bilirubin are -3.72 mu mol / L / day, and -0.39 mu mol / L / day, respectively. Normalizing to the initial values of both variables, one obtains a 2.1% relative decrease per day for the total bilirubin fraction and a 2.8% relative decrease per day for the direct bilirubin fraction. It seems to me that direct bilirubin decreases as well as total bilirubin. Accordingly, the ratio between direct and total bilirubin does not change. Similarly, at the individual level the mean change in total bilirubin was -10 mu mol / L, in direct bilirubin was -4.5 mu mol / L, whereas their ratio does not change significantly.
Can the Authors comment on this? Why do the abstract and the discussion session seem to contradict the result section?

This is a very well raised point that is gratefully acknowledged. Here the reviewer places more emphasis on the linear regression analysis rather than the graphical representation of percentiles whereas we had originally done the reverse. The reviewer's inclusion of relative decrease per day is particularly inciting. In fact their well-reasoned argument has persuaded us that the linear regression analysis is more useful in interpreting the trend of split bilirubin with age whilst the graphs serve better as a practical reference tool for clinicians. Accordingly we have altered our abstract and discussion to concord with the linear regression analysis and conclude that both total and direct bilirubin decrease with age. We have also included the concept of relative decrease which also emphasises the importance of a direct bilirubin that is failing to decrease in an individual patient.

Whilst these amendments alter the content of our abstract and discussion, they do not alter our conclusions and points of practical application for clinicians.

f) Results - The individual level: direct-total bilirubin ratio is a dimensionless number.

Duly amended.

g) Can the Authors report the confidence interval of the 95th percentile of bilirubin values?

Duly added to Table 1.