PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Laboratory reference intervals in the assessment of iron status in young children
AUTHORS	Parkin, Patricia; Hamid, Jemila; Borkhoff, Corneilia; Abdullah, Kawsari; Atenafu, Eshetu; Birken, Catherine; Maguire, Jonathon; Azad, Azar; Higgins, Victoria; Adeli, Khosrow

VERSION 1 - REVIEW

REVIEWER	Ridefelt, Peter Dept of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden Competing interests: None
REVIEW RETURNED	14-Jun-2017

GENERAL COMMENTS	This manuscript describes the definition of reference intervals for parameters used to assess iron status in children. Blood has been sampled from a large number of healthy children, >4 300. The methods used are up to date, and statistics used appropriately. Reference intervals are established by measurements on blood samples from healthy individuals. Medical decision limits are cut-offs, often suggested in statements papers from national or international expert bodies. Such decision limits are at the best derived from outcome studies, but in reality often come from published professional recommendations residing lower in the hierarchy described in reference 9. My main concern is that the paper highlights the risk of misclassification of children as the principal message. That is a true and real risk, however, this paper only deals with defining reference intervals for healthy children, age 10 days to 10 years. It has not studied clinical outcomes, which is the most preferred method (ref 9). Adopting the contents of that reference would even the put medical decision limits from AAP in a lower category than the presently used strategy for defining reference intervals. The true essence of the paper only become clearly evident in the very last sentence of the last paragraph in the Discussion; …, we have described the differences between reference intervals and decision limits, and have used hemoglobin and serum ferritin as examples to highlight the potential for misclassification when using reference intervals ALONE." Thus, in my opinion the main advancement in this paper is a solid and accurate work describing reference intervals in young children in iron status parameters. How labs and clinicians should present and use laboratory measurements have not been addressed by the present way of organizing the study. The present study is limited to children under 10 years of age. This is a major drawback when these figures should be used in other labs. The motto of the CALIPER initiative, to fill the gaps, is clearly

not met in the present study. This limitation should be discussed. Also, the statement paper from AAP only deals with infants and toddlers up to 35 months of age. The present paper also includes children 3-10 years old. The authors does not discuss decision limits for children over 3 years of age.

The paper use fixed intervals when partitioning for age; 10days – 1 year, 1 – 3 yrs….
Age- and gender partitioning of reference intervals is a challenging task, and there is no universal and accepted method. However, from a physiological point of view it is unlikely that all these parameters would optimally fit a model with fixed age groups. The CALIPER Colantonio paper (ref 10) use one age group for iron between 0 and 14 years of age. What was the objective behind this treatment of age groups?

Minor comments

Materials and methods describes instruments used. It would be preferred if also reagents and calibrators used were stated to facilitate the transferability of these figures to other labs.
There are two different traceability chains for ferritin methods (IS 80/602 and IS 94/572), which one was applicable here?

It is stated that parametric or non-parametric statistics were used for final calculation of reference intervals after testing for skewness and kurtosis. However, Table 2 does not specify the details of how these calculations were made, and when non-parametric vs parametrics were used.

Number of significant digits are inconsistent, e.g the upper range for months is 126.95 in Table 1. Hemoglobin upper limit for males is 132.9, and 133 for the two youngest groups, Table 2.

The saw tooth appearance in the figure suggests that age was not random. Were the visits to the checkups mostly done around the birthday of the child?

CRP was measured. Page 6, line 9 state that one reason was that ferritin is an acute phase reactant. However, nothing is mentioned about the outcome. Were children excluded due to high CRP? If so, which cut-off was used?

Lower limits of CRP are highly uncertain, even in hsCRP methods. The figure 0.09 mg/L pops up in numerous places in Table 2, hinting that this might have been a LOD (lower level of detection) or LOQ (lower limit of quantification). What were the LODs or LOQ for the Roche method? Nevertheless, even lower numbers than 0.09 is stated in the table, e.g. 0.036 mg/L for lower limit in males 1-3 yrs.

Table 3. STfR is mentioned in the foot note, but no data are present in the table.

The presentation of current knowledge in the Introduction about reference intervals for iron status parameters in children is fastidious. Table 3 is adequate when describing the Canadian perspective with six recent publications on iron status related parameters from the CALIPER and CHMS studies. However, these six publications are the only ones mentioned in the Introduction,

	page 6, line 13. The Introduction does not hint that the papers cited are only Canadian studies. There are other relevant studies published.

VERSION 1 – AUTHOR RESPONSE

	Comment	Response to Comment
R1	My main concern is that the paper highlights the risk of misclassification of children as the principal message.	In our objectives (Introduction and Abstract), Statistical Analysis and Results the risk of misclassification is described as our secondary objective. The primary objective of this study was to establish reference intervals for hematologic and biochemical tests commonly ordered by clinicians to assess the iron status in young children using the CLSI guidelines.
R2	It has not studied clinical outcomes, which is the most preferred method.	As highlighted by the Reviewer, the hierarchy described by Sikaris (reference 6) indicates that the highest quality of evidence for decision limits is on clinical outcome studies. We agree with the Reviewer that we did not study decision limits derived from clinical outcome studies. To address this, we have modified the manuscript throughout, using the term 'cut-off values' when referring to the recommendations from the American Academy of Pediatrics (AAP); and using the term 'decision limits' when referring to values derived from clinical outcome studies.
R3	The present study is limited to children under 10 years of age. This limitation should be discussed.	We have added to limitations that the present study is limited to young children under 10 years of age. To fill the gaps in pediatric reference intervals, other initiatives, such as CALIPER, have large sample sizes for children older than 10 years.
R4	The authors do not discuss decision limits for children over 3 years of age.	We have added to limitations that we have not addressed cut-off values for children 3 to 10 years.
R5	The paper use fixed intervals when partitioning for age; 10 days – 1 year, 1 – 3 yrs. What was the objective behind this treatment of age groups?	In Methods we have written: Age partitions were selected based on epidemiologic knowledge of the changing prevalence of iron deficiency and overall child growth and development. We have added: It is well established that the age group of 1 to 3 years is an age of peak prevalence for iron deficiency, thus we included a specific partition for this age group.
R6	It would be preferred if also reagents and calibrators used were stated to facilitate the transferability of these figures to other labs.	To address this, we have added Supplemental Table 1 to the Sample Collection and Analysis section.
R7	There are two different traceability chains for ferritin methods (IS 80/602 and IS 94/572), which one was applicable here?	The traceability chain used for serum ferritin was IS 80/602. This has been added to the Methods-Sample Collection and Analysis section.
R8	When were non-parametric vs parametrics used?	We have added to Methods-Statistical Analysis: The parametric method was used if the normality assumption was met or the sample size in each partition was large; the non-parametric method was used if the normality assumption failed or the sample size was limited. To show which method was used for each analyte

		Supplemental Table 2 has been added to Statistical Analysis section.
R9	Number of significant digits are inconsistent, e.g the upper range for months is 126.95 in Table 1. Hemoglobin upper limit for males is 132.9, and 133 for the two youngest groups, Table 2.	We have edited so that results are presented with a maximum of 2 or 3 digits.
R10	Were the visits to the checkups mostly done around the birthday of the child?	To clarify, we have added a sentence: Health supervision visits occur at 2 weeks, 2, 4, 6, 9, 12, 15, 18 months and then annually around the birthday of the child.
R11	Were children excluded due to high CRP?	We did not exclude any children due to high CRP.
R12	What was the lower level of detection for CRP for the Roche method?	We have added to the Methods-Sample Collection and Analysis section: The lower level of detection for CRP is 0.15 mg/L.
R13	Table 3. STfR is mentioned in the foot note, but no data are present in the table.	STfR has been removed from the foot note.
R14	The presentation of current knowledge in the Introduction about reference intervals for iron status parameters in children is fastidious.	We have shortened the paragraph regarding reference intervals for iron status in children in the Introduction.
R15	Six publications are mentioned in the Introduction, page 6, line 13. The Introduction does not hint that the papers cited are only Canadian studies.	We have added two publications from theNordic Reference Interval Project 2000 (NORIP) which present data for iron related laboratory tests.
E1	How representative is this sample of the wider population that the authors feel the results are generalizable to?	To address generalizability, we have modified the sentence in Methods to: The profile of this open longitudinal cohort has been previously described, and children with blood samples and without blood samples are similar with respect to demographics and health outcomes. Furthermore, the prevalence of iron deficiency is similar to other Canadian studies of this age group. Readers are referred to Table 1 in Reference 18 which describes the characteristics of the entire cohort, as well as according to with/without blood sample subgroups.
E2	The authors state that “the non-parametric or parametric (as appropriate) method was used. What method was used for each of the examined analytes?	To show which method was used for each analyte Supplemental Table 2 has been added to Statistical Analysis section.
E3	“Some children contributed a blood sample to more than one age-group” – is this because they were seen repeatedly?	To address repeated sampling, we have modified the sentence in Methods to: The profile of this open longitudinal cohort has been previously described, and children with blood samples and without blood samples

		are similar with respect to demographics and health outcomes. We have also added to Methods: we have added a sentence: Health supervision visits occur at 2 weeks, 2, 4, 6, 9, 12, 15, 18 months and then annually around the birthday of the child.
E4	Does this influence the findings?	If two or more samples were obtained from a child, these data were used for estimating reference intervals for different age groups. Data within a specific age group remain independent and do not violate the independence assumption. This allowed us to use all available data and increases the precision of our estimates. We would like to highlight that this is a common approach used when estimation is only needed at age group level, and when the focus is not the longitudinal nature of the data.
E5	I am concerned about the conclusions of this study. The authors state that "clinical laboratories may consider adopting the reference intervals presented here", but go on to say that "reference intervals may misclassify (underestimate) children with iron deficiency as compared with decision limits" However they acknowledge that "…limited by the low quality of evidence used to establish the currently recommended decision limits". It seems to me that there appears to be no gold standard to compare to, so any comparison is therefore questionable.	The Editor highlights that there is no 'gold standard' to compare with. Similarly, the Reviewer (R2 above) highlighted that the highest quality of evidence for decision limits is on clinical outcome studies. To address this, we have modified the manuscript throughout, using the term 'cut-off values' when referring to the recommendations from the American Academy of Pediatrics (AAP), which would not be considered 'gold standard'; and using the term 'decision limits' when referring to values derived from clinical outcome studies, which would be considered highest quality (ie, 'gold standard'). We have also modified the Discussion to address the Editor's concerns regarding the conclusions.