Estimation of age-specific reference intervals for laboratory blood tests in children

Mohammad Monsurul Hoq

ORCID 0000-0002-0504-9335

Doctor of Philosophy

November 2020

Department of Paediatrics

Faculty of Medicine, Dentistry & Health Sciences

The University of Melbourne

A thesis submitted in total fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Introduction: Reference intervals (RI) are an important clinical assessment tool used by physicians and laboratory professionals to interpret test results. A RI is commonly defined as a range between the 2.5th and 97.5th percentile and represents the middle 95% of the reference population. The estimation of RIs is a complex process which involves defining a reference population, selecting reference individuals, collecting, and testing reference samples, and applying appropriate statistical methods.

Estimation of RIs in children is challenging for a number of reasons. There is a need for adequate representation of the general population in the reference population. Analytes are influenced by the nutrition, development, and growth of a child. Hence, RIs should reflect these age-dependent changes in the analytes as children grow older. Most laboratory tests are instrument dependent and variation between different analysers has been documented in several studies. In addition, the best statistical methods for estimating and reporting RIs varying with continuous age are still not widely agreed upon. Finally, there are no available guidelines for laboratories to validate continuous age-specific RIs for their use.

Hence, this project addressed the following four gaps: 1) the lack of appropriate and reliable continuous age-specific paediatric RIs, 2) an absence of head-to-head comparison of laboratory results on different analysers, 3) the lack of clear evidence and accompanying guidelines regarding the best statistical methods for constructing continuous age-specific paediatric RIs, and 4) the lack of guidelines for validating continuous age-specific RIs.

Methods: This research project consists of four studies. Data collected as part of the HAPPI Kids study, from children 30 days to < 18 years, were used to compare laboratory test results

on different analysers and to establish and validate continuous age-specific RIs. Differences in mean test results of analytes by analyser types were investigated using mixed-effects regression analysis and by comparing maximum variation between analysers with analyte-specific allowable total error reported in the Westgard QC database. Continuous age-specific paediatric RIs were estimated using quantile regression where power variables in age were selected based on fractional polynomial regression for the mean, with modification by sex when appropriate. The continuous age-specific paediatric RIs were considered valid if more than 90% of test results from a routine laboratory were within the intervals. A systematic review was conducted to examine the range of statistical methods used over the past 25 years for the estimation of age-specific RIs and to identify trends in usage and reporting. A simulation study was conducted to evaluate and compare statistical methods for constructing continuous age-specific RIs in children under different scenarios and for different sample sizes.

Results: The variation in the mean test results across five analysers was not clinically significant for 24 out of 30 biochemistry analytes examined and common continuous age-specific RIs were established for children 30 days to 18 years. The continuous age-specific paediatric RIs were validated in a routine laboratory after initial analysis for most analytes and after secondary analysis for a few analytes.

According to the results of the systematic literature review, a wide variety of statistical methods have been used for estimating age-specific RIs in children over the past 25 years. However, there has been insufficient uptake of modern statistical methods in estimating continuous age-specific paediatric RIs for analytes that change with age. The results of the simulation study show that in comparison to the commonly used methods for estimating

continuous age-specific paediatric RIs, the applied novel method that we applied performed better in most scenarios for sample sizes of 400 and over.

Discussion: The comparison of test results for common biochemistry analytes should provide evidence of lack of differences between analysers and in turn improve clinical interpretation. It is recommended that continuous age-specific RIs are used as a gold standard for interpreting test results. Laboratories should estimate or validate continuous age-specific paediatric RIs for their use by applying appropriate statistical methods based on the evidence provided in this research.

Declaration

I, Mohammad Monsurul Hoq, declare that this thesis, “Estimation of age-specific paediatric reference intervals for laboratory blood tests in children” and the work presented in it are my own. I confirm that:

- The thesis comprises only my original work toward the PhD except where indicated in the preface,
- Due acknowledgement has been made in the text to all other materials used, and
- The thesis is fewer than 100,000 words in length, exclusive of tables, figures, references and appendices.

Mohammad Monsurul Hoq

Date: November 16, 2020

Preface

This thesis makes use of data from the HAPPI Kids study. The HAPPI Kids study is conducted in collaboration among 1) four tertiary hospitals in the greater Melbourne area: Royal Children's Hospital, Royal Women's Hospital, Northern Hospital, and Western Health-Sunshine Hospital, 2) six laboratories: Royal Children's Hospital, Royal Melbourne Hospital, Melbourne Pathology, Monash Health, Australian Clinical Labs and Dorevitch Pathology, and 3) Murdoch Children's Research Institute in Melbourne, with advice provided by a consortium of leading researchers, clinicians, biochemists, chemical pathologists, biostatisticians etc. The ethical approval and sample collection of the project started before the commencement of the PhD. However, I led the process of publishing the study protocol in BMJ Open and highlighted the uniqueness of the protocol. These unique aspects of the protocol are 1) prospective collection of samples from healthy children, 2) justification of the sample size based on the existing literature, and 3) a novel statistical method for estimating continuous age-specific paediatric reference intervals (RIs). The protocol was published in BMJ Open and is included as the body of Chapter 2. The contribution of many people involved in the study is noted in the Supplementary Table 1 of the appendix B and C.

One of the main aims of the thesis was to estimate continuous age-specific paediatric RIs for 30 biochemistry analytes. The existing literature suggests that the traditional approach of estimating discrete age-group-based paediatric RIs is misleading and does not appropriately address the changes in analytes with age. Addressing this gap, I have applied a novel statistical method estimating RIs varying with continuous age using data available from The HAPPI Kids study. I have also applied a combination of statistical and clinical approaches to propose common RIs across five commercial analysers. This study was published in Clinical Chemistry and is included as the body of Chapter 3.

The collaborative focus of this PhD meant that the initial work was to analyse and report continuous age-specific RIs for the HAPPI Kids data, but this highlighted three critical issues for practical application and reliability. These were 1) validity of the continuous age-specific paediatric RIs and process of validation, 2) adequacy of the sample size for estimating continuous age-specific RIs for analytes with varying complexity of the relationship with age, and 3) performance of the novel statistical method in comparison to other commonly used statistical methods. Addressing these gaps, I led the following three studies.

The Clinical Laboratory and Science Institute (CLSI) recommends that established RIs should be validated before use in a routine laboratory. However, there are no guidelines for validating continuous RIs. Hence, we conducted the validation study immediately after the estimation of continuous RIs. I have demonstrated that the estimated continuous age-specific paediatric RIs were valid for use in routine laboratories using the data from the HAPPI Kids study, tested in two different routine laboratories applying the same analytical principal and analyser. In the process of validating the continuous age-specific paediatric RIs, I have proposed some modifications to the existing validation guidelines. The validation study was published in The Journal of Applied Laboratory Medicine and is included as the body of Chapter 4.

To assess the adequacy of the sample size and compare performance of our novel statistical method, I needed to explore commonly used sample sizes and uses of previously proposed statistical methods in the estimation of continuous age-specific paediatric RIs. Hence, I led a systematic review of published papers reporting age-specific paediatric RIs. Specific recommendations in terms of reporting statistical methods and estimating continuous age-specific RIs are made based on the findings of the review. This study was presented at a

conference as a poster and was published in Clinical Biochemistry. The final submitted version of the paper is included as the body of Chapter 5.

Finally, I have led a statistical simulation study that aimed to explore the performance of our novel proposed method under different scenarios and for different sample sizes compared to other methods identified in the systematic review. The findings of the simulation study were presented at three conferences and submitted for publication in Biometrical Journal. The latest submitted version of the manuscript is included as the body of Chapter 6.

A review of current status, gaps, challenges and future considerations regarding paediatric RIs was published in Clinical Biochemist Reviews in response to a request from the editor. My contributions to the manuscript are included as the body of the thesis on pages 1, 4 - 10, 16 - 17, 47 - 48, 79 - 81, 102, 133 and 163 - 169. A copy of the author accepted manuscript is provided in Appendix F.

I am the primary co-author of the six publications and contributed more than half of their content. I was responsible for data management, statistical analysis (including programming) and drafting the initial manuscript for all six papers. In addition, I was responsible for planning the validation, systematic review, and simulation studies. I have contributed to the final manuscripts with feedback from the journal editors and reviewers, and support from my supervisors and co-authors.

The references to all six articles are:

1. Hoq M, Karlaftis V, Matthews S, Burgess J, Donath SM, Carlin J, Monagle P, Ignjatovic V.A prospective, cross-sectional study to establish age-specific reference

intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: the HAPPI Kids study protocol. *BMJ Open*. 2019; 9(4):e025897.

2. Hoq M, Matthews S, Karlaftis V, Burgess J, Donath S, Carlin J, Yen T, Ignjatovic V, Monagle P. Reference values for 30 common biochemistry analytes across five different analyzers in neonates and children 30 days to 18 year olds, *Clinical Chemistry*. 2019; 65(10): 1317 - 1326.
3. Hoq M, Matthews S, Karlaftis V, Burgess J, Donath S, Carlin J, Ignjatovic V, Monagle P. Validation of the HAPPI Kids Continuous Age-Specific Pediatric Reference Intervals. *The Journal of Applied Laboratory Medicine*. 2020: jfaa045
4. Hoq M, Canterford L, Matthews S, Khanom G, Ignjatovic V, Monagle P, Donath S, Carlin J, Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review. *Clinical Biochemistry*. Aug 2020.
5. Hoq M, Monagle P, Donath S, Carlin J. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. *BMC Research Methodology*. 2021. (Submitted)
6. Hoq M, Matthews S, Donath S, Carlin J, Ignjatovic V, Monagle P. Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations. *Clinical Biochemist Review*. 2020; 41 (2).

Based on the work of this thesis, I have presented the primary findings in three conferences. The references to these conference abstracts are:

1. Hoq M, Monagle P, Donath S, Carlin J. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. in Young Statistician Conference. 2019. Canberra ACT.

2. Hoq, M, Canterford L, Khanom G, Monagle P, Donath S, Carlin J. Statistical methods applied in the estimation of age-specific paediatric reference intervals for a laboratory blood test: A systematic review. in Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference. 2019. Adelaide, SA: *Clinical Biochemist Review.* 2020; 41 (1) Supplement.
3. Hoq, M, Monagle P, Donath S, Carlin J. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. in Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference. 2019. Adelaide SA: *Clinical Biochemist Review*. 2020; 41 (1) Supplement.
4. Hoq, M, Monagle P, Donath S, Carlin J. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. In 41st Annual Conference of the International Society for Clinical Biostatistics. 2020. Virtual.

My PhD research was supported by the Royal Children's Hospital Foundation (RCHF, grant number 233, 2014), The Victorian Centre for Biostatistics (ViCBiostat), and a National Health and Medical Research Council (NHMRC) Postgraduate scholarship (grant number APP1168363, 2019).

Acknowledgement

I would like to thank my PhD supervisors. Without their supervision, guidance and support my PhD journey would have been difficult. I thank A/Prof Susan Donath for her motivation to do the PhD. I could share my challenges and success with her. I am grateful to Prof Paul Monagle for sharing his knowledge in paediatrics and laboratory medicine. The numerous discussions we had helped me improve my understanding of the topic. I am indebted to Prof John Carlin for continuously challenging me to improve my statistical knowledge, communicating the message clearly and refining my English. I would also like to thank my PhD advisory committee, Prof Harriet Hiscock and Prof Julie Simpson for their feedback, reflection and direction to complete the PhD on time.

I would like to thank all my co-authors (Vera Ignjatovic, Susan Matthews, Janet Burgess, Vicky Karlaftis, Tina Yen, Jessica Cowley, Gulshan Ara Khanom and Louise Canterford) for their contributions in drafting the manuscripts, which gave me a wonderful opportunity to learn more from their expertise. A special thanks to Dr Susan Matthews for sharing her expertise in clinical biochemistry. Without her support and insight in clinical biochemistry it would have been challenging for me to interpret the data. I would also like to thank Prof Vera Ignjatovic for her supervision and guidance at the initial stages of my PhD and contributing as a co-author in different publications.

I extend my gratitude to Dr Margarita Moreno-Betancur, Rushani Wijesuriya, and Diana Zannino for their technical support in implementing the statistical simulation study and Dr Sarah Arnup for sharing her expertise in systematic reviews of statistical methods. I am also grateful to the Graduate Research Hub, Graduate Students' Association, Library of The University of Melbourne and Research Students' Association and Communication Team of

Murdoch Children’s Research Institute for their capacity building program and technical support during my PhD.

The mental health and wellbeing support received from my colleagues in the Clinical Epidemiology and Biostatistics Unit and Haematology Research group of Murdoch Children’s Research Institute, and the Trans20 study team at The Royal Children's Hospital made my PhD Journey enjoyable and memorable. I would like to thank all of them.

I appreciate the encouragement and motivation of my friends, especially Dr Matthew Reeve, Dr Felicity Reeve, and Abu Hanif Mohammad Abu Hasib. I am indebted to my mother, Mahmuda Sultana and two sisters, Shamima Shultana and Abida Shultana, who made a lot of sacrifices for my achievements, which in turn gave me the strength to complete my PhD. Lastly, no word is enough for my lovely wife, Gulshan Ara Khanom, who greatly supported me during the time of the PhD.

Table of Contents

ABSTRACT ..... **II**

DECLARATION ..... **V**

PREFACE ..... **VI**

ACKNOWLEDGEMENT ..... **XI**

TABLE OF CONTENTS ..... **XIII**

LIST OF FIGURES ..... **XVI**

ABBREVIATIONS ..... **XVII**

CHAPTER 1: INTRODUCTION **1**

1.1 Reference intervals and their clinical importance *1*

1.2 Paediatric reference intervals *3*

1.3 Challenges in estimating paediatric reference intervals *4*

1.4 Current status of paediatric reference interval initiatives *8*

1.5 Statistical analysis of the laboratory results to estimate reference limits *11*

1.6 Validation and transference of estimated paediatric RIs *18*

1.7 Synopsis of the research problem *19*

1.8 Research aims and objective *20*

1.9 Thesis structure *21*

CHAPTER 2: STUDY DESIGN, DATA COLLECTION METHOD AND ANALYSIS PLAN OF THE HARMONISING AGE PATHOLOGY PARAMETERS IN KIDS (HAPPI KIDS) STUDY **23**

2.1 Introduction *23*

2.2 Publication *24*

2.3 Discussion *49*

CHAPTER 3: REFERENCE VALUES FOR 30 COMMON BIOCHEMISTRY ANALYTES ACROSS FIVE DIFFERENT ANALYSERS IN NEONATES AND CHILDREN 30 DAYS TO 18 YEAR OLDS 52

3.1 Introduction *52*

3.2 Publication *52*

3.3 Discussion *83*

CHAPTER 4: VALIDATION OF THE HAPPI KIDS CONTINUOUS AGE-SPECIFIC PAEDIATRIC REFERENCE INTERVALS **86**

4.1 Introduction *86*

4.2 Publication *86*

4.3 Discussion *106*

CHAPTER 5: STATISTICAL METHODS USED IN THE ESTIMATION OF AGE-SPECIFIC PAEDIATRIC REFERENCE INTERVALS FOR LABORATORY BLOOD TESTS: A SYSTEMATIC REVIEW 107

5.1 Introduction *107*

5.2 Publication *107*

5.3 Discussion *137*

CHAPTER 6: COMPARISON OF STATISTICAL METHODS FOR CONSTRUCTING CONTINUOUS PAEDIATRIC REFERENCE INTERVALS: A SIMULATION STUDY **140**

6.1 Introduction *140*

6.2 Publication *141*

6.3 Discussion *168*

CHAPTER 7: DISCUSSION & CONCLUSION **170**

7.1 Summary of findings *170*

7.2 Translation of research findings into practice *172*

7.3 Further research addressing the gaps and challenges *176*

7.4 Conclusion *177*

REFERENCES **179**

APPENDICES **188**

Appendix A: Supplementary documents for the publication included in Chapter 2 (Hoq *et al*. A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: the HAPPI Kids study protocol. *BMJ Open*. 2019; 9(4):e025897) 188

Appendix B: Supplementary documents for the publication included in Chapter 3 (Hoq *et al*. Reference values for 30 common biochemistry analytes across five different analyzers in neonates and children 30 days to 18 year olds, Clinical Chemistry. 2019; 65(10): 1317 - 1326.) 190

Appendix C: .... Suplementatry documents included in Chapter 4 (Hoq *et al*. Validation of the HAPPI Kids Continuous Age-Specific Pediatric Reference Intervals. The Journal of Applied Laboratory Medicine. 2020: jfaa045) 241

Appendix D:. Supplementary documents of the publication included in Chapter 5 (Hoq *et al*. Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review. *Clinical Biochemistry*. Aug 2020.) 271

Appendix E: . Supplementary documents of the publication included in Chapter 6 (Hoq *el al*. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. Biometrical Journal. 2020. (Submitted)) 333

Appendix F: Author accepted manuscript published in Clinical Biochemist reviews (Hoq *et al*. Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations. *Clinical Biochemist Review*. 2020; 41 (2).) 352

List of Figures

Figure 1-1: Distribution of test results of patients with or without the disease [5] 2
Figure 1-2: Changes in serum creatinine with age and sex, adapted from Colantonio *et al* [15] 4
Figure 7-1: Visual comparison of trends in reference values over time between discrete age-group and continuous age-specific RIs. 175

Abbreviations

ALT	Alanine aminotransferase
ALP	Alkaline phosphate
AST	Aspartate aminotransferase
AACB	Australasian Association of Clinical Biochemists
AHRIP	Australasian Harmonised Reference Intervals for Paediatrics
CALIPER	Canadian Laboratory Initiative on Paediatric Reference Intervals
CHILDx	Children's Health Improvement through Laboratory Diagnostics
CDL	Clinical Decision Limit
CLSI	Clinical Laboratory Standards Institute
CI	Confidence Interval
GGT	Gamma-glutamyl transferase
GAMLSS	Generalised Additive Models for Location Shape and Scale
HAPPI Kids	Harmonising Age Pathology Parameters in Kids
C-RIDL	IFCC Reference Intervals and Decision Limits Committee
IVDs	In-vitro diagnostics
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
ISO	International Organisation for Standardisation
KiGGS	The German Health Interview and Examination Survey for Children and Adolescents
LIS	Laboratory information systems

LDH	Lactate dehydrogenase
LMS	lambda, mu and sigma
LOOK	Lifestyle Of Our Kids
MAD	Median absolute deviation about the medians
NORICHILD	Scandinavian Initiative for the Establishment of Paediatric Reference Intervals
PRINCE	Paediatric Reference Intervals in China
RCPA	Royal College of Pathologists of Australasia
RI	Reference Interval
RR	Reference Range
SD	Standard deviation

Chapter 1: Introduction

1.1 Reference intervals and their clinical importance

Laboratory test results are critical to modern clinical medicine, with claims that nearly 70% of physicians' medical decisions are based on information provided by laboratory test reports [1]. The importance of laboratory test results is further evident from the sheer number of pathology tests performed. According to the Royal College of Pathologists of Australasia (RCPA), approximately 500 million laboratory tests are performed each year in Australia and more than 11 million Australians have at least one test per year [2].

A blood test result by itself is of little value unless it is reported with a reliable and appropriate clinically related interpretation. A reference interval (RI) is one such clinical assessment tool used frequently to interpret test results. Statistically, a RI is defined as a range within which (100 - α) % ($0 < \alpha < 100$) of the observations fall, where lower and upper limits are (α/2 and (100 - α/2) centiles respectively [3]. The interval between two extreme percentiles represents the central part of a defined reference population, in most cases consists of healthy individuals. The basic concept of a RI is that observations that are outside the interval are considered unusual or abnormal, with the understanding that this will falsely classify α% of the reference population as "unusual". Soldin *et al* further explained this concept arguing that usually there will be some overlap in the distribution of test results between diseased and non-diseased population (Figure 1-1) [4]. The classification of α% of the reference population as "unusual" will be of varying clinical value, depending on the distribution of the analyte in whatever disease population is of clinical interest.

The value α is arbitrary. However, 95% RIs are commonly used in which case 5% of the test results will be outside the 2.5^{th} and 97.5^{th} percentile. There is no clinical justification for

considering the central 95% test results as physiologically normal. The value of α could be 80%, 90% or even 99% based on the clinical interest specific to the analyte.

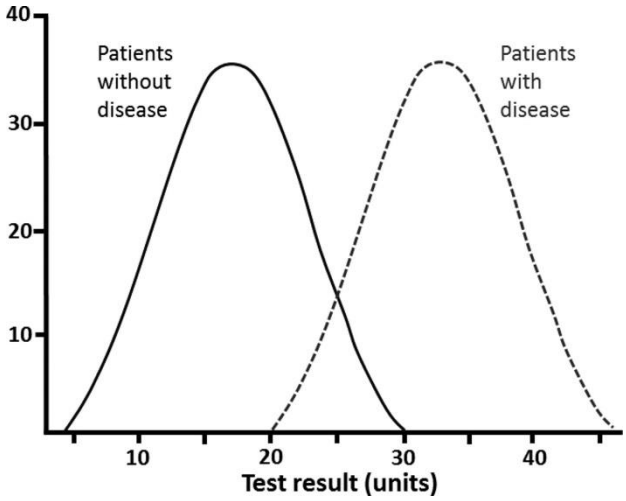


Figure 1-1: Distribution of test results of patients with or without the disease [5]

RI is also known as a reference range (RR). Unlike a range defined by the maximum and minimum values, a RI is defined by an interval between two percentiles. RI is sometimes confused with clinical decision limit (CDL). However, there is a clear distinction between the two assessment tools. Statistically, a RI refers to the distribution occurrence in a reference population, while a CDL indicates disease status associated with adverse clinical outcomes [6]. Clinically, a person with a test result outside a RI may or may not need a decision regarding clinical intervention while test results below or above a CDL depending on the analyte will need a decision regarding clinical intervention [6].

The appropriateness of any RI is a critical factor in the clinical interpretation of a test result and subsequent patient diagnosis and treatment. For this reason, the International Organisation for Standardisation (ISO) has specified requirements for all aspects of laboratory testing, including pre-analytical, analytical and post-analytical phases [7, 8].

According to these requirements, clinical laboratories need to periodically demonstrate the appropriateness of the RIs to the population served and the measurement system in use [7, 8] Hence, it is important that RIs are estimated with appropriate recognition of variation by age and sex for the population [9].

RIs are used in other disciplines of health for interpreting, for example, heart rates, respiratory rates, blood pressure, kidney and renal pelvis measurement, and bone turnover markers. The child growth curve is the most commonly used RI worldwide [10].

1.2 Paediatric reference intervals

In children, changes in specific analytes (e.g. creatinine and alkaline phosphatase (ALP)) are associated with normal physiological growth and development (Figure 1-2) [11-13]. These changes in analytes are often linked to the four stages of childhood development: neonatal, infancy, childhood and adolescence [14]. In the early stages of life, the developing organ systems contribute to a considerable change in the concentration of biomarkers. Similarly, the onset of puberty contributes to changes in analytes, which may start as early as eight years of age in females and around 10 to 12 years of age in males [14]. For example, serum creatinine concentrations are elevated in newborns due to the known contribution of maternal creatinine, but rapidly decline in the first month as renal function improves. Subsequently, the creatinine concentrations rise continuously from one month of age and differ by sex during the onset of puberty due to body growth and muscle mass. Hence, RIs must reflect these age-dependent variations from birth through adolescence. The age-specific changes is one of the most important factors in the interpretation of paediatric RIs [14].

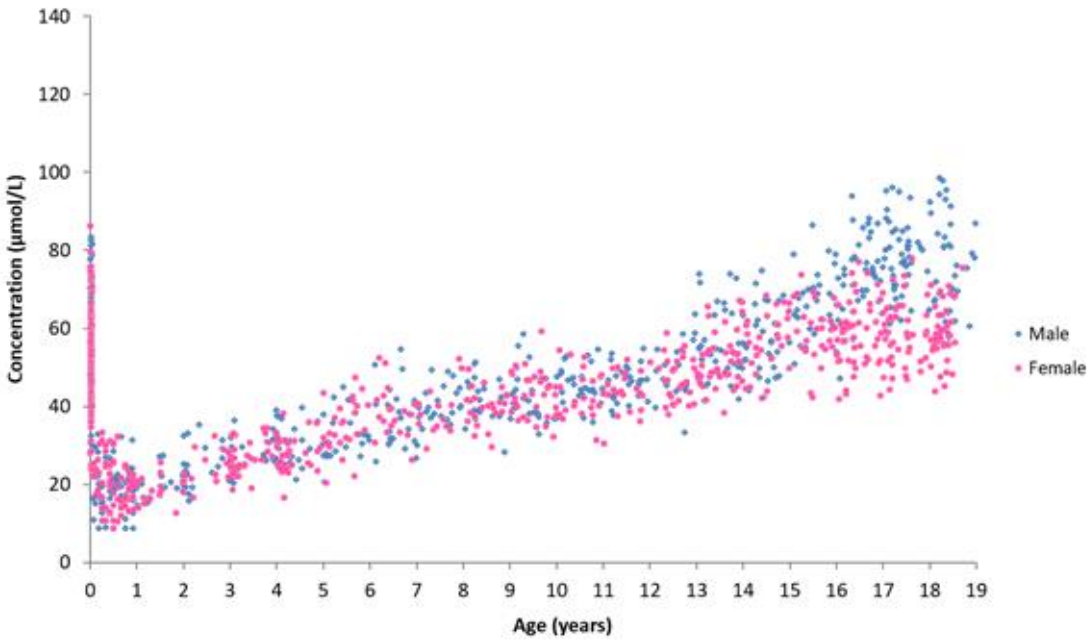


Figure 1-2: Changes in serum creatinine with age and sex, adapted from Colantonio *et al* [15]

1.3 Challenges in estimating paediatric reference intervals

The estimation of RIs, in particular paediatric RIs, is a complex and time-consuming process, [16]. The RI estimation process is broadly classified into five steps: (i) definition of a reference population; (ii) selection of the reference individual; (iii) collection of reference samples; (iv) laboratory testing of the reference samples according to routine, standardised operating procedures; and (v) application of statistical methods to the laboratory results generated in step (iv) [16]. For children, the estimation of RIs is difficult because it requires recruitment of adequate numbers of healthy children, collection of relatively large blood volumes proportional to the patient's blood reserve and an understanding of the physiological changes in analytes associated with normal growth and development [17-19]. These challenges are discussed below in brief.

1.3.1 Recruitment of healthy children

One of the main steps in establishing paediatric RIs is recruiting healthy children representative of the community that a specific laboratory or hospital serves [20]. Ideally, each child in the community should have the same probability of being selected as a representative of the reference population. Hence, the recruitment of children should occur prospectively from a community, having no current interaction with a health service.

The recruitment of children prospectively from a community is extremely difficult, both ethically and pragmatically. A reasonable compromise is to actively recruit otherwise healthy children who present to health care services for minor surgical procedures, who would usually not be having clinically investigative blood tests because they are assumed to be normal. From a neonatal perspective, healthy term newborns on the post-natal wards of hospitals represent an ideal sampling population. However, very low consent rates, restrictions in the amount of blood that can be safely collected from children with respect to body weight, a requirement for specialist phlebotomy procedures and staff combined with other ethical (e.g. single collection attempt) and logistical challenges (e.g. storage, transportation and cost), limit the ease with which many laboratories worldwide can attain prospective collection of paediatric blood samples [21, 22]. This is particularly true for smaller, non-specialised laboratories that service regional health centres.

An alternative approach has been to conduct retrospective studies, with data extraction, often referred to as “data mining”, from existing laboratory information systems (LIS) or testing leftover samples from laboratory storage [11, 12, 23, 24]. Data mining has the advantage of generating large datasets compared to the direct approach of obtaining and testing samples for the single purpose of establishing RIs [24]. Considering the high likelihood that some test

results will derive from individual or patients who have diseases, rigorous statistical analysis are conducted for removing outlier or selecting a subset of test results under the assumption that the individuals were healthy or without specific conditions. Hoffmann method, Bhattacharya method or the method developed by Arzideh are commonly used to estimate the RIs [23]. A detailed discussion of different methods and recommendations on indirect methods has been published by Jones *et al* [23]. However, despite the large sample size and rigorous statistical analysis, the representativeness of the findings, and thus generated RIs is influenced by the fact that the majority of the blood tests were initially conducted for investigation of possible disease states, an issue that may affect paediatric testing more so than adult testing [23, 24].

1.3.2 Analytical factors

Pre-analytical and analytical aspects of testing blood samples is another crucial factor in the estimation of paediatric RIs [19]. The highest source of error in the entire testing process is attributed to the pre-analytical phase, which involves biological (e.g., fasting status, recent exercise, diet) and methodological factors (e.g., type of sample container, preservatives in the sample container, length of time between sample collections, how sample is collected, and storage conditions) [19]. Hence, to reduce error in laboratory testing due to pre-analytical issues, blood samples from the reference sample should be tested according to the same procedure used for testing in general [19].

Biological variation, physiological changes that occur within an individual and between individuals, is another critical factor in interpreting the test results and in the estimation of paediatric RIs [25]. Biological variation provides information on the clinical significance of a change in the concentration of a specific biomarker [25]. The information can also be used to

establish a quality specification and provides insights regarding the usefulness of RIs for different analytes [25]. The biological variation and total error acceptable for different analytes reported in the Westgard database, mostly determined in adults, are used as a quality indicator worldwide [26]. In Australia, the Royal Australasian College of Pathologists (RACP), has also defined the allowable limits of performance for several general serum chemistry analytes based on a set of analytical quality specification [27]. However, only a few studies have reported biological variation for serum biochemistry analytes specific for the paediatric population [25, 28].

Despite recent activities to achieve standardisation of laboratory results, between-method differences remain for the same tests [29-31]. As sequential paediatric specimens on the same patient are often tested by different laboratories, the application of appropriate reagent and analyser-specific RIs for reliable clinical decision is imperative. Most importantly, the clinical implication of any inter-analyser variation must be understood when making a decision based on analyser-specific RIs [32].

There are several initiatives to harmonise paediatric RIs used between different laboratories and/or different analysers. The decision to harmonise is often based on clinical consensus between pathologists [33] or analysing secondary data applying Bhattacharya analysis [34, 35], a method for isolating Gaussian distribution in the presence of other data. However the decision to harmonise should be based on a clinical evaluation of test results that were outside the RIs [25].

1.3.3 Changes in analytes with growth and development

Traditionally, RIs were estimated for different discrete age-groups to reflect the age-specific changes in biomarkers with growth and development [16, 32]. However, several studies have identified that discrete age-group RIs are misleading, specifically when testing children whose age is close to the defined age-group cut-off values [11, 12]. For example, the continuous increase of creatinine concentration with increased muscle mass (and therefore age) or increase in ALP associated with the osteoplastic activity during the pubertal growth spurt, make it evident that discrete age-group RIs alone are not sufficient to capture these physiologically based changes [12, 36]. An additional challenge in estimating age and sex-specific paediatric RIs is that a large sample size is needed to enable reliable estimation of changes with age.

1.4 Current status of paediatric reference interval initiatives

1.4.1 Population-based, prospective paediatric studies

The quality of RIs depends on the selection and recruitment of the individuals chosen to provide a specific RI, the control and quality of the laboratory's pre-analytical, analytical and post-analytical process, and the statistical methods applied. While controlling all these aspects has been challenging, several national and international initiatives have demonstrated that estimating paediatric RIs using samples collected prospectively from population-based studies is feasible. According to the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), several countries and regions have taken initiatives to estimate paediatric RIs [37]. In Germany, the German Health Interview and Examination Survey for Children and Adolescents (KiGGS) presented the age and sex-specific percentile distribution of key laboratory parameters based on more than 14,000 blood and serum samples [38]. In the USA, the Children's Health Improvement through Laboratory Diagnostics (CHILDx) study is

focusing on estimating paediatric RIs for several clinical laboratory assays by collecting both blood and urine specimens from healthy subjects [39]. In Scandinavian countries, the Scandinavian Initiative for the Establishment of Paediatric Reference Intervals" (NORICHILD) has reported discrete age-group paediatric RIs based on blood collected from nation-wide community surveys [40].

The Canadian Laboratory Initiative on Paediatric Reference Intervals (CALIPER) is one of the largest paediatric RIs project completed to date, and its numerous publications of paediatric RIs have contributed significantly to improving the clinical diagnosis and interpretation of paediatric test results [32]. CALIPER has collected thousands of samples from healthy children in Canada across several ethnic groups and has estimated paediatric RIs for several analytes on Abbott Architect analysers [15]. CALIPER has also mathematically transferred the estimated paediatric RIs to other analytical platforms except for immunoassays, published a systematic review of the statistical methods applied in the estimation of RIs, and performed a simulation study comparing statistical methods for estimation of discrete age-group RIs [41-43].

Addressing the lack of accurate paediatric RIs in China, the Paediatric Reference Intervals in China (PRINCE) project aims to establish and verify paediatric RIs for 31 analytes using data collected from approximately 15,000 healthy children from 10 children's hospitals [44]. To date, the project has evaluated the feasibility and efficiency of the method using data from 602 children [44]. The project has also conducted a study comparing different statistical methods for estimating continuous age-specific RIs [45].

In Australia, the Lifestyle Of Our Kids (LOOK) study, a community-based longitudinal study, followed a cohort of 852 healthy eight-year-old children and reported RIs for 37 analytes at age 8, 10 and 12 years [46].

1.4.2 Analyser-specific or common paediatric reference intervals

Addressing inter-analyser variation, the CALIPER study mathematically transferred and validated discrete age-group paediatric RIs from Abbott Architect analysers to other systems including the Beckman Coulter DxC800, Ortho Vitros 5600, Roche Cobas 6000, and Siemens Vista 1500 analysers for common biochemical and immunological analytes [15, 43]. The authors reported that the transferred RIs were similar for most analytes to the RIs estimated on the Abbott Architect c800, while in some cases (e.g. lipase on Siemens Vista and apoliopoprotein B on Ortho Vitros) the reference limits were higher than those for the Abbott Architect [43]. The Australasian Association of Clinical Biochemists (AACB) and RCPA endorsed harmonised paediatric RIs for 9 biochemical analytes (sodium, serum and plasma potassium, chloride, bicarbonate, creatinine, calcium, phosphate, magnesium and ALP for children aged 0 week to < 18 years), based on the paediatric data from 15 laboratories (i.e. over 200,000 test results) using Bhattacharya analysis [31, 34]. However, the comparison of analytical methods to estimate variability between manufacturers was performed using adult samples and may not be representative of the paediatric population, particularly for neonates [31].

1.4.3 Discrete age-group or continuous age-specific reference intervals

Considering the challenges in clinical diagnosis based on the discrete age-group RIs for biomarkers that change with age, several studies have estimated continuous age and sex-specific paediatric RIs [11, 12, 36, 47]. Bussler *et al* estimated continuous RIs for alanine

aminotransferase (ALT), aspartate aminotransferase (AST), and gamma-glutamyl transferase (GGT) as 3rd and 97th percentiles using a widely used semi-parametric statistical method ("LMS"), based on data from 1746 healthy children (3131 cases) aged 11 months to 16 years [47]. Zierk *et al* reported continuous age and sex-specific RIs for 22 haematological and biochemical analytes also using an LMS-type method based on data available from a hospital laboratory database for children aged 6 months to 18 years [12]. Recently, CALIPER reported continuous RIs for 38 biochemical markers using non-parametric quantile regression based on the data from the CALIPER database for children aged 6 months to < 19 years [36].

More details on different national and international initiatives and approaches have been published earlier in 2015 [48].

1.5 Statistical analysis of the laboratory results to estimate reference limits

Statistical analysis of the laboratory results is the final critical step for obtaining paediatric RIs. The analysis process can be broadly divided into four steps consisting of identifying outliers, assessing the distribution of reference values, estimating the reference limits, and calculating the 95% confidence interval (CI) of the reference limit. These steps are briefly discussed below and commonly used statistical methods in each step of estimating age-specific RIs are reported in Chapter 5 along with my recommendations.

1.5.1 Identifying outliers

It is assumed that the reference values come from the same underlying population of test results. However, in some instances a few values that lie outside the range of the reference values, may arise from a different population of test results with an underlining condition or due to the analytical issues, are easily identified as "Outliers" [16].

Identification and treatment of outliers is an integral part of data processing and has been used for centuries [49]. It is also an essential step in the estimation of RIs as it removes the effect of the outlier on the estimated reference limits. Horn *et al* (2001) revealed that the RIs were on average 10% narrower when outliers were removed compared to when outliers were not removed [50].

There are several statistical methods available for identifying outliers. However, the Clinical Laboratory Standards Institute (CLSI) recommends tests proposed by Dixon and Reed, and Tukey [16]. Both methods assume data follow a normal distribution and identify outliers based on the distance from other values [16].

1.5.2 Assessing the distribution of the reference values

For most statistical methods, one of the main underlying assumptions is that data are normally distributed. However, there are alternative methods that do not require data to be normally distributed. Hence, distribution of the data should be assessed before applying statistical methods that require data to be normally distributed.

Several formal statistical tests of univariate normality are available for assessing the distribution of the reference values [51]. However, a visual inspection of the frequency distribution is useful in examining the distribution.

1.5.3 Estimating reference limits

There are two well-known approaches of estimating age-specific RIs. One approach has adopted age partitioning to obtain RIs for discrete age groups. However, there are no well-defined statistical methods for partitioning age into small subgroups. Traditionally, the

reference values are divided into age-groups of 6 to 12 months or 2 to 5 years, and mean reference values for these groups are estimated. The subgroups are then combined if the difference in mean is not statistically significant i.e. 95% CI includes 0 or p value <0.05. Commonly used statistical methods for comparing mean, median or distribution such as t-test, Kruskal-Wallis test, Mann-Whitney test etc. were used. However, for large sample any difference will be statistically significant regardless of its clinical significance. Considering the challenges in partitioning reference values into subgroups of age, the CLSI guidelines recommended that existing literature i.e. pertinent physiological information and potential usefulness of the subclass RIs in clinical practice should be considered [16]. Another approach has considered age as a continuous variable to obtain reference curve.

There is a wide variety of statistical methods available for estimation of reference limits depending on the approach. Wright and Royston (1999) provided a brief review of these methods and discussed issues related to goodness-of-fit, sample size, and confidence bands [52]. Based on a systematic review of 22 articles, Daly *et al*. (2013) reported that statistical methods such as nonparametric, parametric, robust, and semi-parametric curve-fitting approaches are the commonly used methods in the estimation of paediatric RIs [41]. A brief description of these statistical methods used in the estimation of paediatric RIs for laboratory measurement is provided below.

Estimating a reference limit

Nonparametric method: To standardise the process, the CLSI has recommended a nonparametric method for estimating RIs due to the simplicity of the method and that it avoids the need for an assumption about the distribution of the values in the population [16]. The nonparametric method involves ordering the values from smallest (1) to largest (n) and

simply calculating *0.025(n+1)th and 0.975(n+1)th* ranked observation. Since, in most cases, the values of these two ranks are non-integers, the two limits are estimated by interpolating between the data points corresponding to the ranks on either side. A minimum sample size of 120 is recommended for this method, in which case the lower and upper limits will be the 3rd and 118th rank observations, respectively [16, 20]. However, the estimates are sensitive to outliers.

Kernel method: Addressing large variation in the estimates determined by the nonparametric method, a weighting function called a kernel is applied to all of the quantiles in the sample to produce a weighted average quantile as the new estimator [20]. The kernel function is expressed as the expected values of the *i*th order statistics, $\zeta_\rho = \sum_{i=1}^{n} W_{n,i}(p)x_i$ where the weights are controlled by smoothing parameters and $\sum_{i=1}^{n} W_{n,i}(p) = 1$ [20]. A kernel estimator is a smooth device whereby the estimate at a specified percentile of the population values is a weighted average of all the values observed [20]. Compared to the nonparametric method, this method is rarely used in recent times [41].

Parametric method: Where reference values follow a normal distribution, the 2.5th and 97.5th percentiles may be estimated by subtracting and adding 1.96 times the standard deviation (SD) from the mean, respectively [16]. Nevertheless, in most cases, the analytes measured in clinical laboratories do not follow a normal distribution[16]. In such cases, simple *log* transformation could convert the skewed distribution to approximately normal distribution [16]. Box-Cox transformation has also been used to transform skewed distribution to approximately normal distribution [16].

Robust method: The CLSI's recommendation for a nonparametric test is 120 samples. However, 120 samples are not readily available in all cases for a specific group. In such cases, when the sample size is fewer than 120 samples, a robust method is recommended as a potential alternative [16]. The robust method is an iterative process. Initially, the median (T_{bi}) and median absolute deviation about the medians (MAD) are calculated using the actual observations. Subsequently observations down-weighted according to their distance from the median, and median and MAD are calculated using the weighted observations. This process continues until the change in median in the consecutive iterative process is minimal [16]. With median calculated, the 95% RIs are calculated using $T_{bi} \pm t_{n-1}^{0.025}\sqrt{S}$, where t is the upper 2.5th percentile of a student's t with $(n-1)$ degrees of freedom, n is the sample size and S is a the bi-weight estimator for range of tuning constant and variability of $T_{bi.}$ [16].

Estimating continuous age-specific reference limits:

Healy et al method.: Healy *et al* described a nonparametric approach to estimating continuous age-specific RIs, based on the assumption that the centiles of the distribution at any age will be a smooth function of age and the distribution as a whole will change slowly and smoothly with age [20, 52]. According to the method, reference limits are estimated using a nonparametric method for a series of overlapping age-groups [20, 52]. The centiles are then modelled by a polynomial of degree p, where coefficients are a function of a standard normal variable corresponding the ith centile of another polynomial of order q [20, 52]. An advantage of this method is that age is considered as a continuous variable. However, selecting the appropriate polynomial of degree, p requires experience [20, 52].

Cole's lambda, mu and sigma (LMS) method: For this method, age is divided into integer ages (e.g. 1, 2, 3, …). For each age-group, the parameters of the Box-Cox transformation i.e. the skewness (L), median (M) and coefficient of variation (S) of a measurement are estimated. These estimates are then smoothed by using any regression smoothing procedure e.g., polynomial regression or kernel estimation or cubic splines. However, there are no clear guidelines regarding model selections [53]. Centiles of the transformed values are calculated and back-transformed to the original scale [53]. A modification to the method was suggested by Cole and Green (1990), where the parameters were estimated as a smooth function of continuous age using cubic splines removing the need for the partition of age into small groups [54].

Generalised Additive Models for Location Shape and Scale (GAMLSS) method: Rigby and Stasinopoulos (2005) proposed a new method, GAMLSS, which might be considered a modification of the LMS method. The GAMLSS method can be applied for any distribution of the measurement and the parameters of the distribution (i.e. mean, variance, skewness and kurtosis) can be modelled as a linear, nonlinear or smooth function of continuous age [55, 56].

Royston's method: Royston (1991) recommended lower order polynomial regression for estimating continuous age-specific RIs based on the assumption that data are normally distributed, or can be easily converted to normality by logarithmic transformation at each value of age [3, 57]. According to this method, the mean reference value is modelled as a low order polynomial (linear, quadratic or cubic) function of age, and RIs are constructed by adding or subtracting a multiple of the standard deviation, which is assumed to be constant with age[57]. In 1998, Royston extended this method and proposed fractional polynomials of

age to model the mean reference values for more complex relationships between the biomarker and age, and to model absolute residual for biomarkers with SDs of reference values that are dependent on age [58].

Quantile regression: In recent years, quantile regression has been applied in the estimation of RIs. Gannoun (2002) proposed a non-parametric quantile regression for estimating continuous age-specific RIs [59]. Wei (2006) demonstrated how quantile regression based on spline functions could be used in estimating the child growth chart [60]. Cameron and Trivedi applied quantile regression where the conditional distribution of the quantile is expressed as a linear function of age and sex [61].

1.5.4 Calculating a confidence interval for the limits

A confidence interval (CI) for the reference limit indicates the reliability of the estimated reference limit. It is a quantitative measure of the variability of estimates under the assumptions that the reference samples are selected randomly either from the community or from a described pool and independent of each other [16]. According to Harris and Boyd (1995), the width of the confidence interval should be acceptably small relative to the width of the reference limit [16]. For a discrete age-group RI, their recommended width of the 90% CI is 0.2 times the width of the 95% RI [16]. For a continuous age-specific RIs, Royston (1991) suggested that the standard error of the limit of the RIs should be no more than 10% of the residual SD from the parametric model and demonstrated how to determine a sample size that would achieve the desired precision [57].

1.6 Validation and transference of estimated paediatric RIs

Considering the challenges in establishing paediatric RIs as discussed in section 1.3, the process is likely to be practically limited to large scale paediatric specialist centres with dedicated research resources or multi-national diagnostic manufacturers. Laboratories that analyse a small number of paediatric specimens amongst predominantly adult cohorts and/or smaller regional laboratories may only be able to adopt previously published paediatric RIs. Hence, the verification and validation of RIs in paediatric populations play an important role in ensuring appropriate RIs are in use at local laboratory sites.

To assess the validity of estimated discrete age-group RIs in a laboratory, the CLSI recommends a less costly and more convenient process [16]. Under the assumption that the test subject population and analytical system are comparable, the acceptability of the RIs is assessed by testing samples from 20 children [16]. The estimated discrete age-group RIs are considered valid for use in the laboratory if 18 or more test results fall within the interval [16]. If the estimated RI of any age-group is valid, the estimated RIs for all age-groups for the biomarkers are considered valid.

The CLSI guidelines for validation are appropriate for the validation of discrete age-group RIs [16]. Considering several studies have published continuous age-specific paediatric RIs, there are no current recommendation or endorsed guidelines for the validation of continuous age-specific RIs by a routine laboratory. So, it is uncertain to what extent the existing guidelines apply to continuous age-specific RIs.

To transfer the estimated RIs between two different analytical methods or analysers in a laboratory, CLSI recommends a linear regression model is fitted for the test results from the

new analytical methods or analysers based on the test results from the existing analytical methods or analysers [16]. The fitted model then can be used for estimating the RIs of the new analytical methods or analysers based on the existing RIs.

1.7 Synopsis of the research problem

Reliable and appropriate RIs are essential for correct interpretation of laboratory tests, which in turn is increasingly important for improving the quality of clinical assessment and care of children [15]. In Australia, the AACB has defined harmonised RIs for selected analytes in children [31]. However, there is inadequate uptake of these RIs for routine pathology analytes in children [62].

The process of estimating accurate and reliable RIs is complex, especially for children [15]. Several issues need to be addressed to create reliable, accurate RIs for children. First, there is a need for adequate representation of the general population in the reference population. Second, biomarkers measured from blood samples (such as creatinine and alkaline phosphatase) are influenced by the development, and growth of a child. To address this issue, RIs should be viewed as continuously changing with age, which creates a potential statistical challenge. Third, laboratory tests are instrument dependent and variation between different analysers has been documented in several studies. However, no comparison of RIs by analysers for the same reference population has been reported. Fourth, the best statistical methods for estimating and reporting continuous age-specific RIs are still not widely agreed upon. In biochemistry, observed $\alpha/2$ and $(100 - \alpha/2)$ percentiles are frequently used as the RIs, whereas various more sophisticated curve-fitting approaches are applied in anthropometry and other disciplines of clinical medicine. And last, there is no available guideline for laboratories to validate the estimated continuous RIs for their use.

Hence, the four main knowledge gaps that this project aimed to address are: 1) the lack of appropriate and reliable continuous age-specific paediatric RIs in Australian population for common biochemistry analytes for commercially available analysers, 2) an absence of head-to-head comparison of laboratory results on different analysers using aliquots from the same blood sample from a healthy participant, 3) the lack of clear evidence and accompanying guidelines regarding the best statistical methods for constructing continuous age-specific paediatric RIs under various scenarios, and 4) the lack of guidelines for validating continuous age-specific RIs.

1.8 Research aims and objective

Aim: Evaluation and application of methods for estimating continuous age-specific RIs for laboratory blood tests in children.

Specific objectives:

- To estimate and report continuous age-specific paediatric RIs in Australian population for common biochemistry analytes including direct head-to-head comparison of reference values on five major analysers.
- To validate the estimated continuous age-specific paediatric RIs for use in a routine laboratory.
- To investigate the range of current statistical practices in this area and how they have changed over recent decades, if at all.
- To evaluate and compare statistical methods for constructing continuous age-specific paediatric RIs for common biochemistry analytes.

1.9 Thesis structure

Chapter 2 describes the Harmonising Age Pathology Parameters in Kids (HAPPI Kids) study protocol and highlights the importance of the prospective collection of blood samples from healthy children. The chapter describes the sample collection, handling, and processing, which are crucial for the reliability of the estimated RIs and generalisability of the findings. An examination of what the sample size implied for the variability of the estimated RIs is provided.

Chapter 3 describes a novel statistical approach for estimating continuous age-specific RIs and reports the RIs as a function of age and sex for 30 biochemistry analytes in children using data from the HAPPI Kids study. The chapter also reports the head-to-head comparison of reference values across the five major biochemistry analysers and provides a quantitative and qualitative measure of the extent to which the reference values vary across five analysers.

Chapter 4 describes the process of validation for the continuous age-specific RIs based on some modification to the validation process recommended by the CLSI. The results of the proposed validation method are applied to assess validation (in a routine lab) of the continuous age-specific RI for 18 analytes established in chapter 3.

Chapter 5 reports the findings from a systematic review of original research publications from the past 25 years estimating paediatric RIs, with a focus on the statistical methods applied. This chapter highlights whether the statistical methods applied in different steps of the estimation process have been reported, the choice between estimating discrete age-group or continuous age-specific RIs and its implication for the statistical methods applied, the trend in practices and emerging methods, and the sample sizes used for different methods.

Chapter 6 reports the findings from a simulation study that aimed to provide guidance on the most appropriate statistical methods for estimating continuous age-specific paediatric RIs under different scenarios. Four statistical methods for estimating continuous age-specific RIs were compared based on root mean-squared error using hypothetical true values based on multiple scenarios including different sample sizes. This chapter reports the findings of the simulation study and justifies the novel approach and sample size used in the estimation of the HAPPI Kids continuous age-specific paediatric RIs.

Chapter 7 provides a detailed discussion of the key findings of the research project including challenges and future priorities.

Chapter 2: Study design, data collection method and analysis plan of the Harmonising Age Pathology Parameters in Kids (HAPPI Kids) study

2.1 Introduction

Previously documented paediatric RI studies have frequently been based on the retrospective analysis of laboratory data or paediatric hospitals data [11, 12]. The retrospective collection of data has been justified since collecting blood samples from children is challenging, the blood volume that can be safely collected from children is small, and ethical concerns surrounding obtaining blood from healthy children [23, 24]. However, concerns have been raised regarding the generalisability of the RIs estimated using laboratory results from children who were initially referred for clinical investigation [23, 24]. Addressing this, several studies have demonstrated the feasibility of collecting blood samples prospectively from children in the community [32, 38, 40]. However, there has been no RI study in Australia which collected blood samples from children aged 0 – 18 years in the community. Addressing this gap, the HAPPI Kids study collected blood samples from healthy children in the hospital outpatient services in Victoria, Australia.

Considering the changes in biomarkers with age, several studies have estimated RIs for discrete age groups while a few studies estimated RIs for continuous age. While CLSI recommend nonparametric method for discrete age-group RIs, there are no guidelines for statistical methods for estimating continuous age-specific RIs [16]. Coles LMS, GAMLSS and Royston's methods are commonly used. However, all three methods rely on transformation to normal distribution and require estimating parameters (e.g. mean, median, SD, etc) of the transformation to calculate reference limits [55, 58, 63]. The integration of continuous RIs in the existing LIS has also been challenging since except for Royston's method Coles LMS and GAMLSS do not provide any equation which could be used for

calculating RIs for any age. Addressing these challenges while analysing the HAPPI Kids data our proposed novel statistical method appeared to be effective as the method does not require the data to be normally distributed and expresses the continuous age-specific RIs as functions of age. Our method also appeared worth comparing with other commonly used methods and motivated us to do a simulation study.

This chapter describes the HAPPI Kids study protocol, highlighting a) the feasibility of prospective collection of blood samples from healthy children, b) the sample collection and testing procedure, c) the rationale for the sample size, and d) the proposed novel statistical approach for the estimation of continuous age-specific RIs. The study protocol was published in BMJ Open in April 2019 and the author approved manuscript is presented as the body of this chapter [64].

This chapter highlights the biochemistry analytes for which i) sample aliquots from the same healthy participants were tested for the same analyte using different automated analysers, ii) sample aliquots of the same sample were tested in a routine laboratory using the same analyser type and test method and c) samples from pre-term children were tested. By the end of 2018, analysis of samples from 30 biochemistry analytes was completed and continuous age-specific RIs for those analytes are reported in chapter 3. In addition, a routine laboratory completed analysis of samples from children (not included as part of the original cohort) for 18 of the 30 biochemistry analytes. The test results for those 18 analytes were used for validation of the established continuous age-specific RIs and are reported in chapter 5.

2.2 Publication

A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: The HAPPI Kids study

Corresponding Author

Associate Professor Vera Ignjatovic, PhD, B.Sc. (Hons)

Co-Group Leader, Haematology Research, Murdoch Children's Research Institute

Principal Fellow, Department of Paediatrics, The University of Melbourne

Address

Haematology Research, Murdoch Children's Research Institute

The Royal Children's Hospital

50 Flemington Road, Parkville, Victoria, 3052, Australia

Tel: +61 3 99366520, Email: vera.ignjatovic@mcri.edu.au

Co-authors

Monsurul Hoq

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, Australia; Department of Paediatrics, The University of Melbourne, Parkville, Australia.

Email: monsurul.hoq@mcri.edu.au

Vicky Karlaftis

Haematology Research, Murdoch Children's Research Institute, Parkville, Australia.

Email: vasiliki.karlaftis@mcri.edu.au

Sue Mathews

Department of Biochemistry, Laboratory Services, The Royal Children's Hospital, Australia.

Email: Susan.Matthews@rch.org.au

Janet Burgess

Department of Pathology Collection, Laboratory Services, The Royal Children's Hospital, Parkville, Australia.

Email: Janet.Burgess@rch.org.au

Susan Donath

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute; Parkville, Australia; Department of Paediatrics, The University of Melbourne; Parkville, Australia.

Email: Susan.donath@mcri.edu.au

John Carlin

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, Australia; Department of Paediatrics, The University of Melbourne, Parkville, Australia.

Email: john.carlin@mcri.edu.au

Paul Monagle

Haematology Research, Murdoch Children's Research Institute, Parkville, Australia; Department of Paediatrics, The University of Melbourne, Parkville, Australia.

Email: Paul.Monagle@rch.org.au

Vera Ignjatovic

Haematology Research, Murdoch Children's Research Institute, Parkville, Australia;

Department of Paediatrics, The University of Melbourne, Parkville, Australia.

Email: vera.ignjatovic@mcri.edu.au

WORD COUNT: 3746

ABSTRACT

Introduction: The clinical interpretation of laboratory tests is reliant on reference intervals. However, the accuracy of a reference interval is dependent on the selected reference population, and in paediatrics, the ability of the reference interval to reflect changes associated with growth and age, as well as sex and ethnicity. Differences in reagent formulations, methodologies, and analysers can also impact on a reference interval. To date, no direct comparison of reference intervals for common analytes using different analysers in children has been published. The HAPPI Kids study aims to establish age-appropriate reference intervals for commonly used analytes in the routine clinical care of neonates and children, and to determine the feasibility of paediatric reference interval harmonisation by comparing age-appropriate reference intervals in different analysers for multiple analytes.

Methods and analysis: The HAPPI Kids study is a prospective cross-sectional study, collecting paediatric blood samples for analysis of commonly requested biochemical, immunological and haematological tests. Venous blood samples are collected from healthy premature neonates (32 to 36 weeks gestation), term neonates (from birth to a maximum of 72 hours post-birth), and children aged 30 days to ≤ 18 years (undergoing minor day surgical procedures). Blood samples are processed according to standard laboratory procedures and, if

not processed immediately, stored at – 800C. A minimum of 20 samples is analysed for every analyte for neonates and then each year of age until 18 years. Analytical testing is performed according to the standard operating procedures used for clinical samples. Where possible, sample aliquots from the same patients are analysed for an analyte across multiple commercially available analysers.

Ethics and dissemination: The study protocol was approved by The Royal Children's Hospital, Melbourne, Ethics in Human Research Committee (34183 A). The study findings will be published in peer-reviewed journals and shared with clinicians, laboratory-scientists and laboratories.

Keywords: paediatrics, neonates, children, reference intervals

STRENGTHS AND LIMITATIONS OF THIS STUDY

1. The prospective design of the study ensures blood samples are collected from healthy neonates and children aged 30 days to ≤ 18 years.
2. Aliquots from the same blood sample are tested for common analytes on multiple commercially available analysers allowing direct head to head comparison.
3. Samples are obtained from a minimum of 20 neonates and then for 20 children within each year of age until 18 years with equal numbers of males and females (total 380 samples per analyte).
4. Parametric statistical methods are to be applied for estimating reference intervals that vary continuously with increasing age.
5. Changes in analytes associated with age are based on samples collected cross-sectionally not based on individuals followed longitudinally.

INTRODUCTION

Reference intervals are important for the correct interpretation of clinical laboratory tests and affect the clinical assessment and care of patients. Reference intervals are used for determining normal versus abnormal for a measurement of interest, and are usually defined as a range encompassing 95% of the reference population, where lower and upper limits are 2.5th and 97.5th centile respectively [1].

Obtaining accurate data from a suitable reference population is an essential element in determining reference intervals, and therefore the characteristics of a reference population need to be matched to a "normal" or a healthy population using strict criteria. However, collecting blood samples from children for different laboratory tests is challenging due to the small blood volume that can be safely collected from children and the ethical concern of obtaining blood from healthy children [2]. For these reasons, previously documented paediatric reference interval studies are frequently based on retrospective analysis of laboratory data, prospective sample collection from school children in the community (i.e. limited age groups) or from paediatric hospitals where children are undergoing minor elective surgery [3-7]. Regardless of how the samples are collected, very few of these studies have included neonates through to 18 years of age [8].

In children, development and growth are among the key challenges in determining reference intervals for different analytes [2]. The dynamic nature of different analytes as children grow and develop influences the representation of paediatric reference intervals, which have often historically been established or partitioned for different age-ranges [8, 9]. This approach can often lead to misdiagnosis, especially when the age of the child that is being investigated is in close proximity to the borders of the partitions. For example, a child aged three years and 360

days may have clinical results that would be classified as outside the reference interval of 0 to 4 years for a blood test of interest, but within the reference interval for children aged 4 to 9 years. Hence, the difference of 5 days results in the test result being labelled as abnormal. Misdiagnosis leads to repeated blood tests for the child, increased utilisation of health services, and is often associated with anxiety for parents and children alike. The determination of the age-specific partition points has often been arbitrary and often there is not an even distribution of the number of reference subjects included across different partitions. For these and other reasons, the process of determining age-specific reference intervals for different analytes based on the artificial partitioning of age has been found to be problematic [4]. A preferred approach, which addresses the age-related changes in analyte levels, is to construct age-appropriate reference intervals considering age as a continuous variable [1, 4, 10].

The laboratory test results for most analytes are dependent on the analyser and reagents used for analysis and therefore, analyser- and reagent-specific reference intervals should be determined. In several studies, mathematical algorithms or “transference” mechanisms have been applied to compare and construct reference intervals for different analysers based on the reference values for one specific analyser [11, 12]. However, to date, direct comparison of the reference intervals for common analytes using different analysers has not been undertaken in neonates and children.

The Harmonizing Age Pathology Parameters in Kids (HAPPI Kids) study is designed to address the existing gaps in reference interval studies with prospectively collected blood samples from healthy neonates and children, from birth until 18 years of age.

METHODS AND ANALYSIS

The HAPPI Kids study: Aims

- Establish age-appropriate reference intervals for commonly used analytes in clinical practice in the setting of neonates and children.
- Determine whether analyser characteristics alter age-appropriate reference intervals for multiple analytes, in order to determine whether paediatric reference intervals can be harmonised.

Study design

The HAPPI Kids study is a prospective cross-sectional study for the collection of paediatric blood samples for commonly requested biochemical, immunological and haematological analysis.

Study subjects

The study subjects consist of healthy premature neonates (32 to 36 weeks gestation), term neonates (from birth up to a maximum of 72 hours post-birth and children aged 30 days to 18 years (undergoing minor day surgical procedures).

Participant recruitment and sample collection

Participant recruitment commenced in February 2015 in four major public hospitals in Melbourne, Australia. Parents or guardians of eligible study participants are approached by Pathology Collection staff at each site and written consent is obtained following a verbal and written explanation of the study and participation requirements. Following patient consent, the Pathology Collection staff interview the parents or guardian using the study questionnaire (Supplementary Table 1 of Appendix A) to confirm study inclusion and exclusion criteria

status (Table 1). In addition, the child's medical record is reviewed to assess and document the child's general health.

Three distinct routes of patient recruitment are utilised:

1. Pre-term neonatal samples are collected from "healthy" neonates about to be discharged from post-natal wards, born at 32 to 36 weeks gestation, at The Royal Women's Hospital. Samples are collected in the first three days of life via direct venepuncture, using a 23-gauge needle, into required blood collection tubes. The pre-term neonates do not have any systemic abnormalities i.e. underlying diseases and comorbidities, abnormal foetal monitoring and do not require any mechanical ventilation.
2. Term neonatal samples are collected from "healthy" term neonates about to be discharged from post-natal wards following routine intramuscular administration of one milligram of vitamin K in the delivery suites at The Royal Women's Hospital, Northern Health and Western Health - Sunshine Hospital. Samples are collected from birth up to a maximum of 72 hours post-birth in hospital via direct venepuncture, using a 23-gauge needle into the required blood collection tubes (Sarstedt). To minimise the number of blood collection procedures performed, the venous sample is also utilised for application to the routine newborn screening Guthrie card.
3. Blood samples from healthy children aged 30 days to 18 years are collected at The Royal Children's Hospital, Northern Hospital and Western Health - Sunshine Hospital. The samples are obtained from "healthy" children prior to minor elective day surgery (e.g. circumcision). Other than elective surgery, these children are deemed "healthy" and are not receiving any medications. All children are fasted before their elective surgery.

Only one attempt to collect blood per participant, per sample is made during the HAPPI Kids study. Blood samples are collected into specific tubes required by the analyser manufacturer depending on the analytes being tested (e.g. serum vs citrate). The participant recruitment and sample collection are currently on going.

Table 1: Summary of the inclusion and exclusion criteria at screening

	Inclusion criteria	**Exclusion criteria (any one)**
Age groups		
Premature neonates	• Gestational age of 32 to 36 weeks • Generally healthy	• Presence of systemic abnormalities • Requires interpreter
Term neonates	• From birth up to the maximum 72 hours post-birth • Gestational age ≥37 weeks • Weight ≥ 2500 grams • APGAR Score ≥7 at five minutes	
Paediatric	• Attending hospital for minor elective surgery requiring general anaesthetic (30 days to 18 years) OR volunteer to participate in the study after seeing a flyer related to the study (15 to 18 years)	
Category		
Haematology		• Presence of coagulation disorders • Family history of coagulation disorders • Currently on anticoagulation medication
Immunology		• Presence of immune system disorder or immune deficiency syndrome • Presence of genetic disorder • Presence of rheumatologic disorder • Family history of rheumatologic disorder or immune deficiency syndrome • Infection or a febrile illness within the last 7 days

	• Infection or a febrile illness within the last 4 weeks • Hospital admission for intravenous (IV) antibiotics to clear an infection on more than 2 occasions in life • Has needed 2 or more months of oral antibiotics more than 2 occasions in their life • Failure to thrive • Recipient of blood products in the last 3 months • Diagnosed with food allergy, asthma, eczema or hayfever • Family history of food allergy, asthma, eczema or food allergy
Biochemistry	• History of liver and renal disease • Presence of endocrine diseases • Presence of metabolic disease • Presence of renal disease • Presence of hepatic disease • Failure to thrive

Sample size

A minimum of 20 samples is collected for 19 specific age groups e.g. neonates and every year of life from 30 days to 18 years [1]. For the biochemical and haematological parameters, a total of 380 samples are analysed. For the immunological analysis, the number of age groups is increased from 19 to 22 (i.e. 1 to 2 months, 2 to 6 months, 6 to 9 months, 9 to 12 months, 12 to 18 months, 18 to 24 months, and every year of life from 3 years to 18 years), equating to a total of 440 samples. For pre-term neonates, a total of 100 samples are collected to facilitate analysis of 20 samples for each week of gestation from 32 to 36 weeks.

The sample size of 380 or 440 for estimating 95% reference intervals of an analyte using parametric methods with samples distributed uniformly across 30 days to < 18 years is considered adequate based on the approach proposed by Royston [1]. Using a formula for the approximate standard error of the reference limit at the mean value of age, he suggested 292 as a suitable sample size for constructing a 95% reference intervals based on restricting the standard error of the limits of the resulting reference intervals to be no more than 10 percent of the residual standard deviation from the parametric model [1].

Sample processing and storage

Samples are thawed as per standard clinical laboratory practice for the analyte in question. There are no repeat freeze-thaw cycles for any of the samples. For some analytes the samples are processed immediately and not frozen, within the same time frame and conditions under which clinical samples would be handled. Analytes that can be frozen will be tested in batches. Testing will be done with samples from all ages included in each batch with randomly assigned orders of testing to avoid any batch to batch bias.

For biochemistry analytes, samples are collected into S-Monovette serum gel tubes (Sarstedt), centrifuged at 5000rpm and 6 °C for 5 minutes, separated and stored in 400 µL aliquots at -80 °C within four hours of collection.

For immunology analytes, samples are collected into either S-Monovette serum gel tubes (Sarstedt), or S-Monovette neutral tubes (Sarstedt), depending on the downstream tests. The samples are spun at 5000rpm and 6 °C for 5 minutes, separated and the resultant serum stored in 400µL aliquots at -80 °C within four hours of collection.

For haematology analytes, samples are collected into S-Monovette citrate 3.2% tubes (Sarstedt). The samples are centrifuged at 3800rpm at room temperature for 5 minutes, and plasma is stored in 500µL aliquots at -80 °C within four hours of collection, using a previously established collection protocol [13].

Samples utilised for determination of blood groups (ABO) and the thalassemia screen are collected into 500uL S-Monovette EDTA tubes and processed in the laboratory within 24 hours of collection.

Cell count samples are collected in S-Monovette EDTA tubes (Sarstedt) and processed by the laboratory within 3 hours of collection.

The details of blood tubes used in sample collection is provided in the Table 2.

Table 2: Blood tubes used in sample collection

Analytes category	**Tube Type**	**Manufacturer**	**Subtype**	**Volume**	**Product Code**	**Age group**
Biochemistry	SST	Sarstedt	S-Monovette	7.5 mL	01.1602.001	Neonate and Paediatric
Haematology	EDTA	Sarstedt	S-Monovette	2.7 mL	05.1167.001	Paediatric
	EDTA	Sarstedt	Micro Tube	0.5 mL	41.1395.002	Neonate
	Sodium Citrate	Sarstedt	S-Monovette	3 mL	05.1165.100	Paediatric
	Sodium Citrate	Sarstedt	S-Monovette	1.4 mL	06.1668.100	Neonate
	Lithium Heparin	Sarstedt	S-Monovette	7.5 mL	01.1608.001	Paediatric
	Lithium Heparin	Sarstedt		0.5 mL	20.1345	Neonate
Immunology	SST	Sarstedt	S-Monovette	7.5 mL	01.1602.001	Neonate and Paediatric
	Neutral	Sarstedt	S-Monovette	7.5 mL	01.1728.001	Neonate and Paediatric

Sample testing

Blood samples collected from the study participants are tested according to documented and accredited standard operating procedures used for the testing of patient samples within the participating laboratories. The biochemical, immunological and haematological analytes tested are listed in table 3. Where indicated (*), sample aliquots from the same patient are tested for the same analyte using different automated analysers commonly in clinical use to facilitate direct head to head comparison. In addition, where indicated in Table 3 (+), aliquots of the same sample are tested in different laboratories using the same analyser type and test method. The sample testing is currently ongoing.

Table 3: List of analytes tested for the HAPPI Kids study.

Biochemical	**Immunological**	**Haematological**
Sodium*+ †	Immunoglobulin G*	Factors: XII, XI, IX, X, II, VII, VIII, V
Potassium*+ †	Immunoglobulin A*	Inhibitors: Protein C, Protein S, Antithrombin, Alpha-2-macroglobulin
Chloride*+ †	Immunoglobulin M*	Von Willebrand factor
Bicarbonate*+ †	Rheumatoid factor*	Collegen binding assay
Urea*+ †	Complement component 3*	Ristocetin cofactor assay
Creatinine *+ †	Complement component 4*	Activated Partial Thromboplastin Time †
Total Bilirubin †	Cystatin C*	Prothrombin Time †
Conjugated Bilirubin*+ †	ANTI-STREPTOLYSIN O TEST*	Fibrinogen †
Alkaline Phosphate*+ †	Thyroid peroxidase*	Thrombin clotting time
Aspartate Aminotransferase *+ †	Thyroglobulin*	D-Dimers †
Alanine Aminotransferase *+ †	Iron*	Full Blood Examination and Reticulocytes
Gamma-Glutamyl Transferase *+ †	Ferritin* †	Red Cell Folate
Total Protein*+	Transferrin*	Glucose-6-phosphate dehydrogenase
Albumin*+	Antinuclear antibody	Active B12
Calcium*+†	Soluble FAS Ligand	Total Homocysteine
Magnesium*+ †	Soluble CD25	Total B12

Phosphate*+ ŧ	Immunoglobulin E	Serum Folate
Lactate dehydrogenase*+	Classic Haemolytic Complement Pathway	
Creatine Kinase *+	Alternative Haemolytic Complement Pathway	
Lipase*+	Mannose-binding lectin Complement Pathway	
Amylase*+		
Uric Acid*+		
Triglycerides *+		
Cholesterol*+		
High-density lipoprotein*+		
Thyroid Stimulating Hormone *+		
Free Thyroxine *+		
Free Triiodothyronine *+		
Anti-Müllerian hormone *		
Oestradiol*		
Sex hormone binding globulin *		
Dehydroepiandrosterone – Sulphate*		
Cortisol*		
Growth Hormone*		
Testosterone*		
25-hydroxyvitamin D*		
High Sensitivity Oestradiol*		
Androstenedione*		
17α-Hydroxyprogesterone *		
Insulin-like growth factor 1*		
Insulin-like growth factor-binding protein-3 *		

* Sample aliquots from the same patient are tested for the same analyte using different automated analysers in common clinical use to facilitate direct head to head comparison.

+ Aliquots of the same sample are tested in different laboratories using the same analyser type and test method.

ŧ Samples from pre-term children are tested.

Data management system

Study data are collected and managed using the REDCap (Research Electronic Data Capture) electronic data capture system hosted at the Murdoch Children's Research Institute

[14]. REDCap is a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.

Data analysis

Test results for different analytes will be plotted according to age in order to identify outliers, explore normality of distributions, and to assess the association between analyte level and age. Outliers will to be excluded from analysis following a detailed examination, including visual inspection, appropriate statistical tests e.g. Tukey's test and assessment of biological implausibility e.g. haemolysis index for biochemistry samples. Reference intervals will be constructed using parametric statistical methods in two steps. In the first step, a fractional polynomial regression model of age will be fitted to the mean of normally distributed or log transformed test results of an analyte, using sex as a covariate [15]. In the second step, the 2.5th and 97.5th centiles will be estimated using quantile regression where the power variables of age from the fractional polynomial regression model are used as covariates [16]. Potential interaction with sex will be examined. The 95% confidence interval of the reference intervals will be estimated based on the fitted model for the reference limits. A combination of statistical testing, i.e. goodness of fit and variance component analysis, and clinical expertise will be used to determine the extent to which the reference intervals of an analyte based on different analysers can be compared.

Stata 15 will be used for data analysis [17].

Study timeframe

The participant recruitment and sample collection commenced on February 2015 and is expected to be completed by August 2021. The testing of samples for 30 biochemical analytes has commenced, while testing of samples for the remaining biochemical analytes and immunological and haematological analytes are currently on going. We plan to complete the testing of sample by February 2021.

Patient and public involvement

The study was conceived in response to patient complaints about having blood tests done at community pathology laboratories and their children requiring repeat testing because age-appropriate normal data was not available. The recruitment and consent processes were piloted and parental/ child feedback incorporated into the main study protocol. A summary of the results will be shared by mail with all the participants involved. If any child has a clinically abnormal result the parents are contacted and appropriate plans for review of the child are made.

ETHICS AND DISSEMINATION

The study protocol has been approved (34183 A) by The Royal Children's Hospital, Melbourne, Ethics in Human Research Committee (HREC) and subsequently approved by the HREC committees of all participating hospitals including, The Royal Women's Hospital, Northern Health and Western Health - Sunshine Hospital. The study is primarily supported by The Royal Children's Hospital Foundation and is based at the Haematology Research Laboratory, Murdoch Children's Research Institute, Melbourne, Australia.

Written consent is obtained from parents or guardians of eligible study participants following a verbal and written explanation of the study and participation requirements by Pathology Collection staff at each site after. Less than 3% of the blood volume is collected from a child considering the amount of blood volume a child has, while only one attempt is made to minimise risk associated with collecting intravenous blood.

Participants' identifying information is replaced by a study number (ID) at recruitment to maintain confidentiality. This ID is used in all laboratory specimens, evaluation forms, reports and other records that leaves the site. Electronic data are stored in REDCap which can only be accessed by the authorised members of the research team. Similarly, paper-based information are stored in a locked filing cabinet and can only be accessed by the designated persons. The blood samples are stored in freezers that are only accessible to authorised members of the research team by using their individual swipe card maintaining a record of who has accessed the samples and at what time.

All results that fall outside the existing reference range used by the clinical laboratory are referred immediately to the study coordinator to determine whether there is any clinical significance that requires follow-up for the child. If necessary, families are contacted and the blood tests repeated, or an appropriate referral will be made and the individual will be followed up by the appropriate specialist.

At the conclusion of the study, blood samples collected as part of this study will be destroyed, unless consent has been given for them to be stored in a bio-bank located at the Haematology Research Laboratory for use in future research.

The study results will be summarised for submission to peer-reviewed journals. Results will be shared with participating hospitals and laboratories and presented at local meetings, national and international conferences.

DISCUSSION

Reference intervals are used on a daily basis by clinicians to interpret measurements obtained from patients [18]. The accuracy and reliability of the reference intervals are highly dependent on the reference population utilised to define those reference intervals [11, 19]. As recommended by Harris and Boyd [20], the characteristics of the sample from the reference population should be similar to those of the patients [20]. Paediatric reference intervals established by laboratories using retrospective data mining techniques provide extensive sample numbers. However, the representativeness of the population is often compromised by using the laboratory results of children who were initially referred for a significant clinical investigation [8]. Samples collected prospectively from pre-schools are useful in constructing reference intervals, but are applicable to pre-school aged children only [6]. Addressing this issue, the Canadian Initiative in Paediatric Reference Intervals (CALIPER) and German Health Interview and Examination Survey for Children and Diagnosis (KiGGS) collected data prospectively using a community-based approach ensuring representativeness [10, 21, 22]. Similarly, the HAPPI Kids study is one of the very few studies collecting data prospectively from pre-defined healthy preterm neonates and term neonates from birth until 72 hours post birth, as well as children from 30 days to 18 years of age.

Samples collected for HAPPI Kids utilise paediatric Pathology Collectors with specific skills and documented competency to collect venous samples from neonates. All collected samples meet RCH laboratory standard operating procedures for collection and storage as documented

in the Specimen Collection Handbook (SCH) for the analytes [23]. All participating laboratories perform daily quality control, participated in external quality assurance programs (Royal College of Pathologists of Australasia Quality Assurance Programs (RCPA QAP)) and were accredited by the National Association of Testing Authorities (NATA). The prospective design of the HAPPI Kids study also ensures quality of sample collection, storage and testing in different laboratories.

The CLSI recommended sample size for establishing reference intervals for a discrete age group is 120 [12]. However, there is no recommendation for sample size when reference intervals are constructed for continuous age. Few papers have previously reported continuous reference intervals across childhood and are based on retrospective data mining of samples taken for clinical purposes over many years [4, 24]. No previous studies have collected prospective samples (which were not required for clinical management) from neonates and children specifically screened to be a "well population" for the purpose of developing continuous reference intervals. In that context the exact numbers required remain unknown. We judged that a similar degree of precision in estimation of the limits of agreement to that proposed by Royston (1991), i.e. requiring the ratio of the SE of the estimated limits to be no greater than 10% of the SD of the variation in the population, would be appropriate in our setting. Planning to recruit 380 children assured us of meeting this criterion. [1]. However, it is our intention to do a simulation study to develop better guidance for appropriate numbers of samples required to allow reliable characterisation of age-specific analyte dynamics.

The direct comparison of reference intervals by analysers is not possible using data mining techniques, nor when testing remainder of clinically driven samples (due to sample aliquot

limitations). To overcome this problem CALIPER program which initially established reference intervals using the Abbott ARCHITECT analyser transferred reference intervals in line with Clinical & Laboratory Standards Institute (CLSI) recommendations by testing only 100 samples on Beckman Coulter DxC800, Ortho Vitros 5600, Roche Cobas 600 and Siemens Vista 1500 using an algorithm [11, 12]. The study used $r2 < 0.70$ as an indicator of non-transference between Abbott ARCHITECT and another platform and reported several analytes as non-transferable between analysers[11]. The challenges in transference from the Abbott ARCHITECT to another analyser were discussed by the authors including the lack of an appropriate number of samples per age group since only 100 samples were tested to predict the linear relationship [11]. The HAPPI Kids study utilises aliquots of the same blood sample from the same patient and tests the aliquots on different analysers, enabling the first documented head-to-head comparison between analysers for the age spectrum covering neonates to children 18 years of age. Therefore, we expect the results of the HAPPI Kids study to be very useful in improving interpretation of pathology results for neonates and children.

REFERENCE

1. Royston, P., Constructing time-specific reference ranges. Statistics in Medicine, 1991. 10(5): p. 675-690.
2. Coffin, C.M., *et al*., Pediatric laboratory medicine: current challenges and future opportunities. American Journal Of Clinical Pathology, 2002. 117(5): p. 683-690.
3. Adeli, K., Closing the gaps in pediatric reference intervals: The CALIPER initiative. Clinical Biochemistry, 2011. 44(7): p. 480-482.
4. Loh, T.P., *et al*., Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. Pathology, 2014. 46(4): p. 336-43.

5. Pediatric reference intervals: Critical gap analysis and establishment of a national initiative. Clinical Biochemistry, 2006. 39(6): p. 559-560.
6. Southcott, E.K., *et al*., Establishment of pediatric reference intervals on a large cohort of healthy children. Clinica Chimica Acta, 2010. 411(19–20): p. 1421-1427.
7. Shaw, J.L.V., *et al*., Validity of establishing pediatric reference intervals based on hospital patient data: A comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clinical Biochemistry, 2014. 47(3): p. 166-172.
8. Colantonio, D.A., *et al*., Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem, 2012. 58(5): p. 854-68.
9. Daly, C.H., *et al*., A systematic review of statistical methods used in constructing pediatric reference intervals. Clinical Biochemistry, 2013. 46(13–14): p. 1220-1227.
10. Kohse, K.P., KiGGS — The German survey on children's health as data base for reference intervals and beyond. Clinical Biochemistry, 2014. 47(9): p. 742-743.
11. Estey, M.P., *et al*., CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens Clinical Chemistry Assays: Direct validation using reference samples from the CALIPER cohort. Clinical Biochemistry, 2013. 46(13–14): p. 1197-1219.
12. CLSI, CLSI Releases Guidelines for Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory. 2008, COMTEX News Network, Inc.
13. Monagle, P., *et al*., Developmental haemostasis. Impact for clinical haemostasis laboratories. Thromb Haemost, 2006. 95(2): p. 362-72.

14. Harris, P.A., *et al*., Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics, 2009. 42(2): p. 377-381.

15. Royston, P. and E.M. Wright, A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. Journal of the Royal Statistical Society: Series A (Statistics in Society), 1998. 161(1): p. 79-101.

16. Wei, Y., *et al*., Quantile regression methods for reference growth charts. Statistics in Medicine, 2006. 25(8): p. 1369-1382.

17. StataCrop. Stata Statisical Software: Release 15. 2017. Colleage Station, TX: Stata Crop LP.

18. Koerbin, G., J.R. Tate, and A.C.f.C.R.I. on behalf of the, Harmonising Adult Reference Intervals in Australia and New Zealand – the Continuing Story. The Clinical Biochemist Reviews, 2016. 37(3): p. 121-129.

19. Ritchie, R.F. and G. Palomaki, Selecting clinically relevant populations for reference intervals. Clinical Chemistry And Laboratory Medicine, 2004. 42(7): p. 702-709.

20. Eugene K. H and James C. B,Statistical Bases of Reference Values in Laboratory Medicine,. New York; Marcel Dekker, Inc., 1995.

21. Adeli, K., *et al*., The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. Crit Rev Clin Lab Sci, 2017. 54(6): p. 358-413.

22. Ridefelt, P., Pediatric reference intervals - the Swedish experience. Clin Biochem, 2014. 47(9): p. 740-1.

23. Melbourne, T.R.C.s.H. RCH Specimen collection handbook. Available from: https://ww2.rch.org.au/specimen/.

24. Zierk, J., *et al*., Age- and Sex-Specific Dynamics in 22 Hematologic and Biochemical Analytes from Birth to Adolescence. Clin Chem, 2015. 61(7): p. 964-973.

Author statement

MH wrote the manuscript and developed the statistical analysis plan. VK wrote the study protocol, obtained ethics approval for the study, co-wrote the manuscript and is the co-ordinator of the study. SM provided Biochemistry expertise in the design of the study and revised the manuscript. JB provided pathology collection expertise in the design of the study, training for the pathology collectors and reviewed the manuscript. JC and SD provided support for statistical analysis plan and reviewed the manuscript. PM conceived the study and contributed to its design and was a major contributor in writing the manuscript. VI contributed to the design of the study and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

This study has been primarily funded by The Royal Children's Hospital Foundation (Proposal ID: 2014 – 233), with supplementary funding from Ortho Diagnostics. In addition, the study received in-kind reagents from Roche Diagnostics. Funding bodies do not have any input in the design of the study, sample collection, analysis and interpretation of data and in preparation of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to the parents and children for their participation in this study. The authors thank members of the Pathology Collection Department of The Royal Children's Hospital for the consent of participants and collection of samples. The authors acknowledge the contribution of the members of the anaesthetic and surgical departments of the Royal Children's Hospital, Northern Hospital and Western Health - Sunshine Hospital, Melbourne, for facilitation of sample collection. In addition, the authors thank members of the post-natal departments of The Royal Women's Hospital, Northern Hospital and Western Health, Sunshine Hospital.

2.3 Discussion

The HAPPI Kids study is one of few studies worldwide, and first in Australia, that collected blood samples prospectively from healthy children from birth until 18 years of age with the aim of estimating continuous age-specific RIs. The study demonstrated that it is feasible to collect blood samples from an adequate number of healthy children for testing in multiple analysers in collaboration with hospitals, laboratories and manufacturers. However, the study had its limitations too.

Accurate and reliable paediatric RIs that reflect dynamic changes during the first few days of life, are still lacking. Difficulties in recruiting adequate numbers of healthy term neonates and infants have been acknowledged by all the major population-based paediatric RI studies [32, 65, 66]. The CALIPER paediatric RIs for neonates and infant (0 to < 1 year) are limited as children were recruited from outpatient clinics [32]. Similarly, studies from Denmark and Sweden addressed the limitation of small sample sizes (6 months to 18 years) by merging datasets to establish paediatric RIs [65]. The HAPPI Kids study successfully collected blood samples from term neonates within 72 hours of birth, albeit with a low consent rate (16%). However, the study was unsuccessful at collecting adequate samples from pre-term neonates born at 32- 36 weeks gestation and no sample was collected from children aged between 72 hours and one month of age [66].

Routine pre-analytical and analytical processes are important factors that require control when establishing paediatric RIs [19]. Following safe clinical practice, blood collection for the HAPPI Kids study was limited to less than 3% of the total blood volume for a child [64]. In addition, ethical consideration of younger children limited multiple or repeated blood collection attempts in this study [66]. Post-collection, standard operating procedures defined

the HAPPI Kids Study specimen processing, but nevertheless data were excluded due to pre-analytical interferences in certain subsets of samples (e.g. sodium, total protein (due to evaporations), and total bilirubin (due to light exposure)) [66]. Adeli *et al* also raised concerns regarding the ability of the automated laboratory equipment to handle small volume specimens, leading to frequent manual processing and difficulties in standardising specimen processing [13]. This problem was also observed in the analysis of the HAPPI Kids Study specimens, where automated processing was overridden to prevent the analyser from aborting automated specimen analysis due to inadequate sample volumes [66].

The samples were collected prospectively and given the ethical difficulties; every sample is precious. Hence, the exclusion criteria were designed to try and identify children in whom there would be no purpose collecting any sample as there would be no analytes that could be considered normal. In many circumstances however, children with some diseases may be likely to be normal for analytes known not to be affected by the disease process. For example, children with thalassemia minor are thought to have normal biochemical parameters.

The magnitude of intra-individual biological variation and total allowable error for different analytes estimated based on adult samples may not be similar to the paediatric population, especially during the first few years of life [13]. Addressing the lack of data on biological variation specific to paediatric populations, the CALIPER study first published paediatric within-day biological variation and quality specification for 38 biomarkers using nested analysis of variation techniques [25]. Considering the challenges in recruiting healthy children using a direct sampling approach, Loh *et al* proposed an indirect method for estimating within-individual biological variation and estimated between-individual biological variation for 22 biomarkers, which was comparable to those reported in the CALIPER study

[28, 67]. In comparison, the HAPPI Kids study was unable to repeat the testing on the sample analysers to estimate technical variation since the aliquots from the blood sample were used for testing in multiple analysers [64]. However, the HAPPI Kids study is one of the few studies that tested the aliquots from the blood samples on five analysers [29, 30, 64].

Chapter 3: Reference values for 30 common biochemistry analytes across five different analysers in neonates and children 30 days to 18 year olds

3.1 Introduction

As presented in the previous chapter, the HAPPI Kids study prospectively collected blood samples from equal numbers of neonates and children at every year of life from age 30 days to 18 years [64]. While the HAPPI Kids study plans to collect specimen for more than 80 analytes and test aliquots from the same sample on different analysers [64], the specimen collection and testing on five commercially available analysers were completed for 30 biochemistry analytes by the end of 2018. This created a unique opportunity to compare the continuous reference values across five analysers and estimate continuous age-specific RIs from neonates to < 18 years.

This chapter describes the findings of the direct head-to-head comparison of clinically important biochemistry assays across these five major analysers used throughout the world. This enabled consideration of the question as to whether results on a patient in one laboratory can be compared to results on the same patient elsewhere, which is important in the current model of the hub and spoke medical systems. The data also addressed the ability of RIs to be harmonised across analysers, a topic which has been of considerable interest recently. This chapter also reports continuous RIs as a function of age and sex. The continuous RIs has potential to substantially improve health care through the ability to more accurately assess children over time. The study finding was published in Clinical Chemistry in September 2019 and the author approved manuscript is presented as the body of this chapter [66].

3.2 Publication

Reference values for 30 common biochemistry analytes across five different analysers in neonates and children 30 days to 18 year olds

Running head: Paediatric reference intervals across 5 analysers

Monsurul Hoq

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Susan Matthews

The Royal Children's Hospital, Parkville, Australia; International Centre for Point of Care Testing, Flinders University, Bedford Park, Australia.

Vicky Karlaftis

Murdoch Children's Research Institute, Parkville, Australia.

Janet Burgess

The Royal Children's Hospital, Parkville, Australia.

Jessica Cowley

Murdoch Children's Research Institute, Parkville, Australia; The Royal Children's Hospital, Parkville, Australia.

Susan Donath

Murdoch Children's Research Institute; Parkville, Australia; The University of Melbourne; Parkville, Australia.

John Carlin

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Tina Yen

The Royal Children's Hospital, Parkville, Australia.

Vera Ignjatovic*

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Paul Monagle*

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia

On behalf of the HAPPI Kids study team

*Contributed equally

Corresponding Author:

Associate Prof Vera Ignjatovic, PhD, B.Sc. (Hons)

Co-Group Leader, Haematology Research Laboratory, Murdoch Children's Research Institute

Principal Fellow, Department of Paediatrics, The University of Melbourne

Address

Haematology Research Laboratory

Murdoch Children's Research Institute

50 Flemington Road, Parkville, Victoria, 3052, Australia

Email: vera.ignjatovic@mcri.edu.au, T: +61 3 9936 6520, M: +61 439 363 891

Key words: pediatric, reference standards, biochemistry, biomarkers, age factors

List of abbreviations

Reference interval (RI)

Laboratory information system (LIS)

Harmonising Age Pathology Parameters in Kids (HAPPI Kids)

Ethics in Human Research Committee (HREC)

Ear, Nose & Throat (ENT)

Thyroid stimulating hormone (TSH)

Free thyroxine (FT4)

Free triiodothyronine (FT3)

National Association of Testing Authorities (NATA)

Analytical performance specifications (APS)

Alanine amino transferase (ALT)

Aspartate amino transferase (AST)

Gamma-glutamyl transferase (GGT)

Inter-quartile range (IQR)

Coefficient of variation (CV)

Confidence intervals (CI)

Alkaline phosphatase (ALP)

Bromocresol green (BCG

Lactate dehydrogenase (LDH)

Creatinine kinase (CK)

Aspartate amino transferase (AST)

High-density lipoprotein (HDL)

Canadian Laboratory Initiative on Paediatric Reference Intervals (CALIPER)

Pyruvate to Lactate (P-L)

Lactate to Pyruvate (L-P)

International Federation of Clinical Chemistry and Laboratory Medicine (IFCC)

Pyridoxal-5-Phosphate (P5P)

Bromocresol purple (BCP)

Australasian Association of Clinical Biochemistry (AACB)

Randomized controlled trials (RCTs)

Food and Drug Authority (FDA)

In-vitro diagnostics (IVDs)

ABSTRACT

Background

Age-specific reference intervals (RIs) have been developed for biochemistry analytes in children. However, the ability to interpret results from multiple laboratories for one individual is limited. This study reports a head-to-head comparison of reference values and age-specific RIs for 30 biochemistry analytes for children across five analyser types.

Methods

Blood was collected from healthy newborn and children aged 30 days to <18 years. Serum aliquots from the same individual were analysed on five analyser types. Differences in the mean reference values of the analytes by the analyser types were investigated using mixed-effect regression analysis and by comparing maximum variation between analysers with analyte-specific allowable total error reported in the Westgard QC database. Quantile regression was used to estimate age-specific RIs using power variables in age selected by fractional polynomial regression for the mean, with modification by sex where appropriate.

Results

The variations of age-specific mean reference values between analyser types was within allowable total error (Westgard QC) for most analytes and common age-specific reference limits were reported as functions of age and/or sex. Analyser-specific reference limits for all analytes on five analyser types are also reported as functions of age and/or sex.

Conclusion

This study provides quantitative and qualitative measures of the extent to which results for individual children can or cannot be compared across analyser types and the feasibility of RI

harmonization. The reported equations enable incorporation of age-specific RIs into laboratory information systems for improving evidence-based clinical decisions in children.

INTRODUCTION

Clinical laboratory test results are critical to evidence-based medicine, with nearly 80% of physicians' decisions based on information provided by laboratory reports *(1)*. A laboratory test result should be reported with a reference interval (RI) appropriate to the patient's demographic that facilitates accurate clinical interpretation. Statistically, RIs are defined by a range between the 2.5^{th} and 97.5^{th} percentile, representing 95% of the "normal" population. However, the establishment of accurate RIs is a complex process.

Reference samples should be selected at random *(2)*. However, access to "healthy" individuals is challenging, especially for neonates and children. Even with access to "normal" paediatric patients, low patient consent rates and small blood volume availability create difficulties *(3)*. Consequently, most paediatric studies have extracted data from laboratory information systems (LISs) to establish RIs *(4, 5)*. Few studies have collected samples prospectively *(6)*.

Pre-analytical and analytical factors such as specimen collection, transport, processing and laboratory analysis must be rigorously controlled and ideally mimic the processes used for patient specimens, to ensure the data used formulating the RIs are accurate. There is an increasing range of automated analyser types, analytical methodologies, and reagent formulations. Several studies reported inter-analyser variation for common biochemistry analytes *(7-9)*. However, there are very few analyser-to-analyser comparisons of reference values using samples from the same patient *(8, 9)*.

Age, sex, and ethnicity can influence RIs *(4-6)*. Studies demonstrated the continual physiological changes of biomarkers in children by establishing RIs for different categories of sex, ethnicity (and other categorical variables) and for discrete age groups by partitioning age into several groups *(2, 3, 6)*. This approach can often lead to misclassification for children, specifically close to the age "cut-off" points *(4, 10)*. These artificial, arbitrary, age groups may not be inclusive of the overall variation of a specific analyte with age *(4, 11)*. An alternative to age-grouped RIs, is paediatric reference curves for biochemical analytes *(4, 5, 12)*. Furthermore, as LISs evolve, the incorporation of centile charts for clinical decisions should ensue.

An understanding of the direct comparison of biochemical results, as measured by different analyser types, will provide evidence based data that will assist in development of common paediatric age-specific RIs. Addressing this clinically important issues, this study recruited equal numbers of participants from neonates and children aged 30 days to < 18 years, aliquoted patient samples for direct comparison across five analyser types and reported age-specific RIs as functions of age and/or sex.

MATERIALS AND METHODS

Study design

The Harmonising Age Pathology Parameters in Kids (HAPPI Kids) study is a prospective, cross-sectional study of paediatric blood samples for commonly requested biomarkers. The study protocol was approved by The Royal Children's Hospital, Melbourne, Australia, Ethics in Human Research Committee (HREC) and subsequently approved by the HRECs of all participating hospitals (HREC 34183 A). An outline of the study protocol is provided and full

details published elsewhere *(13)*. The study contributors are listed in Supplementary Table 1 of the Appendix B.

Study setting and participant recruitment

Participants were recruited after written consent from parents/ legal guardians and conformance with study inclusion criteria. Participation was excluded on the presence of reported acute or systemic illness, endocrine diseases or metabolic disease, history of liver or renal disease and/or failure to thrive or a requirement for an interpreter.

Neonatal blood samples were collected from healthy term neonates following routine intramuscular administration of one milligram of vitamin K in three delivery suites in Melbourne, Australia *(13,14)*. Venous blood samples from healthy children aged 30 days to < 18 years were obtained from Hospital Outpatient Services who were undergoing minor elective day surgery (e.g. orthopaedic, plastic surgery, Ear, Nose & Throat (ENT), minor genitourinary and/or dermatological). Subjects met study inclusion criteria and were not receiving regular medication. All subjects were fasted prior to elective surgery. Only one attempt to collect blood per patient was allowed *(13).* All specimens were collected during the period February 2015 – December 2017.

Sample size

A maximum of 20 samples, with equal numbers of male and female participants, were collected for neonates (birth up to a maximum of 72 hours post-birth) and for every year of life from 30 days to < 18 years, resulting in a total of 380 samples per analyte. However, the blood volumes collected were not adequate to perform tests for all analytes on five analysers. Hence, a total 616 children were recruited.

Sample processing and testing

Standardized laboratory procedures used are described elsewhere *(13)*. Briefly, blood was centrifuged at 5000 rpm (4472 *g*-force) at 6°C for 5 minutes, separated into 400 µL aliquots and stored for a maximum period of 24 months at -80 °C. Pre-analytical processing was completed within four hours of collection.

Samples were transported on ice, in batches to each participating laboratory and processed amongst routine clinical analysis during December 2015 – December 2017. Each laboratory followed identical pre-analytical specimen thawing, specimen mixing and internal quality control protocols. Biochemical analysis was performed according to accredited standard operating procedures used for patient sample testing within the participating laboratories. Specimens within each batch represented all study age groups, to minimize systematic bias.

Samples from the same patient were tested for 24 biochemistry biomarkers on five, automated analysers: Ortho Clinical Diagnostics Vitros 5600, Abbott Architect c8000, Roche Cobas c701, Siemens ADVIA 1800, and Beckman Coulter AU5800, in five collaborating laboratories. Iron, ferritin, and transferrin were measured using Roche Cobas c501/c502, Abbott Architect c800, Roche Cobas c701, Siemens ADVIA 1800, and Beckman Coulter AU5800. Thyroid stimulating hormone (TSH), free thyroxine (FT4) and free triiodothyronine (FT3) were tested on OCD Vitros 5600 (Immunoassay), Abbott Architect (i1000SR, i2000SR, i4000SR) Roche Cobas e602, Siemens ADVIA Centaur/XP/XPT, and Beckman Counter Unicel DXL 600/800. This facilitated direct head-to-head comparisons of the same individual across the five analysers. Analytical methods have been detailed in Supplementary Table 2 of the Appendix B.

All participating laboratories were accredited to ISO: 15189 by the National Association of Testing Authorities (NATA) at the time of testing and met required analytical performance specifications (APS).

Statistical methods

Histograms, scatter plots and boxplots were used to visually check the test results for outliers, distributions, and assessing the association with age and/or sex. Subsequently, linear or non-linear regression model using fractional polynomials in age, with or without sex and/or interaction between age and sex as covariates, as supported by the data and previous literature, were fitted to the mean values of each analyte for children aged 30 days to < 18 years *(4-6, 15-17)*. Data were log transformed for alanine amino transferase (ALT), aspartate amino transferase (AST), gamma-glutamyl transferase (GGT), and ferritin. Afterwards, outliers were identified by calculating the residuals i.e. by subtracting the predicted values from the observed values. If the residual was greater than three times the inter-quartile range (IQR) above or below the 3rd and 1st quartile respectively, the result was considered an outlier *(18)*.

Following outlier exclusion, mixed-effects regression analyses were performed utilizing age, analyser and the interaction between age and analyser as fixed effects and individual as a random effect to investigate whether the mean reference values varied by analyser types. The mean reference value for each analyser type was estimated using the model at each integer age until 18 years and the coefficient of variation (CV) between these estimated means (i.e. between the five analyser types or between analyser types for different methods of creatinine, ALT, AST, and albumin) were calculated at each integer age. The variation by analyser types was considered clinically insignificant, if the maximum CV of the predicted mean values

across the 18 integer age values was less than "the allowable total error" according to the quality requirements specified in the Westgard QC database *(19)*. For analytes, for which the mean reference values did not vary between analyser types in a clinically significant manner, the data for each child across the different analysers were combined and a common mean and RIs were established.

Age-specific RIs were determined for children aged 30 days to < 18 years by using quantile regression estimating the 2.5^{th} and 97.5^{th} percentiles as smooth functions of age, using the same power variables of age that were identified from the fractional polynomial regression model for the mean *(20)*. Sex and interactions between power variables of age and sex as covariates were included in the final models based on the initial model for the mean. The age-specific (and, where necessary, sex-specific) reference limits with 95% confidence intervals (CI) were calculated using the fitted model for children and plotted against age.

Initial visual inspection suggested different ranges for neonates compared to children in early infancy after the neonatal period. Hence neonatal data were included in the regression model only when the observed mean reference value at birth was consistent with the 95% CI for the mean at age zero based on the model estimated using the non-neonatal data. If the observed mean was different, the 20 neonatal data values were summarized separately using median, maximum and minimum values.

Statistical analyses was performed using Stata 15 *(21)*. Further details of the analytic method using alkaline phosphatase (ALP) as an example are provided in Supplementary Document 1 of the Appendix B**.**

RESULTS

The majority (94%) of the study participants (n=616) were born in Australia. The birth country of the parents of participants included Australia (55%), and countries from South-East Asia (6%), South Asia (5%), Europe (5%), Middle East (4%), and Africa (3%). The ethnic distribution of the study cohort was in line with the Australian population *(22).*

Datum from one specimen was excluded from all analytes due to the biological implausibility of creatinine. Data from 35 samples were excluded from calcium, total protein, and ALT; and 135 from sodium as pre-analytical interference was recorded by the testing laboratory. The number of data excluded based on the statistical criteria have been documented in Table 3 of the Appendix B. No sample data were excluded for haemolysis, icterus or lipemia.

The findings of the mixed-effect regression models suggest there were variations in the age-specific mean reference values between analysers for all analytes. However, the maximum CV for mean reference values across age was not clinically significant when compared with "allowable total error" for most analytes (Table 1). In contrast, the variation in the mean reference values between analyser types was both statistically and clinically significant for bicarbonate, total bilirubin, creatinine(Jaffe), albumin (bromocresol green (BCG)), ferritin, transferrin, FT3 and FT4 (Table 1). The findings of the mixed-effect regression analysis and CVs of age-specific mean reference values by analyser types (and for different methods) at each integer ages until 18 years for creatinine and GGT are presented in Supplementary Document 2 of the Appendix B as examples of a) both statistically and clinically significant variations and b) statistically but not clinically significant variation respectively.

Table 1: Comparison of variation of age-specific mean reference values between analyser types and "allowable total error" (Westgard QC).

Analyte	Variation between analyser type	Allowable total error[a]
Sodium	0.7%	0.7%
Potassium	1.7%	5.6%
Chloride	1.2%	1.5%
Bicarbonate	**18.6%**	**10.0%**[b]
Urea	3.9%	15.5%
Creatinine	**27.8%**	8.9%
Creatinine (enzymatic)[c]	8.7%	
Creatinine (jaffe)[d]	**36.0%**	
Calcium	1.3%	2.5%
Magnesium	4.5%	4.8%
Phosphate	5.1%	10.1%
Triglyceride	4.9%	26.0%
Total Cholesterol	3.4%	9.0%
High-density lipoprotein Cholesterol	3.4%	11.6%
Alkaline Phosphatase	11.8%	12.0%
Aspartate Amino Transferase	2.2%	16.7%
Aspartate Amino Transferase (with P5P)[e]	1.4%	
Aspartate Amino Transferase (without P5P)[f]	3.7%	
Alanine Amino Transferase	10%	27.5%
Alanine Amino Transferase (with P5P)[e]	12.4%	
Alanine Amino Transferase (without P5P)[f]	5.7%	
Gamma-Glutamyl Transferase	2.8%	22.1%
Total Protein	2.0%	3.7%
Total Bilirubin	**35.1%**	**26.9%**
Albumin	**5.4%**	**4.1%**
Albumin (BCP)[g]	2.3%	
Albumin (BCG)[h]	**4.9%**	
Lactate dehydrogenase (L-P)	4.2%	11.4%
Creatine Kinase	7.8%	30.3%
Lipase (Wet chemistry methods)	28.5%	37.9%
Amylase	12.4%	14.6%
Uric Acid	5.6%	12.0%
Iron	3.7%	30.7%
Ferritin	**95.0%**	**16.9%**

Transferrin	**7.0%**	**3.8%**
Thyroid Stimulating Hormone	11.0%	23.7%
Free Thyroxine	**15.3%**	**8.0%**
Free Triiodothyronine	**13.0%**	**11.3%**

The analytes for which the age-specific mean reference values vary by analysers types are highlighted by bold font.

a. According to the Westgard QC Database *(19)*
b. According to the Royal College of Pathologist of Australasia Quality Assurance Program
c. Creatinine (enzymatic) on Ortho and Roche
d. Creatinine (jaffe) on Abbott, Roche, Siemens and Beckman
e. AST(with P5P) or ALT (with P5P) on Ortho, Abbott, Roche and Beckman
f. AST (without P5P) or ALT (without P5P) on Roche and Siemens.
g. Albumin (BCP) on Abbott, Roche, Siemens and Beckman
h. Albumin (BCG) on Ortho and Roche

The common reference limits for the 2.5^{th} and 97.5^{th} percentile have been reported as functions of age and/or sex in Table 2 for 24 analytes, for which the age-specific mean reference values do not vary between analyser and presented in Figures 1 & 2. All analyser-specific reference limits as function of age and/or sex are provided in Table 4 and presented in Supplementary Figures 1 through 30 of the Appendix B. Except for potassium, the observed mean values for neonates fall outside the 95% CI of the predicted mean based on the model using non neonatal data. Hence summary statistics of the data from neonates for all analytes are tabulated in Supplementary Table 5 of the Appendix B.

Table 2: Parametric equations for common lower and upper limits of 24 biochemistry analytes for children aged 30 days to 18 years.

Analytes	Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)
Sodium	137.04 + 0.07 Age[a]	142.20+0.09 Age
Potassium	4.03 – 0.02 Age	5.05 – 0.02 Age
Chloride	103.68 – 0.13 Age	109.58 – 0.02 Age
Urea	2.34 + 0.65 Male[b] – 0.02 Age*Male	6.24 + 0.02 Age + 0.08 Male – $5.29*10^{-3}$ Age*Male
Creatinine (enzymatic)	16.70 + 1.72 Age – 1.56 Male + 0.12 Age*Male	26.30 + 3.00 Age – 2.22 Male + 0.98 Age*Male
Calcium	2.22 + 0.03 Age^{-1}	2.55 + 0.03 Age^{-1}
Magnesium	0.80 – 0.03 ln(Age) + 0.01 Male	1.01 – 0.04 ln(Age) + 0.02 Male
Phosphate	1.37 + 0.15 Age^{-1} – $8.36*10^{-5}$ Age^3	2.00 + 0.14 Age^{-1} – $6.82*10^{-5}$ Age^3
Triglyceride	0.36 + 0.01 Age	1.50 + 0.06 Age
Cholesterol	2.57 + $1.25*10^{-2}$ Age^2 – $7.51*10^{-4}$ Age^3	5.13 + $1.64*10^{-2}$ Age^2 – $8.15*10^{-4}$ Age^3
HDL Cholesterol	0.37 + 0.19 Age – 0.06 Age*ln(Age)	1.49 + 0.16 Age – 0.05 Age*ln(Age)
Alkaline Phosphatase	153.100 – 0.146 Age^3 + 0.043 Age^3ln(Age) + 0.129 Male*Age^3 – 0.043 Male*Age^3ln(Age)	387.700 + 0.216 Age^3 – 0.089 Age^3ln(Age) + 0.338 Male*Age^3 – 0.115 Male*Age^3ln(Age)
Aspartate Amino Transferase (with P5P)	30.76 – 0.71 ln(Age) – 1.80 $(ln(Age))^2$ + 0.40 Male	57.07 – 10.17 ln(Age)+ 1.28 $(ln(Age))^2$ + 2.47 Male
Aspartate Amino Transferase (without P5P)	26.85 – 1.01 ln(Age) – 1.22 $(ln(Age))^2$ + 1.57 Male	54.07 – 8.27 ln(Age)+ 0.40 $(ln(Age))^2$ + 3.16 Male
Alanine Amino Transferase (with P5P)	10.73 + 1.48 Age^{-1} + 0.18 Male	37.18 + 0.73 Age^{-1} – 0.75 Male
Alanine Amino Transferase (without P5P)	8.86 + 2.25 Age^{-1} + 0.95 Male	32.48 + 0.25 Age^{-1} + 1.48 Male
Gamma Glutamyl Transferase	5.67 + 0.09 Age	16.05 + 0.82 Age
Albumin (BCP)	34.23 + 0.07 Age	41.47 + 0.31 Age
Total Protein	53.70 + 3.90 ln(Age)	66.05 + 5.40 ln(Age)
Lactate dehydrogenase (L to P)	276.53 – 33.01 $Age^{0.5}$	463.63 – 45.38 $Age^{0.5}$
Amylase	12.89 + 9.22 ln(Age) – $2.57*10^{-2}$ Age^2	76.49 + 33.85 ln(Age) – 0.24 Age^2

Lipase (Wet Chemistry methods)	9.78 +0.12 Age – 1.34 Male	31.41 + 0.36 Age – 0.96 Male
Uric Acid	148.16 + 0.14 Age^2 – 32.67 Male + 0.25 Age^2*Male	319.53 + 0.21 Age^2 + 11.39 Male + 0.50 Age^2*Male
Iron	2.48 + 0.18 Age	22.53 + 0.35 Age
Creatinine Kinase	68.99 – 2.91 Age - 1.83 Male + 1.83 Age*Male	248.60 – 3.30 Age + 28.78 Male + 2.21 Age*Male
Thyroid Stimulating Hormone	0.89	5.20

a. In years

b. Replace male with "1", if subject is male. Otherwise replace male with "0" if subject is female.

Most of the analytes demonstrated substantial relationship with age, especially during the first year of life and the pubertal periods of both male and female subjects. As expected, reference values increased with age for creatinine (due to muscle mass), total protein, albumin (due to blood volume) and iron (due to dietary intake/stores) (Figure 1 and 2) while decreased with age for potassium, lactate dehydrogenase (LDH), and creatinine kinase (CK) (Figure 1 and 2). The RIs decreased substantially during earlier years of life for phosphate and AST (Figure 1 and 2). For triglyceride, GGT, and total bilirubin the upper reference limits increased with age, but there were minimal changes in lower reference limits (Figure 1 and 2). The age-related changes in the RIs may not be of clinical significance for sodium, potassium, chloride, bicarbonate, urea, calcium, total cholesterol, high-density lipoprotein (HDL) cholesterol, lipase, ferritin, transferrin and FT3 (Figure 1 and 2). The age-specific RIs differ by sex for magnesium, FT3 and FT4 consistently over age (Figure 2). Sex-specific differences in the age-specific RIs were emerging during puberty in both males and female study participants for creatinine, ALP, CK, uric acid, ferritin and transferrin (Figure 2). However, the difference in RIs by sex may not be of clinical importance for urea, AST, ALT and lipase (Figure 2).

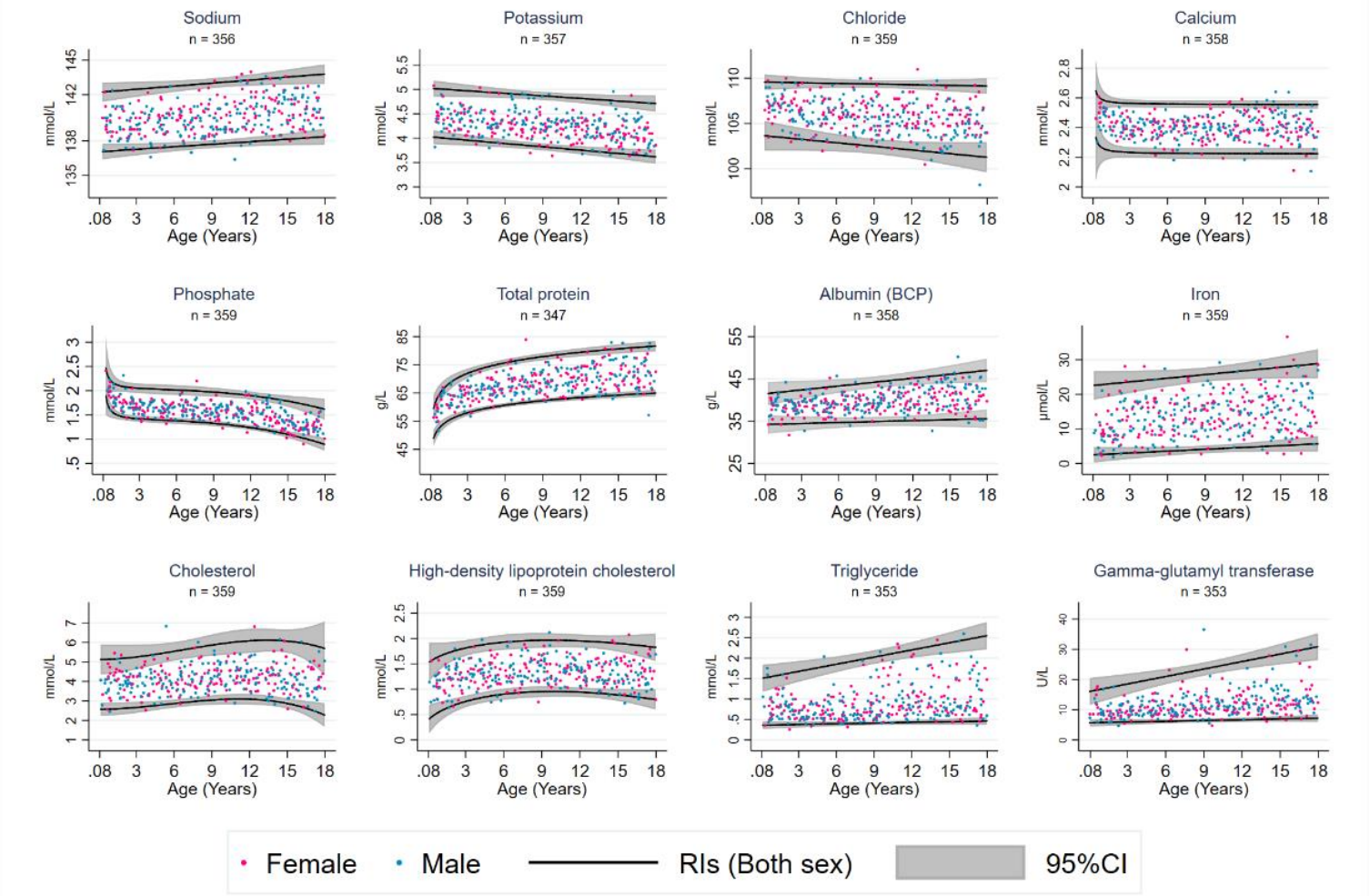


Figure 1: Age-specific paediatric reference intervals for common biochemistry analytes.

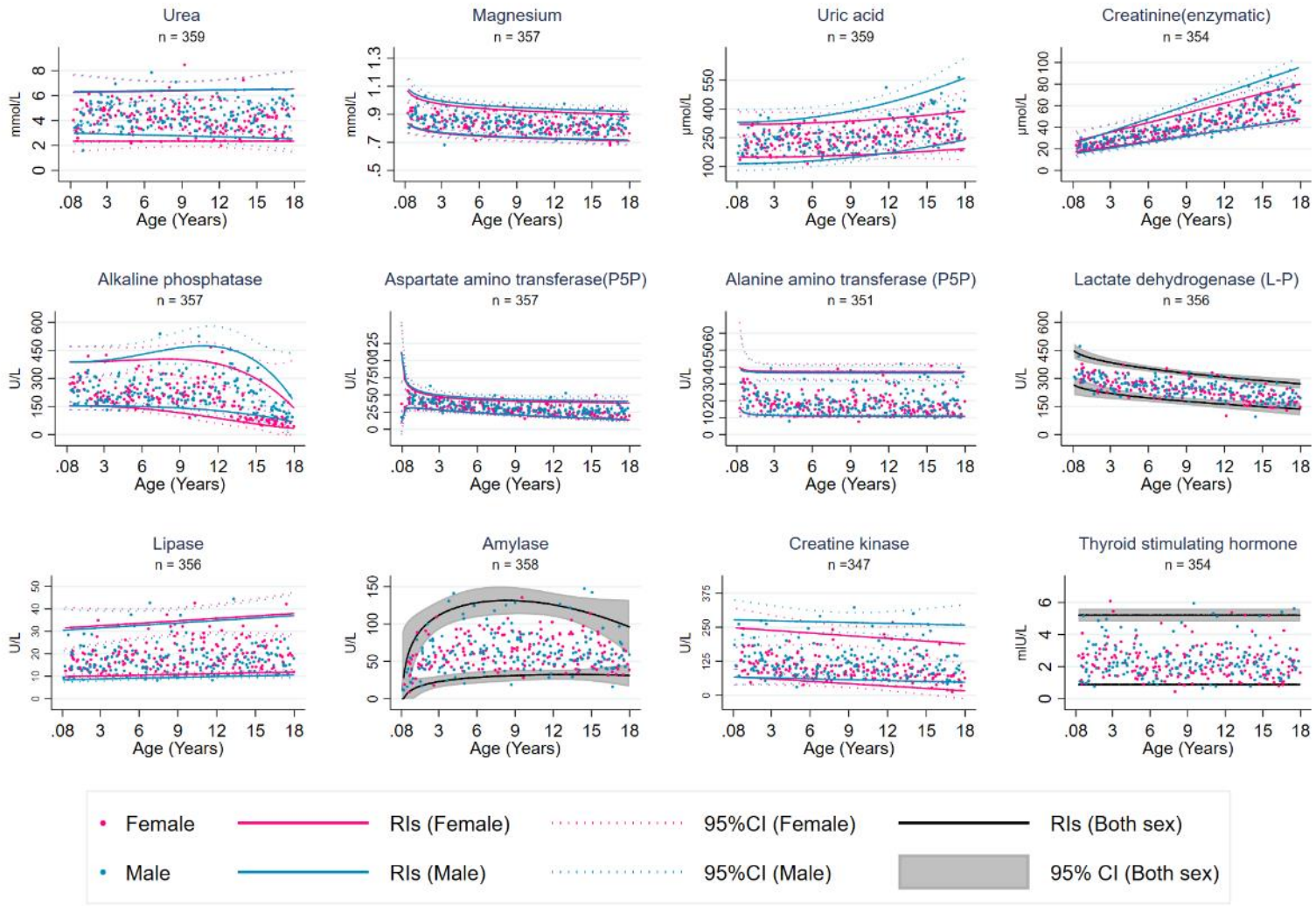


Figure 2: Age-specific paediatric reference intervals for common biochemistry analytes (continued).

The model-based age-specific RIs for creatinine and ALP were similar to the RIs based on discrete age-groups reported by the Canadian Laboratory Initiative on Pediatric RIs (CALIPER) study, in terms of changes with age and during puberty (Figure 3) *(6)*. In contrast, our results were different from age-specific RIs for ALP based on retrospective data from hospital patients in Germany (Figure 4) *(5)*.

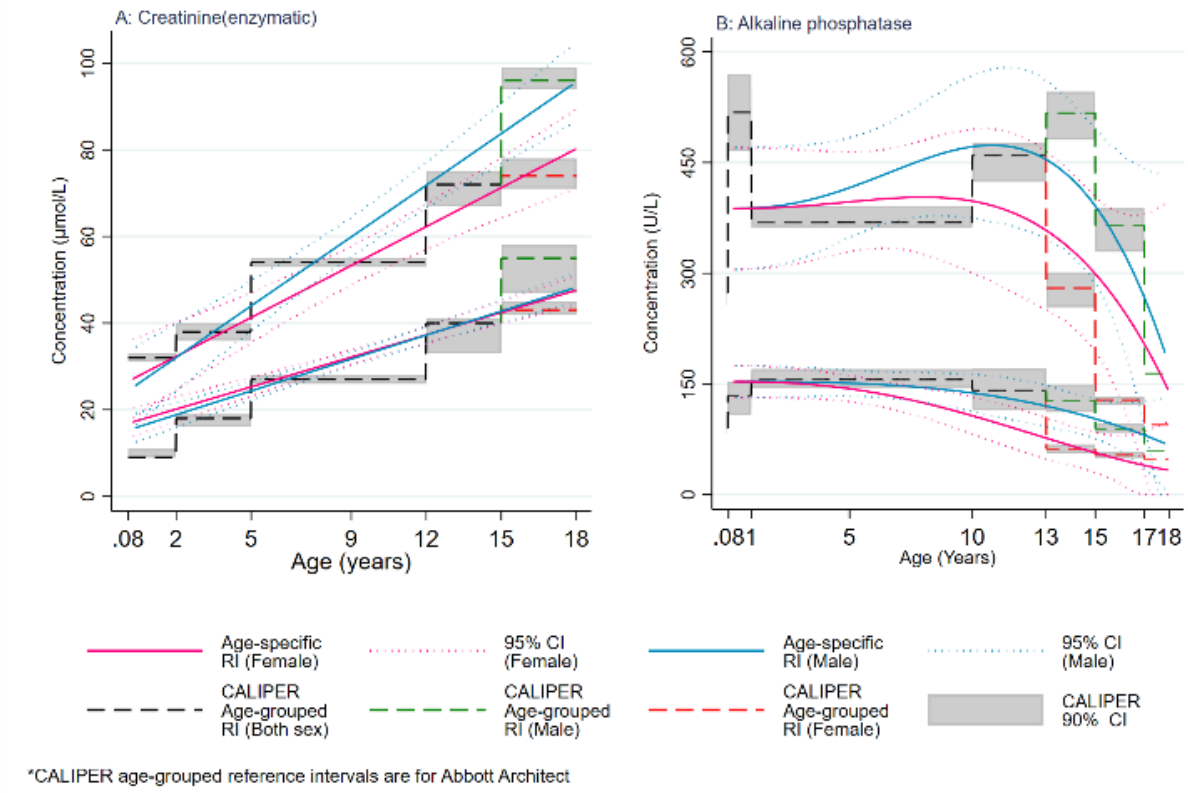


Figure 3: Age-specific (common) compared with CALIPER age-grouped reference intervals.

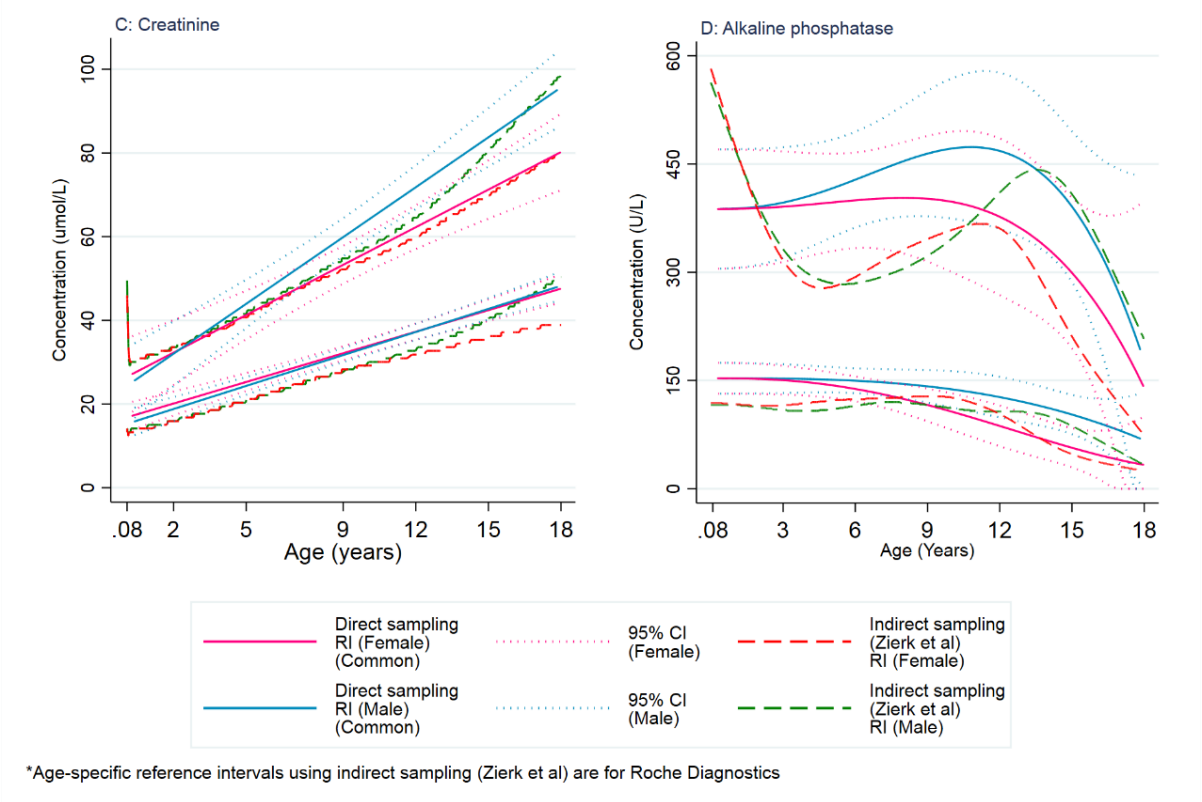


Figure 4: Age-specific reference intervals using direct (common) compared with indirect sampling (Zierk *et al*).

DISCUSSION

This study reports the first head-to-head comparison of paediatric reference values across five analyser types for common biochemistry analytes, using serum aliquots from the same child. The unique study design, the age and even distribution of study participants from 30 days to < 18 years, facilitated estimation of age-specific RIs using statistical models that can be adopted into current LISs. We also present neonatal data for the same analytes and analysers. This study demonstrates the feasibility of obtaining RIs using data from representative healthy children, highlights clinically significant differences in reference values between analyser types for some analytes and reports reference limits as functions of age and/or sex.

Whether patient samples tested at different laboratories, using different analysers, can be directly compared is of global importance, particularly as “hub-and-spoke” models of paediatric care for complex children, (e.g. shared care with tertiary centres (hub) and local health care providers (spoke)) request laboratory testing by different providers for clinical management of the same patient. No previous study has directly compared biochemical reference values on five analysers across the paediatric age span using samples from the same subject. CALIPER collected specimens from healthy children for establishing age-grouped RIs, and its numerous publications have provided important data for clinical care of children *(6)*. However, the majority of the RIs established by the CALIPER study utilized Abbott assays and “transferred” the Abbott intervals to other analysers except for immunoassays *(6)*.

We chose total error to compare the mean reference values between analysers as this has been used in another study *(8).* The total error is also accepted as a reasonable method of comparing results in paediatric population (23).

Of the biochemical analytes tested within this study, the reference values varied across analyser types for bicarbonate, creatinine (Jaffe), total bilirubin, albumin (BCG), ferritin, transferrin, FT3 and FT4. These differences may be due to pre-analytical issues (i.e. bicarbonate) or analysis method or calibration that are non-standardized or not traceable to standardized source. For creatinine, the average intra-analyser coefficient of variation for mean reference values was 11.9%, was higher than between laboratories coefficients of variation (4.37% to 8.74%) reported by Delanghe *et al. (8)*. The differences in reference values between Jaffe and Enzymatic were also similar to the results of Delanghe *et al*., who reported a significant difference between the methods in the earlier years of life and that values of dry chemistry were higher than Jaffe for higher values of creatinine *(8)*. The

differences in reference values for bicarbonate may be explained by the use of frozen instead of fresh samples due to logistical issues and the reported analyser-specific RIs may not be clinically useful *(24)*. In the case of ferritin, since the antibodies to ferritin used in immunoassays may not react well to all forms, RIs should be established for each analyser type. Similarly, free thyroid hormone immunoassays are not standardized, which accounts for method specific differences. For LDH, the analytical method for the OCD Vitros 5600 was pyruvate to lactate (P-L), whereas other analyers utilized lactate to pyruvate (L-P), hence method specific RIs for LDH were expected. Method difference also existed for lipase, the OCD Vitros 5600 utilized dry slide chemistry with 1-oleoyl-2, 3-diacetylglycerol as a substrate, whereas wet chemistry methods using either 1,2-diglyceride or glutaric acid-(6-methyresorufin) ester substrates were utilized on the other analysers.

Differences in reference values were statistically negligible for GGT. All analysers utilized International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) standardized GGT methods. Similarly, Steinmetz el al. found no significant difference in reference values for seven laboratories when compared to the IFCC recommended GGT method *(9)*. The use of pyridoxal-5-phosphate (P5P) as co-factors for the AST and ALT methods are the recommended IFCC methods while bromocresol purple (BCP) as one of the dye-binding methods for measurement of albumin is preferred (25). However, no significant differences in age-specific mean reference values between analysers or in reference values between two different methods on the same analyser (Roche Cobas) for AST, ALT and albumin were reported in this study (See Supplementary Document 3 of the Appendix B). Our findings of clinically non-significant differences in mean reference values between analysers for most electrolytes were similar to the Australasian Association of Clinical Biochemistry (AACB)

(26). Even though we have reported common RIs for those analytes, the clinical significance of the common RIs against analyser-specific RIs in sick children need to be tested.

To avoid clinical misdiagnosis of laboratory results in children of ages close to age cut-offs using age grouped RIs, other studies have endorsed the implementation of continuous paediatric RIs across ages *(4, 5, 12)*. Higgins and Adeli have recently discussed the advantages of using continuous age-related paediatric RIs. The implementation of continuous RIs into LISs was seen as challenging and more innovative, interpretative tools were suggested *(11)*. In an effort to address such challenges, we have applied modern best practice statistical methods such as fractional polynomial and quantile regression, and presented paediatric RIs as functions of age and/or sex *(27, 28).* A similar statistical method has been used in other fields of medicine *(29).*

The physiological trends in age-specific RIs observed in this study for most biochemistry analytes, were similar to centile charts reported by Loh *et al.,* Zierk *et al*., and Bussler *et al*. and to the age-grouped RIs reported by CALIPER Study *(4-6, 12)*. The differences in RIs for ALP from hospital patients in Germany could be due to inclusion of neonates, distribution of the sample over age, the sampling techniques and statistical technique applied *(5).* The changes in measurement with age for sodium, potassium, chloride, bicarbonate, urea, calcium and magnesium reported in this study were not clinically significant and common paediatric RIs across age, similar to those reported by the AACB, could be adopted *(26)*.

Considering the recent trend to presenting RIs for continuous age, the provision of RIs as equations of age and/or sex for common biochemical analytes, will facilitate their integration into current LISs. However, there is considerable variation in LIS systems among vendors

(e.g. age of the software, programming complexity and capacity to present RI graphically). The modern LISs integrated with the electronic medical recording system can incorporate these equations and graphical representation, therefore potentially facilitating widespread access and improving patient care in children.

The different ranges of neonatal values could be explained by wash out of maternal parameters and did not fit the equations for older children. The ranges of neonatal data are similar to previous studies and useful in comparing between analysers *(6).*

The sample size of 360 provided reasonably precise reference limits for each analyte using fractional power variables of age. Royston suggested 292 as a suitable sample size for constructing 95% continuous RIs, based on the standard error of the limits of the resulting RIs being no more than 10% of the age-specific standard deviation, using an even spread of samples over time *(15)*. Bellera *et al*. proposed a sample size of 436 to establish age-specific RIs in the presence of and accounting for inconsistent variation of the analyte with age *(30)*. In contrast, a simulation study suggested sample sizes greater than 500 may be required to identify subtle changes associated with age and sex *(31)*.

Few RI studies are conducted in children not referred for clinical testing, because of perceived difficulties in recruiting reference subjects. These challenges are particularly difficult in small laboratories, who may only perform occasional paediatric testing. Similar challenges in the pharmaceutical industry were assisted by the 1997 Food and Drug Modernization Act (and its subsequent reauthorization) incentive program. There have been almost 1000 paediatric trials since the enactment of the Food and Drug Modernization Act *(32)*. While full discussion is beyond the scope of this paper, the implementation of regulatory incentives for manufacturers of laboratory medicine in-vitro diagnostics (IVDs)

and laboratory information systems to partner with clinical care providers to assist with the development of paediatric RIs may address the paucity of paediatric normative data readily adaptable for clinical care.

In conclusion, the establishment of paediatric, age-specific RIs by collection and testing of specimens from healthy neonates and children of all years of life until 18 extends our understanding of physiological changes in biochemical analytes with age and will improve clinical result interpretation. The head-to-head comparison improves evidence-based, clinical decision making for biomarkers with specific differences between analyser type and / or analytical methodologies. Furthermore, the data clarify the true differences of biochemical values reported by different analysers when considering harmonized paediatric RIs. This should in turn lead to better care of children.

Role of the funding source

This study was primarily funded by The Royal Children's Hospital Foundation, with supplementary research funding from Ortho Clinical Diagnostics. In addition, the study received in-kind reagent supply from Medical and Scientific Affairs at Roche Diagnostics International. Funding bodies did not have any input in the design of the study, sample collection, analysis and interpretation of data or in preparation of the manuscript.

Acknowledgements

The authors thank staff of the Pathology Collection Department at The Royal Children's Hospital for obtaining the consent of participants and the collection of samples. The authors thank staff of the Anaesthetic, Surgical and Post-Natal Departments at the Royal Children's

Hospital, Northern Hospital and Western Health - Sunshine Hospital, Melbourne and collaborating laboratories.

Authors' Contribution

MH conducted the statistical analysis, analysed and interpreted all data and wrote the manuscript. SM and TY provided biochemistry expertise in the design of the study, interpreted biochemical data and revised the manuscript. VK wrote the study protocol, obtained ethics approval for the study, reviewed the manuscript and is the co-ordinator of the study. JB provided pathology collection expertise in the design of the study, training for the pathology collectors and reviewed the manuscript. J. Cowley reviewed the manuscript and one of the co-ordinator of the study. J. Carlin and SD provided advice and support on statistical analysis and reviewed the manuscript.VI contributed to the design of the study and was a major contributor in writing the manuscript. PM conceived the study and contributed to the design of the study and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Conflict of interest

There are no competing conflicts of interest to declare. The authors document supplementary study funding received from Ortho Clinical Diagnostics and in-kind reagent supply from Medical and Scientific Affairs at Roche Diagnostics International during the study.

REFERENCES

1. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: Is there a better way?. Am J Clin Pathol 2010;133:180-6.

2. Clinical and Laboratory Standard Institute (CLSI). Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline - third edition. CLSI document EP28-A3c. Wayne, PA: CLSI; 2010.

3. Daly CH, Liu X, Grey VL, Hamid JS. A systematic review of statistical methods used in constructing pediatric reference intervals. Clin Biochem 2013;46:1220-7.

4. Loh TP, Antoniou G, Baghurst P, Metz MP. Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. Pathology 2014;46:336-43.

5. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, Rauh M. Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clin Chem 2015;61:964-73.

6. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER White Paper. Crit Rev Clin Lab Sci 2017;54:358-413.

7. Schnabl K, Chan MK, Adeli K. Pediatric Reference Intervals: Critical Gap Analysis and Establishment of a National Initiative. EJIFCC 2008;19:115-22.

8. Delanghe JR, Cobbaert C, Galteau MM, Harmoinen A, Jansen R, Kruse R, *et al*. Trueness verification of actual creatinine assays in the European market demonstrates a disappointing variability that needs substantial improvement. An international study in the framework of the EC4 Creatinine Standardization Working Group. Clin Chem Lab Med 2008;46:1319-25.

9. Steinmetz J, Schiele F, Gueguen R, Ferard G, Henny J. Standardization of gamma-glutamyltransferase assays by intermethod calibration. Effect on determining common reference limits. Clin Chem Lab Med 2007;45:1373-80.

10. Wright EM, Royston P. Calculating reference intervals for laboratory measurements. Stat Methods Med Res 1999;8:93-112.

11. Higgins V, Adeli K. Advances in pediatric reference intervals: From discrete to continuous. J Lab Precis Med 2018;3:3.

12. Bussler S, Vogel M, Pietzner D, Harms K, Buzek T, Penke M, *et al*. New pediatric percentiles of liver enzyme serum levels (ALT, AST, GGT): Effects of age, sex, BMI and pubertal stage. Hepatology 2017; 68:1319-30.

13. Hoq M, Karlaftis V, Mathews S, Burgess J, Donath S, Carlin J, *et al*. A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the settings of clinical biochemistry, immunology and haematology: The HAPPI Kids Study. BMJ Open 2019:9:e025897.doi: 10.1136/bmjopen-2018-025897.

14. Attard C, van der Straaten T, Karlaftis V, Monagle P, Ignjatovic V. Developmental hemostasis: Age-specific differences in the levels of hemostatic proteins. J Thromb Haemost 2013;11:1850-4.

15. Royston P. Constructing time-specific reference ranges. Stat Med 1991;10:675-90.

16. Ridefelt P, Hilsted L, Juul A, Hellberg D, Rustad P. Pediatric reference intervals for general clinical chemistry components - merging of studies from Denmark and Sweden. Scand J Clin Lab Invest 2018;78:365-72.

17. Soldin S, Wong E, Brugnara C, Soldin P. Pediatric Reference Intervals. 7th Ed. Washington, DC: AACC Press; 2011.

18. Barbato G, Barini EM, Genta G, Levi R. Features and performance of some outlier detection methods. J Appl Stat 2011;38:2133-49.

19. Westgard QC. Desirable Biological Variation Database Specifications. https://www.westgard.com/biodatabase1.htm (Accessed February 2019).

20. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. Stat Med 2006;25:1369-82.

21. StataCorp. Stata statistical software : Release 15. Vol. College Station, TX: StataCropLP, 2017.

22. Australian Bureau of Statistics. 2016 Census QuickStats. https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/036 (Accessed May 2019).

23. Bailey D, Bevilacqua V, Colantonia DA, Pasic MD, Perumal N, Chan MK, Adeli K. Pediatric Within-Day Biological Variation and Quality Specifications for 38 Biochemical Markers in the CALIPER Cohort. Clin Chem 2014; 60:518-29

24. Burtis CA, Ashwood ER, Tietz NW, Bruns DE. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics. 4th Ed. St. Louis (Mo) : Elsevier Saunders; 2006.

25. Brackeen GL, Dover Js, Long CL. Serum albumin: Differences in assay specificity. Nutr Clin Pract 1989; 4:203-5

26. Tate JR, Sikaris KA, Jones GR, Yen T, Koerbin G, Ryan J, *et al*. Harmonising adult and paediatric reference intervals in Australia and New Zealand: An evidence-based approach for establishing a first panel of chemistryanalytes. Clin Biochem Rev 2014;35:213-35.

27. Royston P, and Altman. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. App Stat 1994; 43: 429–67.

28. Koenker R. Quantile Regression. New York: Cambridge University Press. 2005.

29. Kroon FPB, Ramiro S, Royston P, Cessie SL, Rosendaal FR, Kloppenburg M. Reference curves for the Australian/Canadian Hand Osteoarthritis Index in the middle-aged Dutch population., Rheumatology 2017; 56: 745–52.

30. Bellera CA, Hanley JA. A method is presented to plan the required sample size when estimating regression-based reference limits. J Clin Epidemiol 2007;60:610-5.

31. Griffiths JK, Iles TC, Koduah M, Nix AB. Centile charts ii: Alternative nonparametric approach for establishing time-specific reference centiles and assessment of the sample size required. Clin Chem 2004;50:907-14.

32. Dunne J, Rodriguez WJ, Murphy MD, Beasley BN, Burckart GJ, Filie JD, *et al*. Extrapolation of adult data and other data in pediatric drug-development programs. Pediatrics 2011;128:e1242-9.

3.3 Discussion

Most analytes (e.g sodium, potassium) have standardised methods of testing. As part of the quality assurance the laboratories or manufacturers are required to trace the rest results to the industry standard [4]. Hence, among the 30 analytes tested, the differences in mean test results between analysers were within the allowable limit of performance for 24 analytes. Whereas for the analytes which do not have any traceable industry standard (e.g. thyroid stimulating hormone), the test results differed by analysers [4].

Deciding the statistical method for comparing test results between analysers across age was challenging as there were only a few studies that compared test results between analysers [29, 30]. I initially considered correlation and the Bland Altman test to assess the bias [68]. However, this method was only suitable for comparing test results between two analysers[16]. Hence, I have applied mixed effect regression analysis taking both analyser and between sample variation into consideration where analysers were considered as a fixed effect and individual patient as a random effect. However, deciding whether test results of a biomarker vary among analysers based on the statistical evidence alone was considered inappropriate. Hence, variation between analysers was compared against the allowable limit of performance for the analytes.

Based on the findings of variation in test results across analysers, we estimated common continuous age-specific RIs. This is the first study which reported common continuous age-specific RIs across analysers. This finding will be helpful for small laboratories that do not have the capacity to establish continuous age-specific RIs for common biochemistry analytes. The Royal Children's Hospital, The Royal Melbourne Hospital and The Royal Women's

Hospital are currently validating these common continuous age-specific paediatric RIs for use across the three hospitals.

Several studies have reported continuous age-specific paediatric RIs [11, 12, 36]. However, considering the challenges in incorporating the continuous age-specific RIs into the LIS, these studies reported the RIs for every six month or one year age group [11, 12, 36]. Contrary to these studies, we have reported the continuous age-specific paediatric RIs as a function of age and sex using a novel statistical method [66]. These functions can be used for calculating RIs for any age between 1 month and 18 years and would facilitate incorporation of continuous age-specific RIs into the LIS. We have also reported 95% CIs for the RI limits, providing an indication of reliability, which was possible due to the regression-based parametric methods we applied [66]. Similarly, the CALIPER study also estimated continuous age-specific RIs applying nonparametric quantile regression and reported 95% CIs for the limits [36].

We could not incorporate neonates in the model-based analysis as there were not enough samples, and due to the dynamic nature of change in test results for most analytes between birth and the first month of life [66]. While we have provided summary statistics of the test results for neonates, we would have liked more samples between 72 hours and 30 days to model the change in biomarker with age.

Despite the results demonstrated, incorporating continuous RIs into existing LIS designed for discrete intervals remains challenging [69]. At present, LIS have limited capacity to visually display continuous reference curves or calculate continuous age-specific RIs from equations based on age and sex [66]. Moreover, some laboratories still report test results in a portable

document format (PDF) or hard copy, making it difficult to track sequential changes in test results appropriately over time. In addition to the challenges associated with the translation of continuous RIs into LIS, many pathology results are also viewed in other online systems, including electronic medical records (EMR), and Medical Practitioner desktop or mobile software packages. In these situations, developments are not only required to the LIS, but also to prepare the RIs, display them if this is required in a rendered format (e.g. PDF or directly on-screen) and also systematically transfer the results to receiving systems that have the capacity to accurately and safely render the results. This poses additional complexity for the electronic management of patient results and applicable RIs, but as the diversity in receiving systems and software expands, standardisation is vital. Recent studies endorsing the use of continuous paediatric RIs provide impetus for the development of improved modern LIS functionality and associated integrated EMR and other mobile and desktop software used to view and interpret patient results at the medical practitioner interface [11, 36, 47, 66].

Chapter 4: Validation of the HAPPI Kids continuous age-specific paediatric reference intervals

4.1 Introduction

The validation process is considered an internal quality assurance procedure. RI studies need to demonstrate the validity of the estimated RIs in the reference population in which they were initially intended to be used. However, no study except for CALIPER and Zierk *et al* has validated the estimated continuous RIs (2, 3).

For the HAPPI Kids continuous RIs, reported in chapter 3, to be adopted by routine laboratories, they need to be validated for the local population, or in laboratories using the same analytical methods and analysers [66]. However, there is no recommendation available in the current CLSI guidelines regarding the process of validation for continuous RIs. Addressing this gap, this chapter demonstrates that the current CLSI guidelines can be applied to validate the continuous RIs, with some minor modifications.

This chapter reports the validity of the HAPPI Kids continuous RIs and highlights important issues regarding the validation process for continuous RIs. These findings are useful for laboratories and policymakers in adopting continuous RIs in the clinical/pathology settings. The findings were published in the Journal of Applied Laboratory Medicine in February 2020 and the author accepted manuscript is presented as the body of this chapter [70].

4.2 Publication

Validation of the HAPPI Kids continuous age-specific paediatric reference intervals

Running head: Validation of the HAPPI Kids continuous paediatric RIs

Monsurul Hoq

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Susan Matthews

The Royal Children's Hospital, Parkville, Australia; International Centre for Point of Care Testing, Flinders University, Bedford Park, Australia.

Vicky Karlaftis

Murdoch Children's Research Institute, Parkville, Australia.

Janet Burgess

The Royal Children's Hospital, Parkville, Australia.

Susan Donath

Murdoch Children's Research Institute; Parkville, Australia; The University of Melbourne; Parkville, Australia.

John Carlin

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Vera Ignjatovic*

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia.

Paul Monagle*

Murdoch Children's Research Institute, Parkville, Australia; The University of Melbourne, Parkville, Australia

On behalf of the HAPPI Kids study team

*Contributed equally

Corresponding Author:

Associate Prof Vera Ignjatovic, PhD, B.Sc. (Hons)

Co-Group Leader, Haematology Research Laboratory, Murdoch Children's Research Institute

Principal Fellow, Department of Paediatrics, The University of Melbourne

Address

Haematology Research Laboratory

Murdoch Children's Research Institute

50 Flemington Road, Parkville, Victoria, 3052, Australia

Email: vera.ignjatovic@mcri.edu.au, T: +61 3 9936 6520, M: +61 439 363 891

Key words

Validation, paediatric, reference interval, biochemistry, age-specific

List of abbreviations

Reference Interval (RI)

The Clinical and Laboratory Standards Institute (CLSI)

Harmonizing Age Pathology Parameters in Kids (HAPPI Kids)

Ethics in Human Research Committee (HREC)

Coefficient of variation (CV)

Coefficient of variation within samples (CV_I)

Canadian Laboratory Initiative on Paediatric Reference Intervals (CALIPER)

ABSTRACT

Introduction

To facilitate best possible patient care, reference intervals (RIs) adopted by a laboratory must be appropriate for the population demographics and where applicable, the analytical principle and/or the analytical instrument used. Whilst guidelines from the Clinical and Laboratory Standard Institute recommend a validation process for discrete RIs, there are no current recommendations for the validation process for continuous RIs. This study aimed to validate recently published, HAPPI Kids continuous RIs, in a routine laboratory.

Methods

Initially, the difference in test results between the primary study laboratory that contributed to previous RIs development and a routine laboratory was assessed using specimens from 77 children tested in both laboratories using the Siemens ADVIA 1800 or Centaur/XP/XPT.

Later, validation of the HAPPI Kids RIs was undertaken using 279 paediatric samples tested on the same analyser type in the routine laboratory. The previously published RIs were validated if more than 90% of results in the routine laboratory were within the RIs.

Results

There was minimal evidence of clinically significant differences in test results between the primary and routine laboratories. The continuous RIs were validated after initial analysis for 16 of the 18 biochemistry analytes tested, and after secondary analysis for the remaining 2 analytes.

Conclusion

This study validates the HAPPKI Kids RIs in a routine laboratory, satisfying the laboratory accreditation requirements for evaluation, implementation and sourcing of RIs. In addition, this study presents a modification of the current CLSI method for validation of continuous RIs that will benefit routine laboratories in general.

IMPACT STATEMENT

According to the Clinical Laboratory and Science Institute (CLSI), previously published reference intervals (RIs) need to be validated prior to use. However, there is currently no specific guidelines for the validation process when it comes to continuous RIs. Addressing this gap, our study validates, the recently published HAPPI Kids RIs, in a routine laboratory, satisfying the laboratory accreditation requirements for evaluation, implementation and sourcing of RIs. In addition, this study presents a modification to the current CLSI method for validation of continuous RIs that will benefit routine laboratories in general.

INTRODUCTION

The majority of clinical pathology laboratories do not have the resources to prospectively determine paediatric reference intervals (RIs) specific for their setting (e.g. population, analyser, reagents) *(1)*. For this reason, paediatric RIs used in laboratories are historically derived from: i) clinical decision limits; ii) published professional recommendations; iii) settings by regulatory bodies, iv) post-analytical quality control data, or v) from large scale population based studies *(2)*. The International laboratory accreditation requirements (e.g. ISO15189) mandate that laboratories verify RIs as appropriate for the methodologies and the patient population they serve. The Clinical and Laboratory Standards Institute (CLSI) provides guidelines for assessing the validity of a discrete RI by using 20 samples from healthy children, in the scenario where population characteristics and pre-analytic variables are consistent with the population used for the established RIs *(3,4)*. If 18 or more (≥ 90%) of 20 samples fall within the discrete 95% RI, the established RIs are accepted for use *(3, 4)*.

Recent studies have published continuous RIs and demonstrated improvement in test result interpretation and clinical decision making *(5-8)*. To improve paediatric RIs for Australia, the Harmonising Age Pathology Parameters in Kids (HAPPI Kids) study has published continuous RIs for children aged 30 days to < 18 years for 30 biochemistry analytes *(6)*. For these continuous RIs to become widely adopted, validation by routine laboratories is important *(4)*. However, there are currently no guidelines for validating such RIs. Hence, this study aimed to validate the HAPPI Kids RIs in a routine laboratory with some modification to the current CLSI guidelines.

METHODS

Study protocol and statistical analysis

This study followed-up our previously published HAPPI Kids study reporting continuous RIs for 30 biochemistry analytes *(6)*. The Royal Children's Hospital, Melbourne, Australia, Ethics in Human Research Committee (HREC) and subsequently HRECs of all participating hospitals approved the study protocol (HREC 34183 A). The study contributors are listed in the Supplementary Table 1 of the Appendix C. All statistical analysis was conducted in Stata 15.1 *(9)*.

Assessment of difference in test results between laboratories

Within the recruited paediatric samples used for establishing the HAPPI Kids RIs, aliquots from 77 study participants were subsequently tested in a routine laboratory for as many as 18 analytes as sample volume allowed (Supplementary Figure 1 of the Appendix C). Matched analysis of these results with the paired results available from the HAPPI Kids study, enabled assessment of difference in test results between the primary HAPPI Kids study laboratory and the routine laboratory. Similar to the HAPPI Kids RIs, these aliquots were analysed using a Siemens ADVIA 1800 or a Siemens ADVIA Centaur/XP/XPT for 18 analytes. Due to insufficient blood volume from each participant, sample numbers available from the routine laboratory were less than 77 for many of the analytes. The list of analytes and the analytical methods used in both laboratories were identical and are provided in the Supplementary Table 2 of the Appendix C.

Outliers were removed using Tukey's test based on the difference between the paired results. The mean difference in test results was assessed using 95% confidence interval and paired t-test. The coefficient of variation (CV) for each study participant was defined as half of the

absolute difference (equivalent to the standard deviation in case of n=2) divided by the mean of test results from the two laboratories. The differences in test results between the two laboratories for each analyte were considered clinically insignificant if the median of the CVs was less than individual biologic variation (CV_I) reported in the Westgard QC database, mostly determined in adults *(10)*. If the difference was clinically insignificant the HAPPI Kids RI for the analyte was considered for validation in the routine laboratory.

Assessment of validation of the HAPPI Kids RIs in a routine laboratory

A total of 279 paediatric subjects, aged 30 days to < 18 years, undergoing minor elective day surgery were recruited from the outpatient service of four major public hospitals in Victoria, Australia for the validation of the HAPPI Kids RIs and are referred to as the 'validation cohort' (Supplementary Figure 1 of the Appendix C). Details of the participant exclusion criteria, blood collection method, sample processing and storage condition prior to analysis have been published as part of the HAPPI Kids protocol *(11)*. Briefly, participants with presence of systematic abnormalities were excluded and only one attempt was made to collect the blood per participant. Blood sample for each participant was centrifuged at 5000 rpm (4472g) at 6 °C for 5 min, separated into 400-µL aliquots, and stored at -80 °C within 4 h of collection. These samples were sent to the routine laboratory and were analysed on a Siemens ADVIA 1800 or a Siemens ADVIA Centaur/XP/XPT for a maximum of 18 analytes as specified above (Supplementary Table 2 of the Appendix C**)**.

Outliers were identified for exclusion based on the residual being greater than 3 times the inter-quartile range as a conservative approach considering samples were obtained from healthy children, with residuals defined as the differences between observed and predicted values from a fractional polynomial regression model for the mean. For the remaining data on

each analyte, the proportion of test results from the validation cohort within the HAPPI Kids RIs across age and gender was calculated. The HAPPI Kids RIs were considered valid if more than 90% of the test results from the validation cohort were within the continuous RIs *(3, 4, 12)*. Further secondary analysis of test results in the routine laboratory was done based on age and date of analysis if the HAPPI Kids RIs were initially invalid.

RESULT AND DISCUSSION

The ethnic distribution of the participants was consistent with the HAPPI Kids study participants and the Australian population, as 92.5% were born in Australia with 52.5% of both of the participants' parents born in Australia *(13)*. The samples were uniformly distributed across age and with approximately equal numbers of male and female for every year of life from 30 days to < 18 years.

Agreement between the two laboratories for all analyses are presented in the Supplementary Figure 2 of the Appendix C**.** There was evidence supporting mean differences in test results between the two laboratories for sodium, urea, creatinine, alkaline phosphatase, alanine amino transferase, aspartate amino transferase, calcium, magnesium, phosphate, free triiodothyronine and free thyroxine (Table 1). However, these differences were considered clinically insignificant since the median CVs were less than CV_I reported in the Westgard QC database.

Table 1: The mean difference (95% CI) and median coefficient of variation (in comparison to Westgard) in reference values between two laboratories for 18 biochemistry analytes on Siemens analyser

Analyte	Total number of samples tested, n_1[a]	Outliers excluded, n_0	n = $n_1 - n_0$	rho[b]	Range[c]	Mean difference[d]	95% CI	CV_I[e] (%)	
								Observed (Median)	Westgard *(14)*
Sodium (mmol/L)	76	0	76	0.55	137, 145	**0.76**	**0.44, 1.08**	0.5	0.6
Potassium (mmol/L)	75	2	73	0.97	3.7, 4.9	0.01	0.00, 0.03	0.0	4.6
Chloride (mmol/L)	76	0	76	0.83	101, 111	0.07	-0.20, 0.33	0.6	1.2
Urea (mmol/L)	72	2	70	0.99	2.3, 7.0	**0.07**	**0.030, 0.10**	1.9	12.1
Creatinine (µmol/L)	73	4	69	0.98	11.4, 82.4	**2.22**	**1.65, 2.80**	4.3	6.0
Alkaline Phosphate (U/L)	74	2	72	0.98	47, 513	**13.58**	**11.02, 16.15**	5.1	6.5
Alanine Amino Transferase (U/L)	73	2	71	0.84	9, 43	**3.48**	**2.79, 4.17**	12.9	19.4
Aspartate Amino Transferase (U/L)	73	1	72	0.97	19, 66	**1.33**	**0.77, 1.90**	5.3	12.3
γ-Glutamyl Transferase (U/L)	73	1	72	0.98	5.0, 28.0	0.04	-0.24, 0.41	5.9	13.4
Total Protein (g/L)	74	2	72	0.97	57, 83	0.08	-0.30, 0.46	1.1	2.75
Total Bilirubin (µmol/L)	73	3	70	0.98	1.5, 19.5	0.03	-0.13, 0.19	0	21.8
Calcium (mmol/L)	71	3	68	0.80	2.18, 2.70	**0.03**	**0.02, 0.05**	1.4	2.1
Magnesium (mmol/L)	71	1	70	0.66	0.72, 1.07	**0.05**	**0.04, 0.06**	3.4	3.6
Albumin (g/L)	72	3	69	0.96	34.0, 52.5	-0.17	-0.36, 0.02	1.3	3.2
Phosphate (mmol/L)	71	1	70	0.98	0.88, 2.21	**0.03**	**0.02, 0.04**	1.9	8.2
Thyroid Stimulating Hormone (mlU/L)	68	1	67	0.99	0.42, 4.88	-0.03	-0.06, 0.01	2.9	19.3

Free Triiodothyronine (ρmol/L)	75	0	75	0.90	3.9, 8.2	**0.30**	**0.21, 0.37**	3.0	7.9
Free Thyroxine (ρmol/L)	75	3	72	0.87	10.7, 22.0	**-0.31**	**-0.58, -0.05**	3.3	5.7

The evidence of a difference in test results between the two laboratories is highlighted by bold font.

a. Due to insufficient blood volume from each participant, sample numbers available from the routine laboratory were less than 77 for many of the analytes.

b. Concordance correlation coefficient

c. Range of the average reference values between shared samples

d. Using a paired t-test

e. Coefficient of variation within samples

The proportion of test results from the validation cohort that were within the HAPPI Kids RIs, was more than 90% for all analytes except creatinine and albumin **(**Table 2, Supplementary Figure 3 of the Appendix C). The RIs for albumin were considered valid after exclusion of the albumin test results (n=67 from 278), analysed on 2 individual days with nearly 50% of results outside the RIs (Table 2, Supplementary Document 1 of the Appendix C). The RI for creatinine appears to be valid for children older than 2 years (Table 2, Supplementary Document 1 of the Appendix C).

The assessment of difference in test results between laboratories and/or analysers or reagents is feasible for a laboratory with different testing sites or when a laboratory introduce a new analysers or reagents. However, this may not be feasible when adopting established RIs as samples tested for establishing RIs are not shared for validation. Testing difference in test results is also not essential to validate the established RIs. However, this was useful in explaining the excess of departures from the reference limit at the lower end for both sodium and γ-glutamyl transferase, and the high end for chloride, compensated by a deficit at the other end, i.e. that the 90% was achieved in an asymmetrical way.

Table 2: Proportion of samples that were within the HAPPI Kids RIs for 18 Biochemistry analytes

Analytes (unit)	No. of outlier excluded	No. of sample used for validation	The HAPPI Kids RI *(6)*[c,d]		Proportion of samples within The HAPPI Kids RIs (%)
			Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)	
Sodium[a] (mmol/L)	1	278	137.04 + 0.07 Age	142.20+0.09 Age	91.4 (87.4, 94.2)
Potassium[a] (mmol/L)	4	276	4.03 – 0.02 Age	5.05 – 0.02 Age	91.6 (87.7, 94.4)
Chloride[a] (mmol/L)	3	276	103.68 – 0.13 Age	109.58 – 0.02 Age	93.8 (90.3, 96.1)
Urea[a] (mmol/L)	0	279	2.34 + 0.65 Male – 0.02 Age*Male	6.24 + 0.02 Age – 0.08 Male + $5.29*10^{-3}$ Age*Male	93.2 (89.5, 95.6)
Calcium[a] (mmol/L)	1	278	2.22 + 0.03 Age^{-1}	2.55 + 0.03 Age^{-1}	91.0 (87.0, 93.9)
Magnesium[a] (mmol/L)	2	277	0.80 – 0.03 ln(Age) + 0.01 Male	1.01 – 0.04 ln(Age) + 0.02 Male	91.7 (87.8, 94.4)
Phosphate[a] (mmol/L)	0	279	1.37 + 0.15 Age^{-1} – $8.36*10^{-5}$ Age^3	2.00 + 0.14 Age^{-1} – $6.82*10^{-5}$ Age^3	93.5 (90.0, 95.9)
Alkaline Phosphatase[a] (U/L)	2	277	153.100 – 0.146 Age^3 + 0.043 Age^3ln(Age) + 0.129 Male*Age^3 – 0.043 Male*Age^3ln(Age)	387.700 + 0.216 Age^3 – 0.089 Age^3ln(Age) + 0.338 Male*Age^3 – 0.115 Male*Age^3ln(Age)	92.4 (88.6, 95.0)
Aspartate Amino [a] Transferase (without P5P, U/L)	1	273	26.85 – 1.01 ln(Age) – 1.22 $(ln(Age))^2$ + 1.57 Male	54.07 – 8.27 ln(Age) + 0.40 $(ln(Age))^2$ + 3.16 Male	96.0 (92.8, 97.8)
Alanine Amino Transferase [a] (without P5P, U/L)	10	269	8.86 + 2.25 Age^{-1} + 0.95 Male	32.48 + 0.25 Age^{-1} – 1.48 Male	91.8 (87.9, 94.6)
γ-Glutamyl Transferase[a] (U/L)	10	269	5.67 + 0.09 Age	16.05 + 0.82 Age	91.4 (87.4,94.3)

Albumin [a] (BCP, g/L)	Total	1	278	34.23 + 0.07 Age	41.47 + 0.31 Age	**79.1 (73.9, 83.5)**
	After exclusion		211			**89.6 (84.6, 93.1)**
Total Protein[a] (g/L)		1	278	53.70 + 3.90 ln(Age)	66.05 + 5.40 ln(Age)	90.6 (86.6, 93.6)
Thyroid Stimulating Hormone[a] (mlU/L)			267	0.89	5.20	92.1 (88.2,94.8)
Creatinine (Jaffe, µmol/L)	30 day to 18 years	0	279	12.34 + 1.70 Age + 2.27 Male – 0.46 Male*Age	22.97 + 2.96 Age – 4.72 Male +1.12 Male*Age	**86.4 (81.2, 89.9)**
	2 – 18 years		257			90.2 (86.0, 93.3)
Total Bilirubin[b] (µmol/L)			270	1.67 + 0.19 Age	9.28 + 1.35 Age	97.0 (94.2, 98.5)
Free Thyroxine [b] (ρmol/L)		1	278	$13.13 – 8.20*10^{-3}Age^3 – 0.72$ Male $+ 4.43*10^{-3}$ Male*$Age^{-3} + 2.71*10^{-3}$ $Age^3 \ln(Age^3) - 1.42*10^{-3}$ $Age^3 \ln(Age)$*Male	$20.02 – 6.91*10^{-3}Age^3 –$ 2.00 Male $+ 8.06*10^{-3}$ Male $Age^{-3} + 2.25*10^{-3}$ $Age^3 \ln(Age^3) – 2.42*10^{-3}$ $Age^3 \ln(Age^3)$ Male	95.0 (91.7, 97.0)
Free Triiodothyronine [b] (ρmol/L)		0	279	5.87 – 0.07 Age- 0.97 Male + 0.09 Age*Male	9.43 –0.10 Age– 0.9 Male + 0.07 Age*Male	92.8 (89.1, 95.3)

The proportion < 90% were highlighted by bold font.

a. Common RI for Abbott Architect, Roche Cobas, Siemens ADVIA, Ortho Vitros and Beckman Coulter

b. Analyser specific RI for Siemens ADVIA

c. In years

d. For analytes with difference in reference limit by sex, replace male with “1”, if subject is male. Otherwise replace male with “0” if subject is female, which is the reference category.

The initial failure to validate RIs for albumin prompted further analysis of the routine laboratory albumin results and the albumin quality control (QC) data since pre-analytic, analytic and post-analytical factors have been reported as important when establishing and validating RIs (Supplementary Document 1 of the Appendix C) *(14-15).* Analysis of the laboratory's QC data for albumin demonstrated stable assay performance at two levels over the duration of the specimen testing period (Supplementary Document 1 of the Appendix C). In contrast to the stability of the albumin QC data, the median albumin results from the routine laboratory were higher on 2 individual days of testing, which may be suggestive of variations in pre-analytical specimen handling (e.g. specimen mixing) in the routine laboratory for these batches. Similar analysis of study participant's test results for calcium and phosphate, ions that are albumin bound, did not show the same trend. The CLSI recommendation to test a set of new samples when initial validation attempts fail, was not possible for this study due to restricted sample volume. Instead we excluded the poorly performing batches for albumin for which RIs were considered valid. However, we acknowledge that the best laboratory practice would be to repeat the validation for albumin RIs using a new set of specimens.

The initial failure to validate RIs for creatinine was less concerning, given the known interference of bilirubin in Jaffé creatinine method for children < 1 year, such as that used in this study *(14, 16).* The coefficient of variation between analysers for the Jaffé creatinine method in children < 2 years reported in our previous study suggests that reference values for early years of life vary by > 30% between analysers *(6, 10).* The linear model for the reference limits starting from as early as 30 days perhaps failed to reflect the rapid decline in the first month of life as renal function matured. Hence, the age range was limited to children aged 2 to < 18 years for creatinine RIs and was considered valid.

For validating the continuous RIs, we have adopted modified process of the current CLSI guidelines. We have considered total proportion of samples outside the RIs across all age and sex as a measure of validation *(12)*. The Canadian Laboratory Initiative on Paediatric Reference Intervals (CALIPER) study used a similar approach addressing the limitation of the CLSI guidelines, which considers RIs for all groups are valid if RI for any group is valid *(12).* We have considered 279 samples distributed uniformly across age and with approximately equal numbers between males and females sufficient to estimate a proportion of 0.90 with a standard error of the proportion of 0.018. However, most laboratories cannot test large number of paediatric samples for validation. Hence, based on analysis of probability of validating 95% RIs (i.e. RIs including 95% of a healthy cohort) as a function of sample size when at least 90% of the samples fall within the intervals, we would recommend at least 100 paediatric samples are tested with a 98.9% chance of validating the continuous RIs. This approach is further explained in the Supplementary Document 2 of the Appendix C. The CALIPER study also used 100 samples for validation of the transferred RIs from Abbott analyser to other analytical platforms for given analytes *(12).*

In conclusion, the published HAPPI Kids RIs were validated for 16 of the 18 analytes investigated after primary analysis for use in a routine Australian laboratory, thereby satisfying International laboratory accreditation requirements for evaluation, validation/verification and sourcing of RIs for the laboratory's local population. Validation of the RIs for albumin and creatinine were achieved by exclusion of a subset of data and by limiting the RIs for children aged 2 to <18 years, respectively. Moreover, this study demonstrates that with some modification the current CLSI guidelines for validation of discrete age-grouped RIs can be used for validation of continuous age-specific RIs. This will

be important as laboratory information systems evolve to fully accommodate continuous RI reporting for paediatric patients.

Acknowledgements

The authors thank staff of the Pathology Collection, Anaesthetic, Surgical, and Neonatal and Post-Natal Departments at the Royal Children's Hospital, Royal Women's Hospital, Northern Hospital, and Western Health—Sunshine Hospital, Melbourne. The authors also acknowledge the contribution of chemical pathologists and clinical biochemists at collaborating laboratories: Royal Children's Hospital, Royal Melbourne Hospital, Melbourne Pathology, Monash Health, Australian Clinical Labs, and Dorevitch Pathology.

Research funding

This study was primarily funded by The Royal Children's Hospital Foundation, with supplementary research funding from Ortho Clinical Diagnostics. In addition, the study received in-kind reagent supply from Medical and Scientific Affairs at Roche Diagnostics International. M. Hoq was supported by NHMRC Post Graduate Scholarship.

Role of Sponsor

The funding organizations played no role in the design of the study, sample collection, analysis and interpretation of data or in preparation of the manuscript.

Conflict of interest

There are no competing conflicts of interest to declare.

Authors' Contribution

M. Hoq conducted the statistical analysis, analysed and interpreted all data and wrote the manuscript. S. Matthews provided biochemistry expertise in the design of the study, interpreted biochemical data and revised the manuscript. V. Karlaftis wrote the study protocol, obtained ethics approval for the study, reviewed the manuscript and is the co-ordinator of the study. J. Burgess provided pathology collection expertise in the design of the study, training for the pathology collectors and reviewed the manuscript. J. Carlin and S. Donath provided advice and support on statistical analysis and reviewed the manuscript. V. Ignjatovic contributed to the design of the study and was a major contributor in writing the manuscript. P. Monagle conceived the study and contributed to the design of the study and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

REFERENCE

1. Tate JR, Koerbin G, Adeli K. Opinion paper: deriving harmonised reference intervals - global activities. EJIFCC 2016; 27: 48-65.

2. Sikaris K. Application of the stockholm hierarchy to defining the quality of reference intervals and clinical decision limits. Clin Biochem Rev 2012; 33:141-8.

3. Clinical and Laboratory Standard Institute (CLSI). Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline. CLSI document EP28-A3c. 3rd Ed. Wayne (PA): CLSI; 2010.

4. Jones GRD. Validating common reference intervals in routine laboratories. Clin Chim Acta 2014; 432:119-21.

5. Asgari S, Higgins V, McCudden C, Adeli K. Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally

partitioned reference intervals. [Epub ahead of print] Clin Biochem August 22, 2019 as doi: 10.1016/j.clinbiochem.2019.08.010.

6. Hoq M, Matthews S, Karlaftis V, Burgess J, Cowley J, Donath S, *et al*. Reference values for 30 common biochemistry analytes across 5 different analyzers in neonates and children 30 days to 18 years of age. Clin Chem 2019; 65:1317-26.

7. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, Rauh M. Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clin Chem 2015; 61:964-73.

8. Loh TP, Antoniou G, Baghurst P, Metz MP. Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. Pathology 2014; 46:336-43.

9. StataCrop. Stata statistical software: release 15. College Station (TX): StataCrop LP; 2017.

10. Westgard QC. Desirable biological variation database specifications. https://www.westgard.com/biodatabase1.htm (Accessed February 2019).

11. Hoq M, Karlaftis V, Mathews S, Burgess J, Donath SM, Carlin J, *et al*. A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: The HAPPI Kids study protocol. BMJ open 2019; 9:e025897.

12. Estey MP, Cohen AH, Colantonio DA, Chan MK, Marvasti TB, Randell E, *et al*. CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens clinical chemistry assays: Direct validation using reference samples from the CALIPER cohort. Clin Biochem 2013; 46:1197-219.

13. Australian Bureau of Statistics. 2016 Census Quick-Stats. https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/036 (Accessed May 2019).

14. Cuhanda, S., Preanalytical variables and factors that interfere with the biochemical: a review. OA Biotechnology 2013. 2(2): p. 19.

15. De la Salle, B., Pre- and postanalytical errors in haematology. International Journal of Laboratory Hematology, 2019. 41(S1): p. 170-176.

16. Gencheva II, Ruseva AL. Effects of glucose and bilirubin on the kinetic jaffe's and the enzymatic methods for serum creatinine assay. J Biomed Clin Res 2015; 8:3

4.3 Discussion

Whilst continuous age-specific RIs have improved our understanding of the changes in analyte concentrations with age, there was no recommendation or endorsed guideline for the validation of continuous age-specific RIs for use by a routine laboratory. In contrast, the CLSI's published guidelines for validation are available for the validation of discrete age-group RIs [16]. It is uncertain to what extent these guidelines are applicable to continuous age-specific RIs. As such, we developed a method to validate the continuous age-specific RIs with some modification of the current guidelines [70]. Publication and endorsement of a standard protocol for continuous age-specific RIs validation would provide direction and support for the adoption of continuous age-specific paediatric RIs in local laboratories.

Chapter 5: Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review

5.1 Introduction

Identification of an appropriate statistical method for estimating age-specific RIs depends on the choice between considering age as a discrete or continuous variable [52]. In addition, there are a wide range of statistical methods available for estimating age-specific RIs with no clarity around their relative merits nor the extent to which any method might be regarded as better in practice. Hence, it is important to examine the range of statistical methods used for the estimation of age-specific RIs and identify trends in usage and reporting. These findings may provide a motivation for performing statistical investigation comparing different methods in various scenarios.

This chapter reports the findings from a systematic review of original research publications estimating age-specific paediatric RIs, with a focus on the statistical methods applied. This chapter highlights whether the statistical methods applied in different steps of the estimation process were reported. This chapter also reports the trends in the uptake of statistical methods for producing age-specific RIs, especially for analytes that display strong age-dependence. The findings of the systematic review were published in Clinical Biochemistry and author accepted manuscript is included as the body of this chapter [71].

5.2 Publication

Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: a systematic review

Running title: Statistical methods used in the estimation of PRIs

Monsurul Hoq[a. b], Louise Canterford[a], Susan Matthews[c], Gulshan Khanom[a], Vera Ignjatovic[a,b], Paul Monagle[a,b,d], Susan Donath[a,b], John Carlin[a,b] on behalf of the HAPPI Kids study team

a. Murdoch Children's Research Institute, 50 Flemington Road, Parkville, VIC 3052, Australia
b. The University of Melbourne, 50 Flemington Road, Parkville, VIC 3052, Australia.
c. International Centre for Point of Care Testing, Flinders University, Bedford Park, SA 5042, Australia.
d. The Royal, Children's Hospital, 50 Flemington Road, Parkville, VIC 3052, Australia.

Declarations of interest: none

Corresponding Author:

Monsurul Hoq, MPH, MSc (Statistics), BSc (Statistics)

Biostatistician, Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute

PhD Candidate, Department of Paediatrics, the University of Melbourne

50 Flemington Road, Parkville, VIC 3052, Australia

E: monsurul.hoq@mcri.edu.au, T: +610399366264

ABSTRACT

Introduction

There is an emerging realisation that paediatric reference intervals (RIs) estimated using discrete age-groups may be misleading, especially close to age cut-off values. This limitation has been addressed by estimating RIs that vary continuously with age. This systematic review examines the range of statistical methods used over the past 25 years for estimation of age-specific RIs, and identifies trends in usage and reporting.

Methods

Literature searches were conducted using predefined search criteria for original publications between 1993 and 2018 on the MEDLINE and Embase databases. Data related to sample size, treatment of age (as categorical or continuous), and statistical methods were extracted from the selected publications.

Results

A total of 238 publications were reviewed. Not all publications reported the statistical methods used in different steps. Among the publications, 167 (70%) reported discrete age-group RIs, 54 (23%) reported continuous RIs and 17 (7%) reported both types of RIs. The nonparametric statistical method was commonly used for discrete age-group RIs (64%, n=117), whereas a wide variety of curve-fitting approaches, including Cole's lambda-mu-sigma method (28%, n = 20), parametric curve-based methods (28%, n = 20), generalised additive model for location shape and scale method (13%, n = 9) and quantile regression (11%, n = 8) were used for continuous RIs.

Conclusions

Improvement in the reporting of statistical methods used for estimating age-specific paediatric RIs is required. There has been insufficient uptake of methods for producing continuous RIs, especially for biomarkers that display strong age-dependence.

Keywords: Age-specific, continuous, reference intervals, systematic review, statistical methods

Abbreviations

Reference intervals (RIs)

Confidence intervals (CIs)

The Clinical Laboratory and Science Institute (CLSI)

Standard deviation (SD)

Lambda-mu-sigma (LMS)

Generalised additive model for location shape and scale (GAMLSS)

Inter quartile range (IQR)

Equivalent degree of freedom (EDF)

Laboratory information system (LIS)

Electronic medical recording (EMR)

INTRODUCTION

Reference intervals (RIs) are essential clinical assessment tools used to identify whether a laboratory test result falls within an acceptable range of population values. The conventional statistical definition of a $(100 - \alpha)\%$ RI is the range between the $\alpha/2$ and $(100 - \alpha/2)$ percentiles $(0 < \alpha < 100)$ [1]. Whilst the definition of a RI is uncomplicated, estimating

paediatric RIs, where biomarkers may change with age, involves complexities associated with reference population definition, blood collection, laboratory testing and statistical analysis [2]. The statistical analysis of the laboratory data is a critical component of this process and consists of outlier identification, sample distribution, RI estimation and calculation of 95% confidence intervals (CIs) [2, 3].

Traditionally, the approach to addressing variation in reference values with increasing age, sex or ethnicity in paediatric populations, has been to estimate RIs for specific age groups, sometimes by sex or ethnicity [2, 4]. Estimating RIs for subgroups of sex or ethnicity is reasonable as there is no ordering among the subgroups [5]. However, estimation of these "discrete RIs" by age-group instead of continuous age, without clear clinical justification for the age partitions, may lead to misdiagnosis of children whose ages fall close to the specified age cut-off values [4-6]. The misdiagnosis of children may lead to repeated blood tests, additional follow-ups leading to increased utilisation of health service and anxiety for parents and children, or mistreatment [7]. As such, there is an emerging realisation that the clinical limitations of discrete RIs can be improved by developing RIs that vary continuously with age, which we will refer to as "continuous RIs" [4, 6]. For example, the continuous RIs capture the steady increase in creatinine with age or in alkaline phosphatase during puberty better than discrete RIs [8-10]. The choice between discrete or continuous RIs has implications for each step of the statistical analysis.

Identification of outliers is an important consideration in the estimation of RIs from a dataset. The Clinical Laboratory and Science Institute (CLSI) recommends outlier-screening tests by Dixon and Tukey [2]. These methods are suitable for univariate analysis settings and identify

extreme values based on the overall data. However, for age-dependent biomarkers, methods that take account of changes with age are needed to identify age-specific extreme values.

An assumption of the majority of statistical methods commonly applied is that observations are normally distributed. It is important to examine this assumption against the data in order to select the most appropriate statistical method. The CLSI guidelines do not specifically recommend a normality assessment process, but visual data inspection is preferred [2]. Nevertheless, formal statistical tests of univariate normality [11] are often adopted.

For age-specific RIs, the CLSI recommends partitioning data by age-group followed by nonparametric ranking of observations (smallest to largest) and determination of percentile values ($\alpha/2$ and $(100 - \alpha/2)$) for each age-group [2]. Alternatively, parametric methods based on the normal distribution can be utilised, whereby the p^{th} centile value is estimated as: $\bar{x}$ +/- $z_p s$ (where $\bar{x}$ is the mean value of the measurement; z_p is the p^{th} centile of the standard normal distribution with mean 0 and variance 1 and s is the standard deviation (SD)) [2].

An alternative to estimating discrete age-group RIs was first proposed in 1988 by Cole, who estimated parameters of a Box-Cox transformation for subgroups of age, known as the lambda-mu-sigma (LMS) method, and used these parameters for the estimation of smooth reference centile curves [12]. In 1991, Royston proposed an alternative method for continuous age-specific RIs using flexible parametric curves [1]. In 1992, Cole and Green proposed a modification of Cole's LMS method by estimating the parameters using cubic splines for continuous age to estimate smooth reference centile curves [13]. Most recently, in the mid-2000s, Rigby and Stasinopoulos proposed a new method called the generalised additive model for location shape and scale (GAMLSS), which can be applied for any

distribution type across continuous age [14]. Additionally, quantile regression, which estimates a quantile of the dependent variable (without assuming any parametric distribution) as a smooth function of the independent variable has also been used in estimating continuous RIs [15, 16]. These methods have been commonly used in other disciplines of medicine e.g. child growth chart, kidney function, hand osteopathy, etc to establish centile charts for continuous age [17-19].

Sample size is a major consideration for any study aiming to estimate age-specific paediatric RIs. The recommendation by CLSI for minimum subgroup sample size required for 95% RI (i.e. $\alpha = 5\%$) is 120, because the 2.5^{th} and 97.5^{th} percentiles correspond to the 3^{rd} and 118^{th} ranked observations, respectively. For biomarkers that change with age, discrete RIs usually comprise 2 to 7 age-groups between 0 to 19 years, resulting in 240 to 840 total observations. For continuous RIs, Royston recommended a minimum sample size of 292, based on requiring the standard error of the reference limits to be less than 10 percent of the residual standard deviation from the parametric model. This assumes the samples are uniformly distributed across the age range [1]. Bellera *et al*. recommend greater than 400 samples for estimating continuous RIs, based on the complexity of the relationship of biomarkers with age [20].

In summary, there are several methods available for each steps of the statistical analysis for estimating age-specific paediatric RIs, which vary depending on the choice between estimating discrete or continuous RIs. However, at present no systematic review of the statistical methods used in the estimation of age-specific paediatric RIs for laboratory blood tests has been published. Whilst Daly *et al*. reviewed current practices when constructing paediatric RIs, key search terms were not included (e.g. “reference curve," “centile chart,"

and “centile curve″), which may have influenced the study findings to age partitioning [3]. Hence, the aim of this systematic review was to evaluate age-specific paediatric RI development studies, following the advent of methodology for continuous RIs. This will serve to overview (i) the completeness of reporting of statistical methods used, (ii) the range of statistical methods used, and (iii) the sample sizes used, over the past 25 years.

MATERIALS AND METHODS

Search methods used to identify original research publications

The search strategy utilised in this systematic review aimed to identify all publications that reported the estimation of age-specific paediatric RIs and included keyword search domains “estimation”, “reference interval” and “children” (Figure 1). Within each domain, search terms were either combined using "OR". The “reference intervals” and “estimation” domains were then merged within six words using “AND”, followed by a final merge of this search with “children” domain using “AND” (Figure 1). Original publications estimating age-specific paediatric RIs for laboratory blood tests, written in English, between January 1, 1993, and December 12, 2018, were deemed eligible for this systematic review. Using the search strategy and eligibility criteria defined in Table 1, electronic literature searches of MEDLINE and Embase databases were conducted on December 12, 2018. Full search details have been provided in Supplementary Document 1 of the Appendix D.

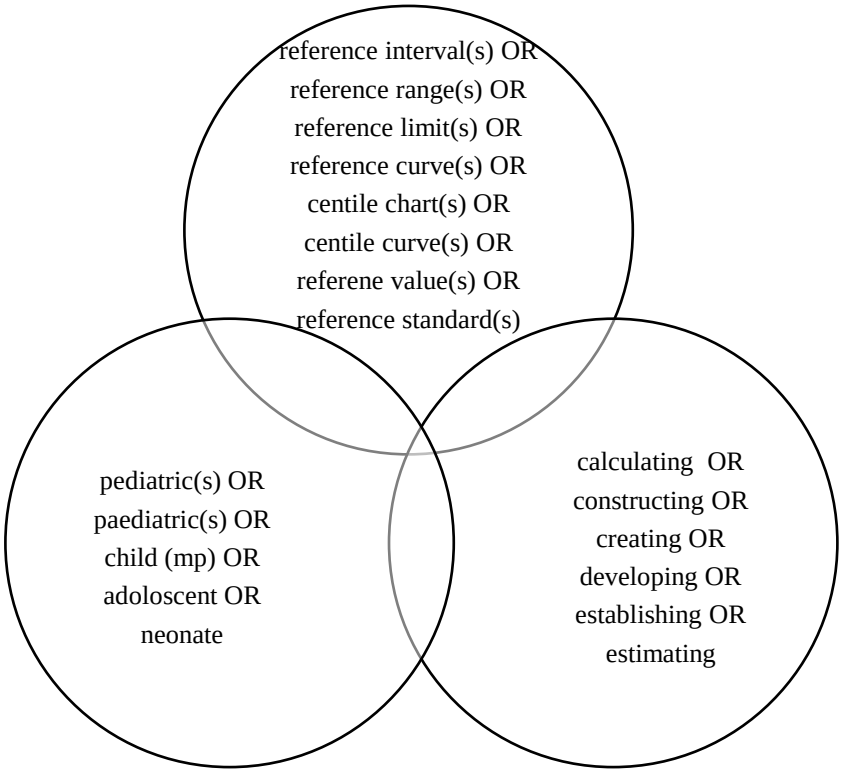


Figure 1: Diagram representing search strategy for the systematic review with keywords [3, 21] defining pre-identified search terms.

Table 2: Inclusion and exclusion criteria for publications in systematic review

Inclusion for Publication	**Exclusion for Publication**
Paediatric RIs estimated for a laboratory blood test	Subjects did not include paediatric population (0 – 18 years)
Written in English	Subjects were from non-human population
Published between January 1, 1993 and December 12, 2018	RIs were not established for blood biomarkers
	Change in RIs with age was not considered
	Did not estimate *de novo* RIs (i.e. validated existing RIs for use in another laboratory or transferred published RIs to another method or analyser)
	Statistical methods paper that estimated paediatric age-specific RIs
	Total samples used for establishing RIs were less than 100

Selection of publications

Resultant publication titles and abstracts were initially screened for duplicates by the primary author (MH) and then assessed against the defined inclusion and exclusion criteria by MH and a co-author (GK). If the eligibility of the publication was unable to be determined from the title and abstract, the full article was examined by MH and GK. An independent review of 10% of the titles and abstracts was performed by an additional co-author (LC) with all discrepancies in eligibility discussed and resolved by MH, GK and LC. The bibliography management, duplicate publication removal process and data storage was performed using EndNote X7.7 [22].

Data extraction

The data extraction questionnaire was developed in line with CLSI guidelines and previous systematic reviews of statistical methods used in the estimation of RIs [2, 3, 23]. The questionnaire was piloted using four articles and then reviewed by MH and LC. The questionnaire included information specific to: a) sample size, b) use of age as a categorical or continuous variable, and c) the statistical methods applied in different steps of analysis. MH reviewed all resultant publications (n=238), while 25% of the publications were independently reviewed by LC to ensure consistency in the review. The authors (MH and LC) discussed and resolved review outcome discrepancies in study design, sample selection, and evaluating the influence of sex on RIs for all publications prior to the final data being entered into a REDCap database [24].

Data analysis

The reporting of statistical methods by the eligible publications and differences in reporting, pre- and post- the release of the 3rd edition CLSI guidelines in 2010, were investigated [2].

Additionally, consideration of age as a discrete or a continuous variable, and its effect on the statistical method applied, were investigated. In summary, discrete age variables were further classified as age-groups of several days, weeks, months, or years e.g. 1 - 3 years, 3 - 6 years, 6 – 12 years etc. or as integer ages of consecutive days, weeks, months, or years e.g. 1, 2, 3, …, 7 days or 1, 2, 3…, 18 years. Compared to age-groups or integer ages with a limited number of categories, integer ages with a wide range of values e.g. 1, 2, 3…, 18 years were treated similarly to continuous age in fitting smooth curve. The number of samples used for the different statistical techniques was also examined. The extracted information was summarised using the frequency and percentage by RI type (discrete vs. continuous) and time period (pre vs. post 2010). Stata 15 and Microsoft Excel 2013 were used for these analyses [25].

RESULTS

Results of the search

The search results are presented diagrammatically in Figure 2. Briefly, from a total of 5274 publications, 2087 duplicates were removed, leaving 3187 unique publication titles and abstracts for screening. Of these, 2362 publications were excluded according to the defined criteria. The remaining 825 full text publications were assessed for eligibility and 557 were subsequently excluded. Whilst 268 publications were deemed eligible for the systematic review, 30 publications from 11 studies were also identified as having multiple publications using the same statistical methods and only one publication from each of these studies was retained [26-35]. The final number of publications systematically reviewed was 238. A summary of all publications reviewed has been provided in Supplementary Table 1 of the Appendix D.

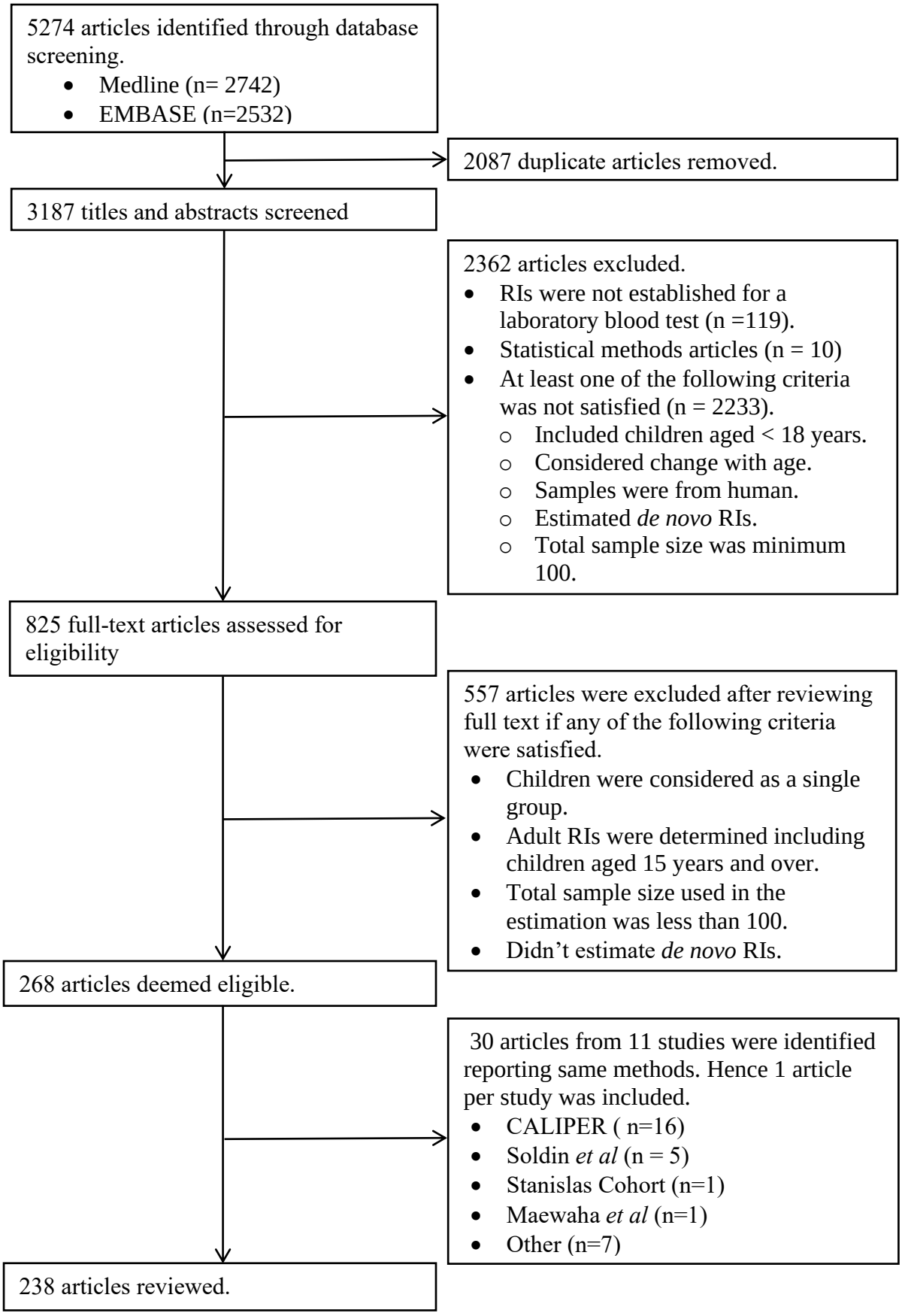


Figure 2: Flowchart outlining the systematic review article section process.

Design characteristics of the study

Among publications (78%, n = 185) that provided reference population description, 90% (n = 167) defined “healthy" using predefined inclusion and exclusion criteria. Few RIs were established for children with conditions such as preterm birth, obesity, kidney or heart disease (8%, n = 18). Among the studies (86%, n = 205) that included population sources, approximately half (42%, n = 87) recruited children from hospital inpatient or outpatient services, approximately two out of five studies (38%, n = 78) used existing data from various sources e.g. laboratory/hospital records and approximately one quarter (24%, n = 49) recruited children from communities or schools. Amongst the publications included in the systematic review, 93% (n = 221) were cross-sectional (using only one measurement per subject).

Discrete versus continuous age

The majority of publications reviewed (67%, n = 160) considered age a discrete variable by partitioning into several age-groups. In addition, approximately one fifth (21%, n=50) of the reviewed publications reported age as a discrete variable using integer ages. Age was treated as a continuous variable in only a small proportion of publications reviewed (17%, n = 40). Among the publications reviewed, a few (5%, n=12) analysed the data for both discrete and continuous age.

Influence of sex on RIs

Among the publications reviewed, one-third (33%, n = 78) did not report the influence of sex on the RIs. RIs were reported separately for male and female in 40% (n = 95) of the publication either based on the existing literatures (7%, n = 16) or did not provide any

justifications (33%, n = 79). One-fourth of the publications (27%, n =65) statistically evaluated the influence of sex on reference values and estimated RIs based on the findings.

Discrete versus continuous RIs

Among the publications reviewed, 167 (70%) reported discrete age-group RIs, 54 (23%) reported continuous RIs and 17 (7%) reported both discrete age-group and continuous RIs. The proportion of publications reporting discrete age-group RIs increased from 2010. In contrast, the proportion of publications reporting continuous RIs plateaued from around 2010 and appears to have decreased since 2014 (Figure 3).

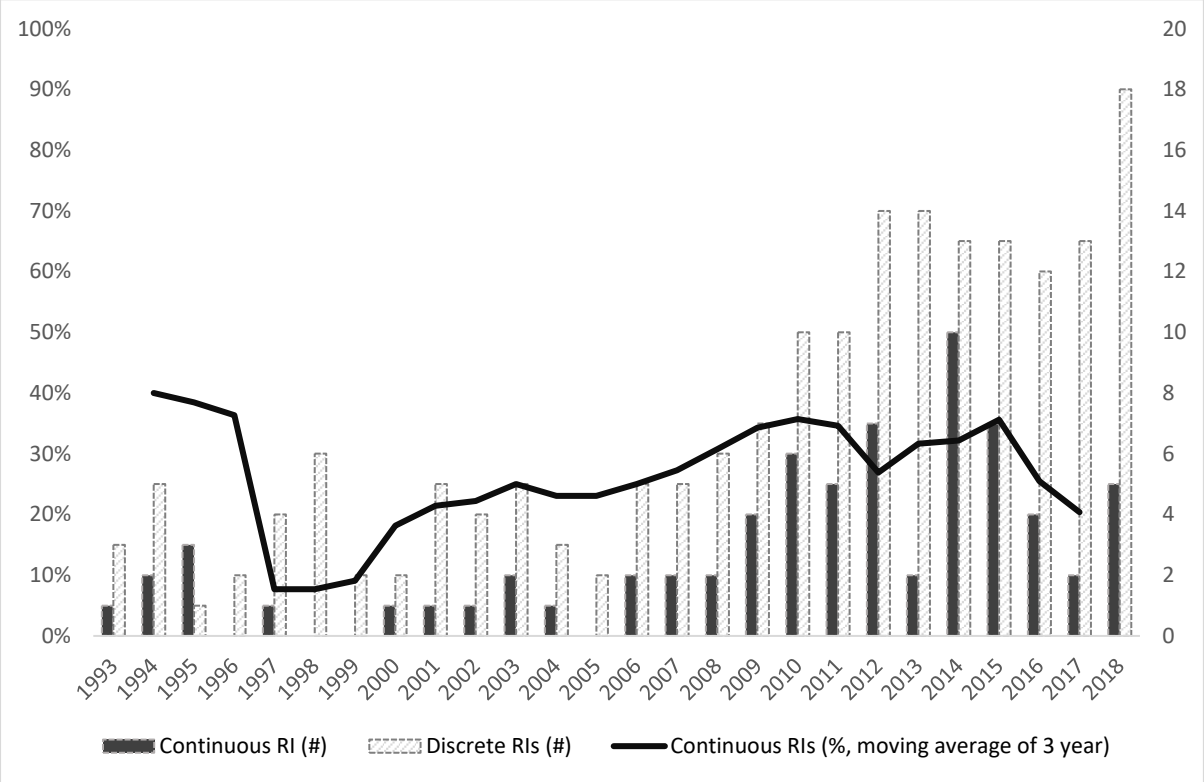


Figure 3: Number and proportion (as %) of publications reporting discrete age-group or continuous-agea paediatric RIs for a laboratory blood test between 1993 and 2018.

a. Publications reporting both discrete age-group and continuous RIs were counted twice as part of discrete age-group and continuous RIs.

The completeness of reporting of statistical methods used

Among the publications reviewed (n=238), 99 (42%) were published in or before 2010 and 139 (58%) were published after 2010. Approximately one third (32%, n = 77) of the publications mentioned the assessment of outliers. The proportion of publications reporting treatment of outliers improved to 40% (n=55) post-2010 from 22% (n = 22) in or before 2010. Half (50%, n = 118) of the publications reviewed reported some assessment of the normality of the distribution. The proportion of publications reporting assessment of data distribution was similar between pre-2010 (52%, n=51) and post-2010 (48%, 67). In terms of the statistical methods used for the estimation of RIs, 12 (5%) publications did not provide adequate details to facilitate study replication, of which 5 (5%) were published pre-2010 and 7 (5%) were published post-2010. More than half (56%, n = 133) of the publications did not provide any reference to guidelines, publications on statistical methods, or justification of the methods used to estimate RIs. However, there was an increase in publications that provided a reference to statistical methods used from 34% (n=34) pre-2010 to 51% (n=70) post-2010. Only 36 (15 %) studies reported CIs for the limits of the RIs, of which 32 (89%) were published post-2010. Similarly, the proportion of papers that reported the statistical software used for the analysis increased from 59% (n=58) in or before 2010 to 91% (n=127) post-2010. A detailed breakdown of the completeness of reporting of statistical methods by RIs type is provided in Table 2.

Table 2: Detailed breakdown of the completeness of statistical methods reported for different steps in estimation of RIs by types of RIs

	Discrete RI[a] n, % (n=184)	**Continuous RI[a] n, % (n=71)**
1. Outlier exclusion	64 (34.8 %)	18 (25.4 %)
Dixon/Reed	23 (35.9 %)	2 (11. 1 %)
Tukey's fence	17 (26.6 %)	3 (16.7 %)
Grubb's test	2 (3.1 %)	0 (0%)

SD method	3 (4.7 %)	1 (5.6 %)
Subjective elimination[b]	15 (23.4 %)	9 (50.0 %)
Others	11 (17.2 %)	4 (22.2 %)
2. Normality of data check	96 (52.5 %)	27 (38.6 %)
Graphical presentation	21 (21.6 %)	2 (7.4 %)
Formal testing	52 (53.6 %)	7 (25.9 %)
No method stated	32 (33.0 %)	18 (66.7 %)
3. Statistical method used		
Nonparametric (Ranks /Percentile)	117 (63.6 %)	-
Robust	10 (5.4 %)	-
Parametric (mean +/- SD)	62 (33.7 %)	-
Lambda-Mu-Sigma	-	20 (28.2 %)
Generalised Additive model for local, scale and shape	-	9 (12.7 %)
Regression based using parametric function of age	-	20 (28.2 %)
Quantile regression using parametric or nonparametric function of age	-	8 (11.3 %)
Not specified[c]	8 (4.2 %)	4 (5.4 %)
Others	25 (13.3 %)	10 (14.1 %)
4. Reported confidence interval	36 (19.8 %)	2 (7.1 %)[d]
5. Justification for statistical method provided	67 (28.0 %)	49 (69.0 %)
CLST/IFCC/ INFCC	42 (62.7 %)	3 (6.1 %)
Statistical methods publication	13 (19.4 %)	39 (79.6 %)
Other publication	12 (17.9 %)	7 (14.3 %)
6. Statistical software used	143 (77.7 %)	59 (83.1 %)
Excel (MicroSoft)	19 (13.3 %)	2 (3.4 %)
*Stata (StataCorp LP***)**	13 (9.1 %)	6 (10.2 %)
R (R Core Team)	15 (10.5 %)	21 (35.6 %)
GraphPad (GraphPad Software, Inc)	8 (5.6 %)	1 (1.7 %)
MedCal (MedCalc Software)	11 (7.7 %)	2 (3.4 %)
SPSS (IBM Corporation)	46 (32.2 %)	14 (23.7 %)
SAS (SAS Institute)	21 (14.7 %)	12 (20.3 %)
LMS ChartMaker (Harlow printing ltd)	-	8 (13.5 %)
Others	39 (27.3)	10 (16.9 %)

a. Some of the studies used more than one statistical methods and are therefore reported for more than one method category.

b. Removed outliers based on clinical or graphical examination.

c. Did not report enough details to classify the method.

d. Among the publications that applied regression based methods (n=28)

Statistical methods applied in the estimation of age-specific paediatric RIs

Among the studies that reported discrete age-group RIs, more than half chose statistical methods used for identifying and excluding outliers that were consistent with the CLSI guidelines, stating the Dixon-Reed (36%, n = 23) or Tukey's (27%, n= 17) methods. In contrast, of the studies that reported continuous RIs, nine eliminated outliers by informal definition as extreme values or clinically implausible. Only four studies considered the influence of age on the reference values, and excluded outliers based on the assessment of residuals [36-39].

For normality assessment, formal testing methods such as the Kolmogorov-Smirnov test, the Shapiro-Wilk test or D'Agostino-Pearson Omnibus tests were preferred (54%, n = 52), compared to visual inspection (22%, n = 21), among the publications that reported discrete age-group RIs. Similarly, formal testing methods were preferred (26%, n=7) among publications that reported continuous RIs. However, Takala *et al.* and Brabant *et al.* assessed the normal distribution graphically by plotting reference values against age [38, 40].

The nonparametric method was used most frequently (64%, n = 117) to estimate discrete age-group RIs. The "robust" statistical method was also used for small sample sizes as per the CLSI guidelines, which involves an iterative process of estimating median and median absolute deviation about the median until the change in consecutive iterations is minimal (5%, n= 10) [2]. Among other methods, the Hoffman method [27] was also applied.

There was a wide variety of statistical methods applied in the estimation of continuous RIs. The LMS method [12, 13] (28%, n = 20) was used back to 1993 and has been applied consistently since then (Figure 4). Within LMS, different statistical methods such as cubic

splines, fractional polynomials, or quantile regressions were used for smoothing the modelling parameters across age. Parametric regression-based methods were reported in 20 (28%) publications since 1993 (Figure 4). Among these, 14 publications applied the method proposed by Royston [1], that of fitting a linear or polynomial model in age for the mean values of the biomarker and estimating the 95% reference limits by subtracting or adding two SDs from the fitted mean values or similar methods. Six publications (8%) estimated discrete RIs for different age-groups of children and fitted linear or polynomial curves for the RIs [35, 41-45]. Since 2004, quantile regression using linear or polynomial models in age or spline curves [15, 16] have been used to estimate the RIs (11%, n=8, Figure 4). In recent years, the GAMLSS method [14] has gained popularity (13%, n = 9, Figure 4). Four studies (5%) that used nonparametric methods or regression-based methods did not provide enough information or only mentioned the statistical software used for the estimation of continuous RIs [33, 46-48].

Range of sample sizes used

The minimum sample size used for any of the age-groups ranged between 4 and 1,654 tests or individuals, with a median of 38.5 (IQR 17 – 92). The total sample sizes used for estimating continuous RIs ranged between 100 and 361,405, with a median of 1,156 (IQR 460, 3127). The median sample size for the parametric regression-based approach was 789 (IQR 394, 1485), for LMS or GAMLSS method was 1300 (IQR 539, 4218), and for quantile regression was 1,029 (IQR 549, 1802).

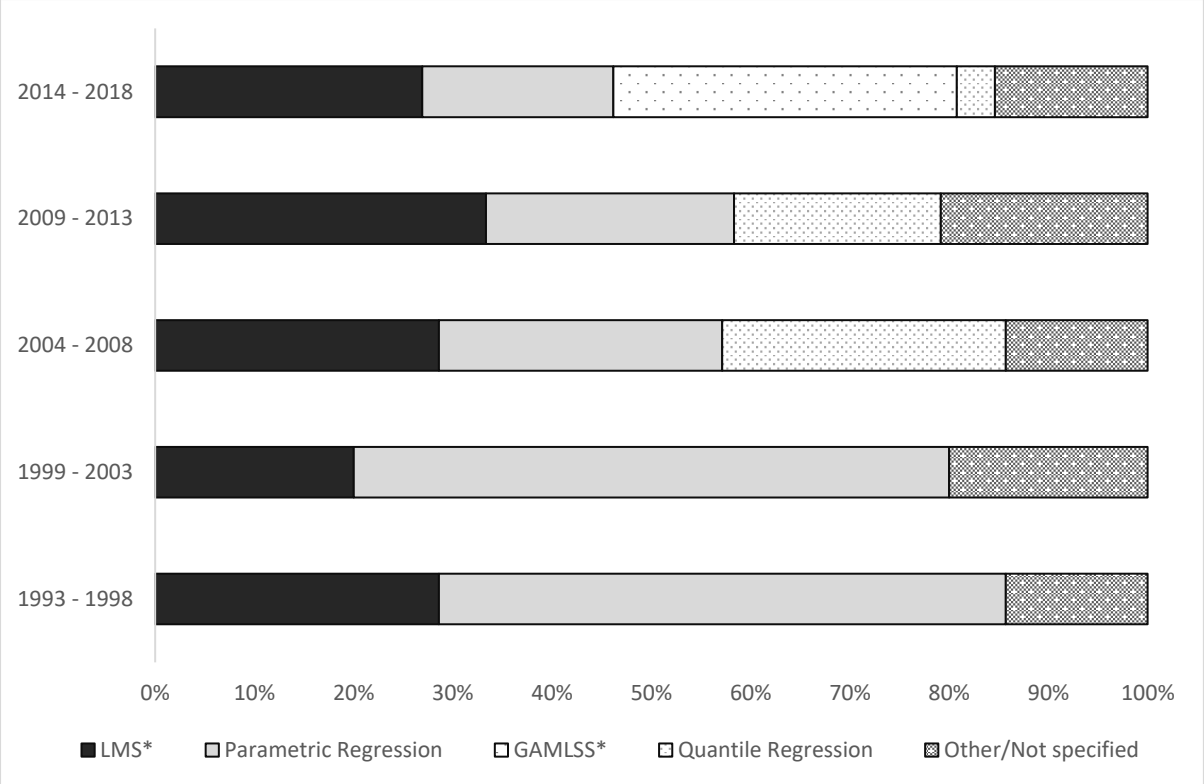


Figure 4: Statistical methods used in estimation of continuous paediatric RIs for a laboratory blood test between 1993 and 2018 (*where LMS stands for Lambda-mu-sigma and GAMLSS stands for Generalised additive model for location shape and scale)

DISCUSSION

This systematic review provides an overview of the statistical methods used to estimate age-specific paediatric RIs over the past 25 years. Overall, it was noted that most publications reviewed did not include full details of the statistical methods used in the various steps of analysis. However, there was an increase in publications reporting the statistical methods used in the estimation of age-specific RIs post 2010. Continuous RIs were reported from 1993 and have been established using several sophisticated curve-fitting approaches. The number (and the proportion) of publications reporting continuous RIs, however, has decreased in recent years. The sample sizes reported in the publications reviewed also varied across a wide range.

The statistical methods used in each step of the RI estimation process must be reported adequately to ensure reproducibility. At a minimum, publications should describe whether (and how) outliers were identified and removed, whether the distribution of the data was assessed and the statistical methods used for establishing RIs. In the absence of this information, it is difficult to determine whether statistical analysis of the data used to estimate RIs was appropriate and thus valid, for the reported RIs. Gaps in reporting of statistical methods, including outlier detection and CI for the RIs, were also highlighted by Daly *et al*. and Arseneau *et al*., and their findings were similar to this study [3, 23].

The Dixon and Reed method and Tukey method for outlier identification are currently recommended by the CLSI and have been commonly used [2]. Apparent outliers can be excluded from the overall data using these methods. However, for RIs estimated for biomarkers that are known to vary physiologically with age (e.g. creatinine, phosphate, etc.), outlier exclusion should be performed for each age-group or be based on the residuals defined as the differences between the observed and predicted values fitted using a model. Sensitivity analysis has also been proposed to assess the influence of apparent outliers [3].

Visual inspection is the most informative method to assess the distribution of data, as it provides a graphical representation of the distribution (e.g. normally distributed, skewed, bi-modal, etc) and aids in the identification of gaps in the data and potential outlying observations. In the case of assessing the normality assumption of the values with respect to age, histograms, quantile-quantile plots, scatter plots of the values against age, or visual inspection of the residual can be useful. However, this systematic review reports that formal testing methods were often preferred over visual inspection for assessing the normality of the data. Such tests lack statistical power to reject the null hypothesis of normality for small

sample sizes, while such methods will reject the null hypothesis for even a small deviation for large samples [11]. Hence, there is limited value in using these formal methods.

The nonparametric, parametric, and robust methods were the three commonly used statistical methods for estimating discrete age-group RIs and except for the parametric method are recommended in the CLSI guidelines [2]. Daly *et al*. recommended laboratories to use the parametric method for normally distributed or transformed data, based on a simulation study comparing these three methods under different scenarios and for different sample sizes [49]. The choice between these should be based on the distribution of the data and available sample sizes, not predefined.

There were several sophisticated curve-fitting methods applied in the estimation of continuous RIs. In all of these methods, the smoothness of the curve depends on the smoothing parameter i.e. the number of 'equivalent degree of freedom' (EDF), order of the polynomial or “knots” and “bandwidths” depending on the methods applied [1, 12, 14, 15]. There have been only a few comparative studies that compared the statistical methods for age-specific RIs [15, 50]. These comparative studies provide an understanding of how the methods perform for a given dataset but do not provide users with a broader perspective on how these methods will perform under different scenarios and for different sample sizes. For the latter, a simulation study comparing methods for estimating continuous RIs across a range of hypothetical scenarios would be valuable [51]. In addition, there is an urgent need for a consensus RI guidelines incorporating the recent development in statistical methods used to estimate continuous RIs.

While continuous RIs improve clinical interpretation for biomarkers that change with age, there are barriers to implementation. One of the key barriers highlighted in several publications is the limited capacity of existing laboratory information system (LIS) to visually display or incorporate equations to calculate continuous RIs [4, 8-10]. Preference in the traditional approach with test results of multiple analytes reported along with the age-group RIs in comparison to viewing multiple charts is another barrier in the implementation of continuous RIs. There is also a lack of clear guidelines in terms of statistical methods for estimating continuous RIs and expertise required in biostatistics. However, with several studies publishing continuous RIs [4, 8-10] and the ability of modern electronic medical recording (EMR) to display continuous RIs, the LIS and EMR developers have a role to play in improving practice in this area [52].

Identifying publications that reported *de novo* age-specific paediatric RIs is difficult as there is no validated search strategy or keyword for RIs. Addressing these issues, we implemented a broad search strategy similar to Daly *et al*., including additional search terms, “reference curve(s)”, “centile chart(s)” and “centile curve(s)”,specific to continuous RIs [3]. We also assessed entire publications that did not specify age-specific changes in the title or abstract, resulting in a comprehensive review.

We were interested in the range of statistical methods used in the estimation of age-specific RIs. Hence, we considered only one publication per study when the same statistical methods were applied in the estimation of RIs across different publications within the same broad project. This ensured that major studies with multiple publications that used the same statistical methods were weighted equally with studies with one publication using a unique statistical method.

CONCLUSIONS

This systematic review highlights that i) the statistical methods used in the estimation of RIs are frequently inadequately reported, and ii) the number of studies that reported continuous RIs has decreased in recent years. Despite the observed lack of consensus on the method to be used, it is important to establish continuous RIs for age-dependent biomarkers given the clinical implications – therefore, researchers should try to select the most appropriate method for their dataset, based on the current literature. For study reproducibility and regulation of RI quality, adequate reporting of statistical methods should be ensured by developing a checklist with a minimum requirement for reporting. This would enhance clinician and consumer confidence surrounding the laboratory assessment, validation and adoption of paediatric RIs in real clinical settings.

Acknowledgments

The authors acknowledge the support of Mr. Patrick Condron from the University of Melbourne in the library search and Mr. Luke Stevens from Murdoch Children's Research Institute (MCRI) in designing the online data collection form in REDCap. The authors also thank Dr. Sarah Arnup from MCRI for sharing her experience with a methodological systematic review.

Funding sources

Mr. Monsurul Hoq was supported by The Royal Children's Hospital Foundation [grant number 233, 2014] and a National Health Medical Research Council (NHMRC) Post Graduate Scholarship [grant number APP1168363, 2019].

Authors Contribution

M. Hoq conceived and designed the systematic review, conducted the initial database searches, screened and reviewed publications, conducted the qualitative and quantitative analysis, analysed and interpreted all the data and wrote the manuscript. G. Khanom and L. Canterford screened and reviewed publications, and reviewed the manuscript. S. Matthews, V. Ignjatovic and P. Monagle provided critical review and input to the manuscript. J. Carlin and S. Donath provided advice on the design of the study, analysis and interpretation of the data, and contributed in writing the manuscript. All authors read and approved the final manuscript.

Ethics and disseminating

No patients were involved in this study.

REFERENCE

[1] P. Royston, Constructing time-specific reference ranges, Statistics in Medicine 10(5) (1991) 675-690.

[2] CLSI, CLSI Releases Guidelines for Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory, COMTEX News Network, Inc., 2008.

[3] C.H. Daly, X. Liu, V.L. Grey, J.S. Hamid, A systematic review of statistical methods used in constructing pediatric reference intervals, Clin Biochem 46(13-14) (2013) 1220-7.

[4] T.P. Loh, G. Antoniou, P. Baghurst, M.P. Metz, Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals, Pathology 46(4) (2014) 336-43.

[5] E.M. Wright, P. Royston, Calculating reference intervals for laboratory measurements, Stat Methods Med Res 8(2) (1999) 93-112.

[6] V. Higgins, K. Adeli, Advances in pediatric reference intervals: from discrete to continuous, Journal of Laboratory and Precision Medicine 3(1) (2018).

[7] M. Hoq, V. Karlaftis, S. Matthews, J. Burgess, S. Donath, J. Carlin, P. Monagle, V. Ignjatovic, A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the settings of clinical biochemistry, immunology and haematology: The HAPPI Kids study, BMJ Open (2019).

[8] J. Zierk, F. Arzideh, T. Rechenauer, R. Haeckel, W. Rascher, M. Metzler, M. Rauh, Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence, Clin Chem 61(7) (2015) 964-73.

[9] M. Hoq, S. Matthews, V. Karlaftis, J. Burgess, J. Cowley, S. Donath, J. Carlin, T. Yen, V. Ignjatovic, P. Monagle, Reference Values for 30 Common Biochemistry Analytes Across 5 Different Analyzers in Neonates and Children 30 Days to 18 Years of Age, Clinical Chemistry (2019) clinchem.2019.306431.

[10] S. Asgari, V. Higgins, C. McCudden, K. Adeli, Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals, Clin Biochem 73 (2019) 82-89.

[11] B.W. Yap, C.H. Sim, Comparisons of various types of normality tests, Journal of Statistical Computation and Simulation 81(12) (2011) 2141-2155.

[12] T.J. Cole, Fitting Smoothed Centile Curves to Reference Data, Journal of the Royal Statistical Society. Series A (Statistics in Society) 151(3) (1988) 385-418.

[13] T.J. Cole, P.J. Green, Smoothing reference centile curves: The lms method and penalized likelihood, Statistics in Medicine 11(10) (1992) 1305-1319.

[14] R.A. Rigby, D.M. Stasinopoulos, Generalized additive models for location, scale and shape, Journal of the Royal Statistical Society: Series C (Applied Statistics) 54(3) (2005) 507-554.

[15] A. Gannoun, S. Girard, C. Guinot, J. Saracco, Reference curves based on non-parametric quantile regression, Stat Med 21(20) (2002) 3119-35.

[16] R. Koenker, Quantile Regression, Cambridge University Press, New York, 2005.

[17] F.P.B. Kroon, S. Ramiro, P. Royston, S. Le Cessie, F.R. Rosendaal, M. Kloppenburg, Reference curves for the Australian/Canadian Hand Osteoarthritis Index in the middle-aged Dutch population, Rheumatology (Oxford) 56(5) (2017) 745-752.

[18] Y. Wei, A. Pere, R. Koenker, X. He, Quantile regression methods for reference growth charts, Statistics in Medicine 25(8) (2006) 1369-1382.

[19] L.S. Chitty, D.G. Altman, Charts of fetal size: kidney and renal pelvis measurements, Prenat Diagn 23(11) (2003) 891-7.

[20] C.A. Bellera, J.A. Hanley, A method is presented to plan the required sample size when estimating regression-based reference limits, J Clin Epidemiol 60(6) (2007) 610-5.

[21] E. Arseneau, C.M. Balion, Statistical methods used in the calculation of geriatric reference intervals: a systematic review, Clin Chem Lab Med 54(3) (2016) 377-88.

[22] EndNote, EndNote X7.7.1 for Mac and Windows, Thomson Reuters, 2016.

[23] E. Arseneau, C.M. Balion, Statistical methods used in the calculation of geriatric reference intervals: a systematic review, Clinical chemistry and laboratory medicine 54(3) (2016) 377-388.

[24] P.A. Harris, R. Taylor, B.L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, S.N. Duda, The REDCap consortium: Building an international community of software platform partners, Journal of Biomedical Informatics 95 (2019) 103208.

[25] StataCorp, Stata Statistical Software: Release 15, StataCorp LLC, College Station, TX, 2017.

[26] K. Adeli, V. Higgins, K. Trajcevski, N. White-Al Habeeb, The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper, Critical Reviews in Clinical Laboratory Sciences 54(6) (2017) 358-413.

[27] O.P. Soldin, J.R. Dahlin, E.G. Gresham, J. King, S.J. Soldin, IMMULITE 2000 age and sex-specific reference intervals for alpha fetoprotein, homocysteine, insulin, insulin-like growth factor-1, insulin-like growth factor binding protein-3, C-peptide, immunoglobulin E and intact parathyroid hormone, Clinical Biochemistry 41(12) (2008) 937-42.

[28] H. Berrahmoune, J.V. Lamont, B. Herbeth, P.S. FitzGerald, S. Visvikis-Siest, Biological determinants of and reference values for plasma interleukin-8, monocyte chemoattractant protein-1, epidermal growth factor, and vascular endothelial growth factor: Results from the STANISLAS cohort, Clinical Chemistry 52(3) (2006) 504-10.

[29] R.K. Marwaha, R. Khadgawat, N. Tandon, R. Kanwar, A. Narang, A. Sastry, K. Bhadra, M. Kalaivani, Reference intervals of serum calcium, ionized calcium, phosphate and alkaline phosphatase in healthy Indian school children and adolescents, Clinical Biochemistry 43(15) (2010) 1216-9.

[30] U. Schauer, F. Stemberg, C.H. Rieger, M. Borte, S. Schubert, F. Riedel, U. Herz, H. Renz, M. Wick, W. Herzog, Establishment of age-dependent reference values for IgA subclasses, Clin Chim Acta 328(1-2) (2003) 129-33.

[31] M.H. Gannage-Yared, N. Balech, V. Farah, M. Antar, R. Saliba, E. Chahine, Pediatric TSH Reference Intervals and Prevalence of High Thyroid Antibodies in the Lebanese Population, International Journal of Endocrinology Print 2017 (2017) 6372964.

[32] A.M. Buchanan, F.J. Muro, J. Gratz, J.A. Crump, A.M. Musyoka, M.W. Sichangi, A.B. Morrissey, K. M'Rimberia J, B.N. Njau, L.J. Msuya, J.A. Bartlett, C.K. Cunningham, Establishment of haematological and immunological reference values for healthy Tanzanian

children in Kilimanjaro Region, Tropical Medicine & International Health 15(9) (2010) 1011-21.

[33] R.D. Christensen, J. Jensen, A. Maheshwari, E. Henry, Reference ranges for blood concentrations of eosinophils and monocytes during the neonatal period defined from over 63 000 records in a multihospital health-care system, Journal of Perinatology 30(8) (2010) 540-5.

[34] W. Tamimi, H. Tamim, N. Felimban, A. Almutair, Y. Altwaijri, I. Alalwan, Age- and gender-specific reference intervals for serum lipid levels (measured with an Advia 1650 analyzer) in school children, Pediatrics International 53(6) (2011) 814-9.

[35] M.W. Elmlinger, W. Kuhnel, H.G. Lambrecht, M.B. Ranke, Reference intervals from birth to adulthood for serum thyroxine (T4), triiodothyronine (T3), free T3, free T4, thyroxine binding globulin (TBG) and thyrotropin (TSH), Clin Chem Lab Med 39(10) (2001) 973-9.

[36] J. Okebe, J. Mwesigwa, S.C. Agbla, F. Sanya-Isijola, I. Abubakar, U. D'Alessandro, A. Jaye, K. Bojang, Seasonal variation in haematological and biochemical reference values for healthy young children in The Gambia, BMC Pediatrics 16 (2016) 5.

[37] A.A. Odutola, M.O. Afolabi, J. Jafali, I. Baldeh, O.A. Owolabi, P. Owiafe, G. Bah, B. Jaiteh, N.I. Mohammed, S.A. Donkor, J. Ndure, J. Adetifa, K.L. Flanagan, M.O.C. Ota, Haematological and biochemical reference values of Gambian infants, Tropical Medicine & International Health 19(3) (2014) 275-283.

[38] T.I. Takala, E. Makela, P. Suominen, J. Matomaki, H. Lapinleimu, L. Lehtonen, A. Rajamaki, K. Irjala, P.M. Lahteenmaki, Blood cell and iron status analytes of preterm and full-term infants from 20 weeks onwards during the first year of life, Clinical Chemistry & Laboratory Medicine 48(9) (2010) 1295-301.

[39] A. Heinze, M.C. Elze, S. Kloess, O. Ciocarlie, C. Konigs, S. Betz, M. Bremm, R. Esser, T. Klingebiel, M. Serban, J.L. Hutton, U. Koehl, Age-matched dendritic cell subpopulations reference values in childhood, Scandinavian Journal of Immunology 77(3) (2013) 213-20.

[40] G. Brabant, A. von zur Muhlen, C. Wuster, M.B. Ranke, J. Kratzsch, W. Kiess, J.M. Ketelslegers, L. Wilhelmsen, L. Hulthen, B. Saller, A. Mattsson, J. Wilde, R. Schemer, P. Kann, K.B. German, Serum insulin-like growth factor I reference values for an automated chemiluminescence immunoassay system: results from a multicenter study, Horm Res 60(2) (2003) 53-60.

[41] U. Lausten-Thomsen, M. Christiansen, C.E. Fonvig, C. Trier, O. Pedersen, T. Hansen, J.C. Holm, Reference values for serum total adiponectin in healthy non-obese children and adolescents, Clinica Chimica Acta 450 (2015) 11-4.

[42] G. Cangemi, N. Di Iorgi, S. Barco, G. Reggiardo, M. Maghnie, G. Melioli, Plasma total adiponectin levels in pediatrics: reference intervals calculated as a continuous variable of age, Clinical Biochemistry 45(18) (2012) 1703-5.

[43] A.K. Brown, L.A. Sleeper, S.T. Miller, C.H. Pegelow, F.M. Gill, M.A. Waclawiw, Reference values and hematologic changes from birth to 5 years in patients with sickle cell disease. Cooperative Study of Sickle Cell Disease, Arch Pediatr Adolesc Med 148(8) (1994) 796-804.

[44] M. Rauchenzauner, A. Schmid, P. Heinz-Erian, K. Kapelari, G. Falkensammer, A. Griesmacher, G. Finkenstedt, W. Hogler, Sex- and age-specific reference curves for serum markers of bone turnover in healthy children from 2 months to 18 years, J Clin Endocrinol Metab 92(2) (2007) 443-9.

[45] R.F. Ritchie, G.E. Palomaki, L.M. Neveux, T.B. Ledue, W.Y. Craig, S. Marcovina, O. Navolotskaia, Reference distributions for apolipoproteins AI and B and the apolipoprotein

B/AI ratios: A practical and clinically relevant approach in a large cohort, Journal of Clinical Laboratory Analysis 20(5) (2006) 209-217.

[46] R.D. Christensen, E. Henry, R.L. Andres, S.T. Bennett, Reference ranges for blood concentrations of nucleated red blood cells in neonates, Neonatology 99(4) (2011) 289-94.

[47] B.C. MacQueen, R.D. Christensen, E. Henry, A.M. Romrell, T.J. Pysher, S.T. Bennett, M.C. Sola-Visner, The immature platelet fraction: creating neonatal reference intervals and using these to categorize neonatal thrombocytopenias, Journal of Perinatology 37(7) (2017) 834-838.

[48] A. Haraldsson, C.M. Weemaes, S. Jonasdottir, O. Olafsson, G. Van De Wiel, J. Goertz, I. Klasen, Serum immunoglobulinD in infants and children, Scand J Immunol 51(4) (2000) 415-8.

[49] C.H. Daly, V. Higgins, K. Adeli, V.L. Grey, J.S. Hamid, Reference interval estimation: Methodological comparison using extensive simulations and empirical data, Clin Biochem 50(18) (2017) 1145-1158.

[50] E.M. Wright, P. Royston, A Comparison of Statistical Methods for Age-related Reference Intervals, Journal of the Royal Statistical Society: Series A (Statistics in Society) 160(1) (1997) 47-69.

[51] M. Hoq, P. Monagle, S. Donath, J. Carlin, Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study, in: H. Amanda (Ed.) Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference, Clinical Biochemist Review Supplement, Adelaide SA, 2019, p. S28.

[52] M. Hoq, S. Matthews, S. Donath, J. Carlin, V. Ignjatovic, P. Monagle, Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations, Clinical Biochemist Reviews 41(2) (2020) 43-52.

5.3 Discussion

Cole's LMS method was first proposed in 1988 and modified in 1992 [63, 72] . Royston's method was published in 1991[57]. Hence 1993 was selected as the starting year of the literature search since two statistical methods were available for estimating continuous RIs. Statistical methods used prior to 1993 for estimating continous RIs are likely to be no longer relevant given more recent methods.

The number of articles that have published continuous RIs has been declining since 2015. However, prior to 2015 there was an upward trend. This could be attributed to the improvement in statistical software that allows estimation of continuous RIs using different packages in R, Stata, MedCal, etc. In the past execpt for LMS there was no such software package available. In addition, the improvement in Electronic Medical Record, which can incorporate RIs curve, created demand among medical practitioners.

Cole's LMS method has been widely applied in the estimation of continuous age-specific RI [72]. The method is based on the Box-Cox transformation with three parameters representing the median, coefficient of variation and shape. The parameters are estimated for different age groups and smoothing using polynomials of age. The method was further improved by estimating three parameter using cubic splines removing the need to initial estimation of parameters by age-groups [54]. However, in both cases the degree of smoothness is controlled by the order of polynomials or number of 'equivalent degree of freedom' (edf). The GAMLSS, a modification of the LMS method, has also been used in the estimation of continuous age-specific reference intervals in recent years [56]. This method can be applied to highly skewed or kurtotic continuous response variable and to estimate other parameters of distribution of the response variable. However, similar to LMS methd, the smoothness of the

curve depends on the edf. Royston's parametric method has also been used where the polynomial regression was used to model the mean of normality distributed or log transformed response variable [73]. The reference limits were calculated using the standard deviation within age-groups for heteroscedasticity. One of the advantage of Royston's method is the z-score may be calculated based on normality assumption. However, not all analytes are normally distributed or can be transformed and the width of age-groups needs careful consideration in case of heteroscedasticity. Nonparametric quantile regression method has been used by few studies, which offers more flexibility in terms of estimating the reference curve. However, the smoothness depend on the choice of knots and bandwidths. The parametric quantile regression which fits a linear model that minimises the sum of the absolute residuals are used to estimate the RIs and have expressed the RIs as a function of age and/or sex.

The median and inter-quartile range of sample sizes reported in this review are useful in understanding the practical implications of recommending sample sizes required for estimating continuous age-specific RIs applying different statistical methods. However, the median sample sizes presented in our review are upper bounds as we have excluded publications with sample sizes less than 100.

Based on the search criteria used in two previously published systematic reviews [41, 74], we had updated our search strategy, which resulted in finding more publications that reported age-specific RIs. While those two studies explored statistical methods in general, our focus was on the statistical methods incorporating the choice between representing age as a categorical or continuous variable. We have also evaluated the applications of methods over time. Daly's systematic review only considered publications which reported paediatric RIs

from January 1, 2021 to May 18, 2012 [41]. The authors reported that only 1 in 22 publications reported paediatric continuous age-specific RIs [41]. In comparison we have reviewed publications that reported paediatric RIs from January 1, 1993 to December 12, 2018. We have reported that 1 in 3 publications reported paediatric continuous age-specific RIs for blood analytes among the publications reviewed. The extensive review allowed us to make recommendations for improving reporting and statistical methods used in different steps.

To ensure the reliability and validity of the estimated age-specific RIs, studies should report the statistical method applied in different steps of the statistical analysis [41, 71]. We recommend that publications at minimum should provide information about identification of outliers, assessment of the distribution of measurement, and the method of estimation of reference limits and corresponding 95% CIs [71]. To improve reporting of the statistical method used in the estimation process, the peak professional bodies (e.g. IFCC Reference Intervals and Decision Limits Committee (C-RIDL)) could introduce reporting guidelines for studies that estimate RIs [37].

The current guidelines of CLSI and IFCC also do not provide any guidance in terms of statistical methods which should be used for estimating continuous age-specific RIs [16, 37]. The poor uptake of continuous age-specific RIs could also be explained by the recommendation of CLSI and IFCC for age-partitioning for analytes that changes with age [16, 37]. Hence, we recommend that these guidelines should be updated and information for using modern statistical methods for estimating age-specific RIs should be incorporated.

Chapter 6: Comparison of statistical methods for constructing continuous paediatric reference intervals: a simulation study

6.1 Introduction

In chapter 3, I reported continuous age-specific RIs calculated by applying a novel statistical method using 360 samples for 30 biochemistry analytes, with varying complexity of association with age [66]. In chapter 4, I have demonstrated that these RIs are valid for use in a routine laboratory and proposed a modification of the existing method for validating continuous age-specific RIs [70]. In chapter 5, I have reported that there are a wide range of statistical methods available for estimating continuous age-specific RIs [71]. However, how our proposed novel method would perform in comparison to other commonly used statistical methods is unknown. The reliability of the continuous age-specific RIs, especially for complex shapes like ALP reported in chapter 3, is also of concern as we only used 360 samples in comparison to the median sample size of 1156 used for estimating continuous age-specific RIs found in our systematic review [71].

The generally accepted requirement of a statistical method is that it should obtain the "true" results "on average" when used repeatedly for samples taken from a population. However, defining true values and repeated sampling are challenging in real life unless we have a very large sample. But we can examine by simulation, where we set up a data-generating model (with random components), use this to create repeated "versions" of hypothetical data and apply the methods to each simulated dataset, comparing the results (between methods) with respect to how close they are on average to the true values implied by the model and how much variance they display across the repeated samples. Using this structured approach, I developed a simulation study evaluating and comparing statistical methods for estimating

continuous age-specific paediatric RIs under different scenarios and for different sample sizes.

This chapter reports the methodology of the simulation study, highlights the performance of our proposed novel method, in comparison to other statistical methods for estimating continuous age-specific RIs, and compares the precision for different sample sizes [75]. The manuscript was submitted to BMC Research Methodology in January, 2021. The author accepted submitted version is presented as the body of this chapter.

6.2 Publication

Comparison of statistical methods for estimating continuous paediatric reference intervals: a simulation study

Monsurul Hoq*[1, 2], Susan Donath[1, 2], Paul Monagle[1, 3], John B. Carlin[1, 2]

[1]Department of Paediatrics, The University of Melbourne, Parkville, VIC 3052, Australia.

[2]Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia

[3]The Royal Children's Hospital, Parkville, VIC 3052, Australia

Correspondence

Monsurul Hoq,

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute,

48 Flemington Road, Parkville, Victoria - 3052, Australia

Email: monsurul.hoq@mcri.edu.au, Phone: + 61 (03) 99366264

Present Address

48 Flemington Road, Parkville, Victoria - 3052, Australia

ABSTRACT

Background: Reference intervals (RIs), which are used as an assessment tool in laboratory medicine, change with age for most biomarkers in children. Addressing this, RIs that vary continuously with age have been developed using a range of curve-fitting approaches. The choice of statistical method may be important as different methods may produce substantially different RIs. Hence, we developed a simulation study to investigate the performance of statistical methods for estimating continuous paediatric RIs.

Methods: We compared four methods for estimating age-varying RIs. These were Cole's LMS, the Generalised Additive Model for Location Scale and Shape (GAMLSS), Royston's method based on fractional polynomials and exponential transformation, and a new method applying quantile regression using power variables in age selected by fractional polynomial regression for the mean. Data were generated using hypothetical true curves based on five biomarkers with varying complexity of association with age, i.e. linear or nonlinear, constant or nonconstant variation across age, and for four sample sizes (100, 200, 400 and 1000). Root mean square error (RMSE) was used as the primary performance measure for comparison.

Results: Regression-based parametric methods performed better in most scenarios. Royston's and the new method performed consistently well in all scenarios for sample sizes of at least 400, while the new method had the smallest average RMSE in scenarios with nonconstant variation across age.

Conclusions: We recommend methods based on flexible parametric models for estimating continuous paediatric RIs, irrespective of the complexity of the association between biomarkers and age, for at least 400 samples.

Keywords: Biomarker, Continuous, Reference intervals, Simulations, Statistics

BACKGROUND

In laboratory medicine, a reference interval (RI) is an important clinical assessment tool, used for identifying patients who may need further clinical investigation. Statistically, a RI is defined by a range between the $\alpha/2$th and $(100 - \alpha/2)$th percentiles where $0 < \alpha < 100$, within which $(100 - \alpha)\%$ of the population values should fall. The definition may be straightforward, but estimating RIs is a complex process which requires appropriately selected samples from a well-defined reference population, collection and testing of blood samples and lastly, appropriate statistical methods for estimation [1]. Although all the steps are important, the issues associated with the selection of appropriate statistical methods are often ignored, and these are complex for biomarkers that change with age especially in children [2].

Traditionally, for biomarkers that vary with age, paediatric RIs have been estimated separately for different subgroups of age. However, this approach is considered misleading as children whose ages are close to the age cut-off values may be misdiagnosed [3,4]. Addressing this, several studies have reported paediatric RIs that vary with continuous age, commonly known as “continuous RIs” [5-8].

There are several statistical methods available for estimating continuous RIs. Cole’s lambda-mu-sigma (LMS) method for fitting smooth centile curves is the most commonly used approach, based on the power transformation of Box and Cox, assuming the age-specific distribution is normal or may be transformed to normality [9]. Royston’s (1998) parametric method, which has also been applied in several studies, is based on fitting a fractional

polynomial model in age for each parameter of a normal, exponential normal (EN) or modulus-exponential normal (MEN) density, with estimation by maximum likelihood [10]. These parameters are then used to transform the data towards normality and centiles are calculated. Quantile regression has also been used in estimating continuous RIs, predicting a conditional quantile of the dependent variable without assuming any parametric conditional distribution [11]. Finally, in recent years, the Generalised Additive Model for Location Shape and Scale (GAMLSS) method has been used increasingly [12]. This method is a modification of the LMS method and can be applied for any distribution of the values and continuous age [12]. A few modifications of the Royston method have also been used in the estimation of continuous RIs [7,13,14].

The choice of statistical method is important as there may be differences in RIs due to the underlying assumptions. Studies investigating the performance of the different statistical methods in estimating continuous RIs using real datasets are unable to compare the various estimated RIs against the "true" RIs [11,15]. Li *et al* compared four methods which require age partitioning, but they did not consider regression-based parametric methods such as Royston's method [16]. On the other hand, only one simulation study compared the performance between two methods involving a nonparametric method and Royston's method [13]. There is also no recommendation for minimum sample sizes required for different statistical methods except for Royston's and Griffith *et al*'s method [13,17].

In summary, there are several sophisticated curve-fitting approaches available for estimating continuous RIs. However, there is no comparison of the statistical methods across different scenarios to provide a clear understanding of the way in which sampling variation may affect the capacity of each method to reproduce the true underlying pattern of variation with age.

Guidelines for minimum samples sizes required are also needed. Hence, we developed a simulation study to investigate the performance of four statistical methods used in estimating continuous RIs for a range of (true) relationships with age and varying sample sizes. In addition, we provided an illustration of differences between the RIs obtained by the four methods for two biomarkers using data from the HAPPI Kids study [18], for two different sample sizes.

METHODS

We performed a simulation experiment comparing four statistical methods for estimating continuous paediatric RIs. We examined how these methods performed under five scenarios chosen to provide examples of biomarkers that displayed increasing complexity in their age dependence e.g. potassium, creatinine, total protein, alkaline phosphatase (ALP) and phosphate (Figure 1) and four sample sizes (n = 100, 200, 400, and 1000) for each scenario. The details of the simulation design are provided below, following reporting guidelines proposed by Morris *et al* [19].

Data generation

For each replication (n=1000), we first generated two covariates i.e. age and sex. The age variable was generated between 1 month and 18 years based on a uniform distribution, while half of the samples were assigned as male and the other half as female. The reference values of the biomarkers were generated using the models presented in Table 1 based on the relationships with age that were observed in the HAPPI Kids study: see Figure 1 [7].

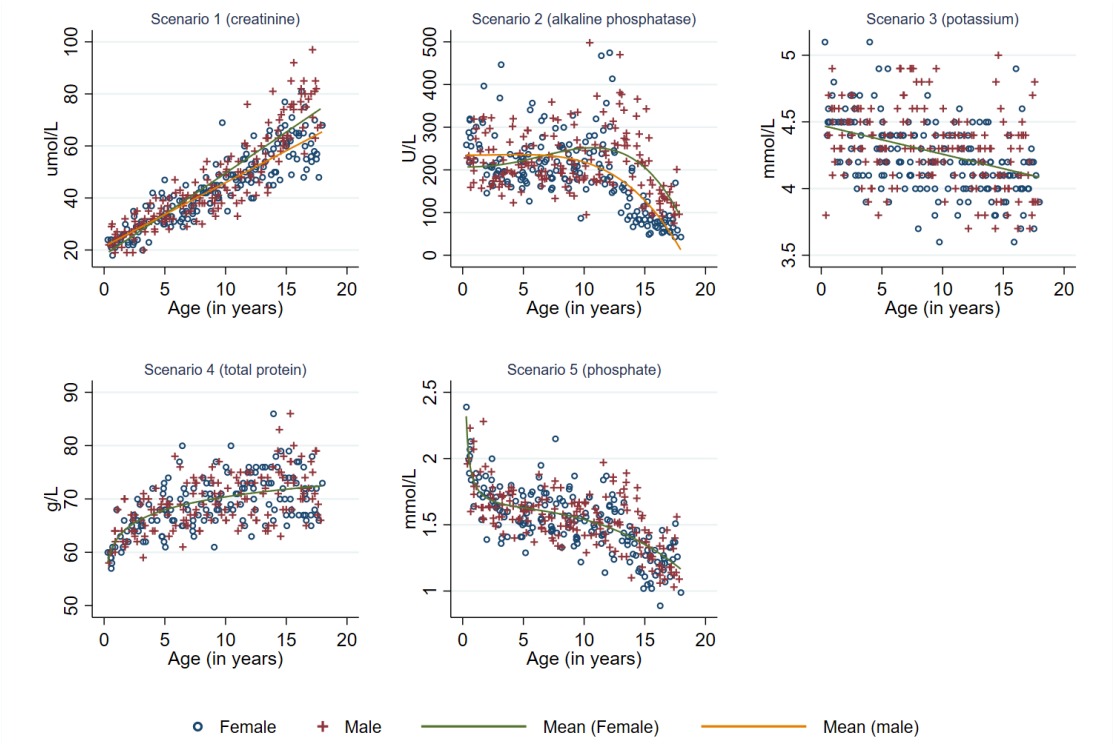


Figure 1: Distribution of measurement across age (in years) for five biomarkers from the HAPPI Kids Study

Target of analysis

The parameters of interest were the 2.5th and 97.5th percentiles at each integer age from 1 to 18 years and by sex. The true 2.5th and 97.5th percentiles denoted by $\theta_{jk}^{(\alpha)}$(where α =2.5 or 97.5) at each age j for sex k were calculated using $\mu_{jk} - 1.96 * \sigma_{jk}$ and $\mu_{jk} + 1.96 * \sigma_{jk}$ respectively, where μ_{jk} is the mean and σ_{jk} is the standard deviation (SD), using the parametric data-generating equations in Table 1.

Table 1: Mean response and standard deviation, as functions of age and sex, for the simulated scenarios

Scenario	Relationship with age and sex	Mean response	Standard deviation
1	Linear in age (separately by sex), SD increasing linearly with age e.g. creatinine	$21.50 + 2.50 \times age - 3.00 \times sex + 0.75 \times age \times sex$	$2.50 + 0.50 \times age$
2	Nonlinear in age (separately by sex), SD increasing nonlinearly with age e.g. alkaline phosphatase	$233.200 + 0.082 \times age^3 - 0.042 \times age^3 \times \ln(age)$ if sex is male	$54.0 + 0.0713 \times age^3 - 0.0268 \times age^3 \times \ln(age)$
		$209.064 + 0.301 \times age^3 - 0.112 \times age^3 * \ln(age)$ if sex is female	$45.0 - 0.167 \times age^3 - 0.0586 \times age^3 \times \ln(age))$
3	Linear in age, constant variance, no sex difference e.g. potassium	$4.50 - 0.01 \times age + 0.10 \times sex$	0.3
4	Nonlinear in age, SD increasing linearly with age, sex difference e.g. total protein	$62.15 - 0.01 \times age + 3.50 \times \ln(age) + 0.10 \times sex$	$3.00 + 0.15 \times age - 0.2 \times sex$
5	Nonlinear in age, constant variance, no sex difference e.g. phosphate	$1.28 + 0.53 \times age - 0.5 + 0.01 \times age^2 - 0.0035 \times age^2 \times \ln(age)$	0.15

Statistical methods for estimating continuous RIs

The statistical methods to be compared were selected based on a systematic review of research reporting estimates of age-specific paediatric RIs [20]. The review found that Cole's LMS, Royston's method, GAMLSS and quantile regression were commonly used methods [12'17,20,21] A brief summary of these four statistical methods for estimation of continuous RIs is provided below.

Royston's method [10]

This method is described by Royston and Wright [10]. Initially, the best fitting model for the mean and SD values are identified from 44 different models fitted using the fractional powers of age from the set (-2, -1, -0.5, 0, 0, 1, 2, 3), following the model selection procedure described by Royston and Altman [22]. In summary, the best fitting model for the mean values is chosen to use powers of age for a given degree of fractional polynomial model according to the criterion of identifying the model with lowest deviance. The best fitting model is selected based on the evidence of a difference in deviance between higher and lower degree fractional polynomial models [23]. A generalised least squares method is used for identifying the best fractional polynomial models of mean and SD, due to its availability in most statistical software. Later, using the best fitting powers of mean and SD, normal, EN (with a constant value for skewness) and MEN (with constant values for skewness and kurtosis) models are fitted by maximum likelihood method. The final model between normal, EN and MEN is selected by comparing deviance between these models.

As an extension of the Royston approach, in order to evaluate the influence of sex on the RIs in our context, the difference in mean values by sex and the interaction between age and sex are evaluated for the selected model. If the regression coefficient for sex reaches conventional

statistical significance (i.e. $p < 0.05$), sex is included in the final model, otherwise it is excluded. If the difference in deviance between the model with and without an age-by-sex interaction is significant, the data are analysed separately by sex.

Hoq *et* al's method [7]

We have used this method to estimate continuous paediatric RIs for 30 biomarkers [7]. Initially, the best fitting fractional polynomial model for the mean values of the biomarker is identified following the model selection procedure described as part of Royston's method [10]. Subsequently, the influence of sex on the RIs is evaluated applying the procedure described above. However, instead of analysing the data separately by sex, the interaction between sex and the fractional polynomial representation of age is included in the final model when the age-by-sex interaction is significant. Finally, quantile regression is used to estimate RIs representing the quantiles of the biomarker, constrained to follow the same power variables of age as identified (with sex interaction where indicated).

Cole's lambda, mu and sigma (LMS) method [9]

For this method, measurements are divided into each year of life from birth to 18 years. For each year, the parameters of the Box-Cox transformation i.e. the median, coefficient of variation and skewness of a measurement are estimated. These estimates are then smoothed by using cubic splines. Centiles of the transformed smoothed values are calculated and back-transformed to the original scale.

In this simulation study, we have used generalised additive models (GAM) to obtain predicted values of the parameters of the Box-Cox transformation due to their flexibility for modelling complex nonlinear relations, similar to the approach proposed by Cole and Green

[9, 14]. The smoothness of the curve depends on the number of equivalent degrees of freedom (EDF) selected for each of the three parameters. Hence, to decide on the optimal smoothing parameter, we apply a statistical approach known as cross-validation to minimise error in predicting the new data for the selected EDF. As part of the cross-validation, we omit data for each year of life once and estimate the average error in predicting the mean measurement of the omitted group. This process is repeated for EDF from 2 to 6, and the value producing minimum error in predicting the mean measurement is considered for predicting the three parameters for each year of life. The RIs were estimated separately for each sex.

Generalised additive models for location scale and shape (GAMLSS) [12]

The GAMLSS is a modification of Cole's LMS method where age is considered a continuous variable and curve fitting is controlled by the values of four parameters. This method applies to a wide variety of distributional forms. To estimate continuous RIs, we have used cubic spline functions to fit the age trend for each parameter. The cubic spline is chosen because the main feature of the method is to allow greater flexibility than polynomials or fractional polynomials for modelling complex nonlinear relations [24]. Since the simulated datasets were generated using a normal distribution, we have used Box-Cox t distribution as the data distribution [25]. The smoothing of the cubic spline function depends on the EDF and optimum EDF is identified based on minimising the generalised Akaike information criterion (GAIC), $-2l(\widehat{\theta}) + (\xi . df)$, where $l(\widehat{\theta})$ is the log likelihood function and ξ is the penalty for each EDF [26]. We have used the Schwarz Bayesian information criterion (SBC) where $\xi = \log n$ [12]. The influence of sex on the RIs is assessed by backward stepwise search among models with or without sex and with interaction between age and sex, with the preferred model based on the difference in GAIC [26].

Performance measures

The performance of each method was evaluated based on the average root mean squared error (RMSE) over age and sex i.e. $\frac{\sum_{k=0}^{1}\sum_{j=1}^{18}\sqrt{\mathrm{E}\left(\hat{\theta}_{jk}^{(\alpha)}-\theta_{jk}^{(\alpha)}\right)^{2}}}{(18*2)}$ where the true parameter is denoted $\theta_{jk}^{(\alpha)}$, its estimate $\hat{\theta}_{jk}^{(\alpha)}$, j is the integer age from 1 to 18, k = 0 or 1 stands for male or female respectively, α indexes the 2.5th and 97.5th percentiles and $\mathrm{E}\left(\hat{\theta}_{jk}^{(\alpha)}-\theta_{jk}^{(\alpha)}\right)^{2}$ is estimated by averaging the squared difference between estimated and true values across 1000 simulated datasets. Coverage and confidence interval (CI) width were also compared between Hoq *et al* and Royston's method, for which 95% CIs for the RI limits were available.

Software packages

Stata 15.1 was used for data generation and analysis of data [27]. The Stata commands –mfp- [23], -mfpi- [28] and –qreg- were used for Hoq *et al*'s method, and –fp-,-lrtest-,-xrigls- [10] and –xriml- [10] were used for Royston's method. To estimate continuous RIs using Cole's LMS method, we modified the –colelms- [29] package to incorporate cross-validation. Except for –mfp- and –qreg-, all the other commands are user-written packages. For the GAMLSS method, we used the –gamlss- package in R [26]. The Stata and R source code to reproduce the results is available in the supplementary document 1 A-D of the Appendix E.

RESULTS

Results from the simulaton study

The average RMSE across age and sex for the four methods in the five scenarios, each with four sample sizes, is presented in Table 2**.** The average estimated and true RIs for five scenarios are presented in the Supplementary Figure 1-5 of the Appendix E.

Table 2: Average RMSE for each of four methods for estimating RIs across integer age and sex, for five scenarios and four sample sizes. Bolded figures indicate minimum values across methods.

Performance Measure			**Average RMSE**							
Limit			**Lower limit**				**Upper limit**			
Sample size			**100**	**200**	**400**	**1000**	**100**	**200**	**400**	**1000**
Scenario	**Range**	**Method**								
1[a]	20 - 100	**Hoq *et al***	3.87	2.74	1.90	**1.16**	3.80	**2.65**	**1.80**	**1.17**
		Royston	2.89	**2.07**	**1.59**	1.29	**3.64**	2.88	2.43	2.17
		Cole's LMS	5.82	4.53	2.43	1.27	7.80	5.97	2.48	1.60
		GAMLSS	**2.84**	2.19	1.78	1.45	4.00	3.13	2.59	2.23
2[b]	0 - 500	Hoq *et al*	30.91	23.54	16.80	10.88	51.30	29.96	**18.77**	**12.19**
		Royston	26.39	21.67	14.93	8.75	40.77	28.48	20.17	13.90
		Cole's LMS	49.95	39.56	25.67	15.16	100.48	76.59	42.54	30.65
		GAMLSS	**24.05**	**16.32**	**11.50**	**8.04**	**32.69**	**24.86**	20.32	17.23
3[c]	3.5 – 5	Hoq *et al*	0.12	0.08	0.06	0.04	0.12	0.08	0.06	0.04
		Royston	**0.07**	**0.05**	**0.04**	**0.02**	**0.07**	**0.05**	**0.04**	**0.02**
		Cole's LMS	0.37	0.30	0.16	0.07	0.36	0.31	0.15	0.06
		GAMLSS[6]	0.11	0.07	0.05	0.03	0.11	0.07	0.05	0.03
4[d]	50 – 90	Hoq *et al*	1.78	1.27	0.98	0.78	1.83	1.31	1.03	0.86
		Royston	**1.60**	**1.04**	**0.73**	**0.58**	**1.61**	**1.15**	**0.87**	**0.71**
		Cole's LMS	5.42	4.53	2.15	1.07	4.88	4.11	1.90	0.85
		GAMLSS	1.92	1.44	1.07	0.78	1.96	1.44	1.07	0.80
5[e]	1 – 2.5	**Hoq *et al***	**0.07**	0.05	0.04	**0.04**	0.08	**0.06**	**0.05**	**0.04**
		Royston	0.07	0.06	0.05	0.04	**0.07**	0.06	0.05	0.05
		Cole's LMS	0.18	0.15	0.09	0.05	0.20	0.19	0.11	0.08
		GAMLSS	0.07	**0.05**	**0.04**	0.04	0.10	0.08	0.08	0.08

a. Linear in age (separately by sex), SD increasing linearly with age e.g. Creatinine. Values < 0 or > 150 were excluded.
b. Nonlinear in age (separately by sex), SD increasing nonlinearly with age e.g. Alkaline Phosphatase. Values < 0 or > 100 were excluded.
c. Linear in age, constant variance, no sex difference e.g. Potassium. Values < 3 or > 7 were excluded
d. Nonlinear in age, SD increasing linearly with age, sex difference e.g. Total Protein. Values < 0 or > 100 were excluded.
e. Nonlinear in age, constant variance, no sex difference e.g. Phosphate. Values < 0 were excluded.

The regression-based parametric methods, i.e. Hoq *et al* and Royston's methods, estimated the continuous RIs with lowest average RMSE for all scenarios except scenario 2. For scenario 1, the better performing of the two regression-based parametric methods varied between the lower and upper limits. For scenario 2, GAMLSS performed better for the lower limit, while Hoq *et al*'s method performed better for the upper limit for larger sample sizes. For scenarios 3 and 4, Royston's method estimated the continuous RIs with lowest average RMSE for all sample sizes. For scenario 5, all methods except Cole's LMS performed very similarly for the lower limit, while regression-based parametric methods had the lowest average RMSE for the upper limit.

The RMSEs by age, sex, scenario, sample size and method are presented in Figures 2 - 3 and in the Supplementary Figure 6 – 8 of the Appendix E. As shown in these figures, the RMSE varied by age with the lowest values for ages in the middle of the range. However, as the sample size increased the variation in RMSEs by age reduced. The differences in RMSE by sex were negligible.

The average coverage of estimated CIs for the RI limits and average CI width across integer age and sex for Hoq *et al* and Royston's method are presented in Supplemental Table 1 and 2 of the Appendix E respectively. For both methods the coverage was poor except for scenario 3. CI widths were narrower for Royston's method compared to Hoq *et al*'s.

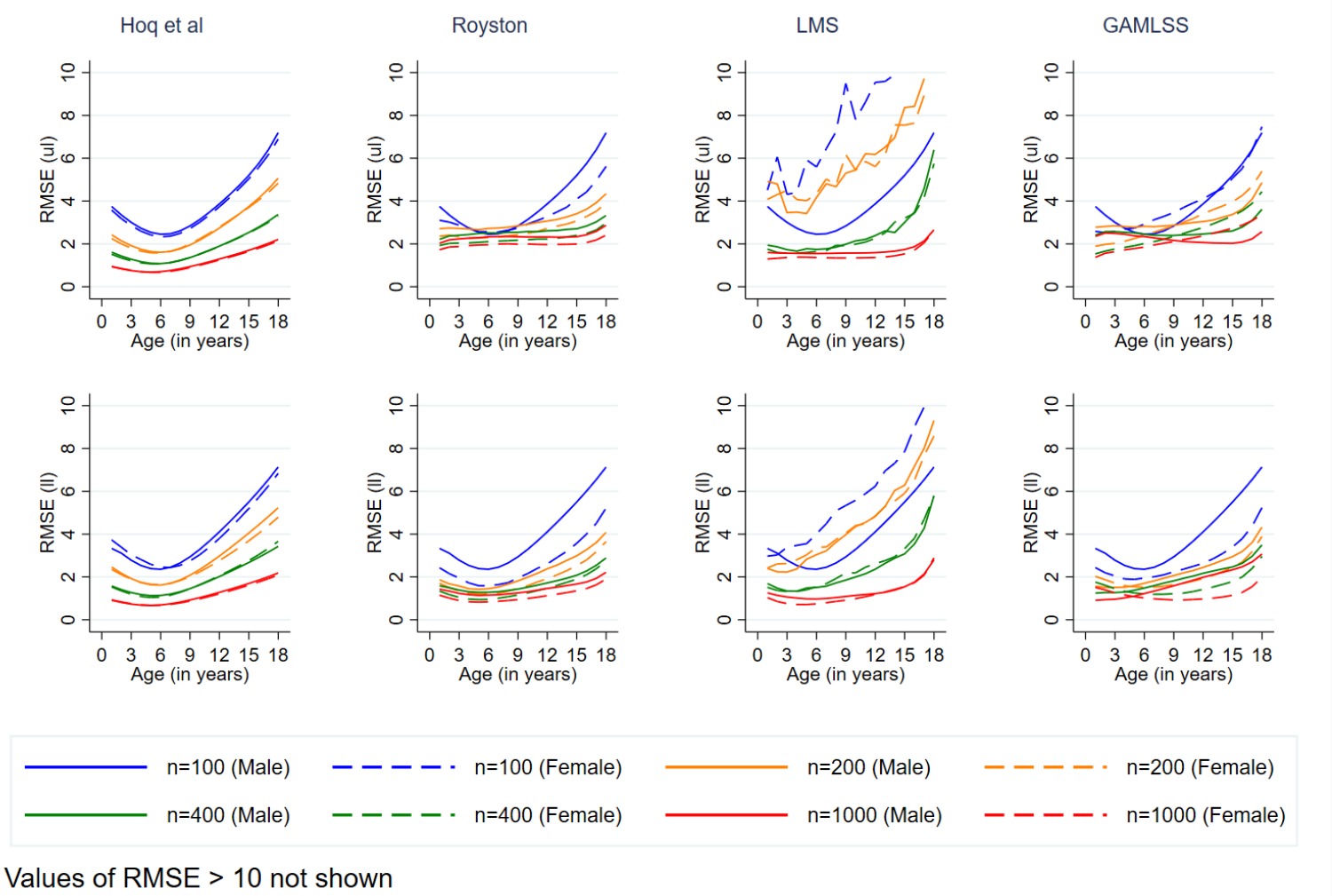


Figure 2: Root Mean Square Error (RMSE) by Age, sex, sample size and method for scenario 1. The first row represents RMSE of the upper limit and second row represents RMSE of the lower limit.

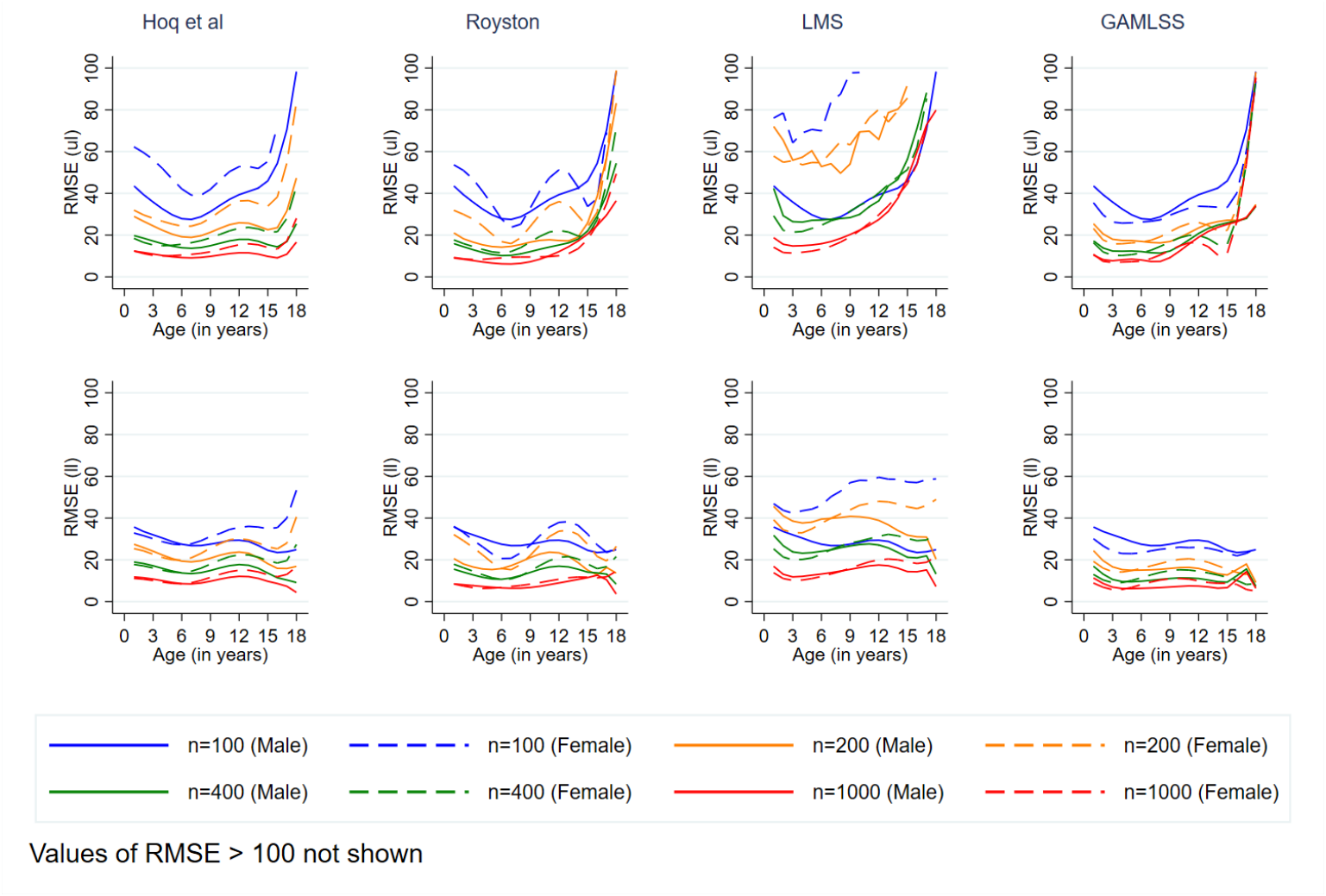


Figure 3: Root Mean Square Error by Age, sex, sample size and method for scenario 2. The first row represents RMSE of the upper limit and second row represents RMSE of the lower limit.

The required number of samples depends on the pattern of age dependence and the statistical methods applied. Considering the decrease in average RMSE and overall coverage as sample size increases, a sample size of 200 appears to be sufficient for any of the statistical methods except Cole's LMS for a biomarker with reference values that are linearly associated with age with SD independent of age (as in Scenario 3). However, for a biomarker with SD that varies nonlinearly with age and a level that is influenced by sex (Scenarios 2, 4), a minimum of 400 samples may be required for Hoq *et al*, Royston's and GAMLSS method. For Cole's LMS method only achieved similar average RMSEs for most scenarios (1, 3, 4 and 5) at sample size 1000.

Results from the HAPPI Kids Study

Finally, we provide an illustration of differences between the RIs obtained by the various methods for two biomarkers, creatinine and ALP in data from the HAPPI Kids (Hoq, Karlaftis, *et al*., 2019) study, for sample sizes 350 and 630 (Figures 4 and 5). The differences in the continuous RIs for creatinine across the four methods were minimal. For ALP, the upper limits of the continuous RIs varied a little by method while the differences were negligible for the lower limits. For both creatinine and ALP, the difference in RIs applying regression-based parametric methods were minimal between sample sizes 350 and 630.

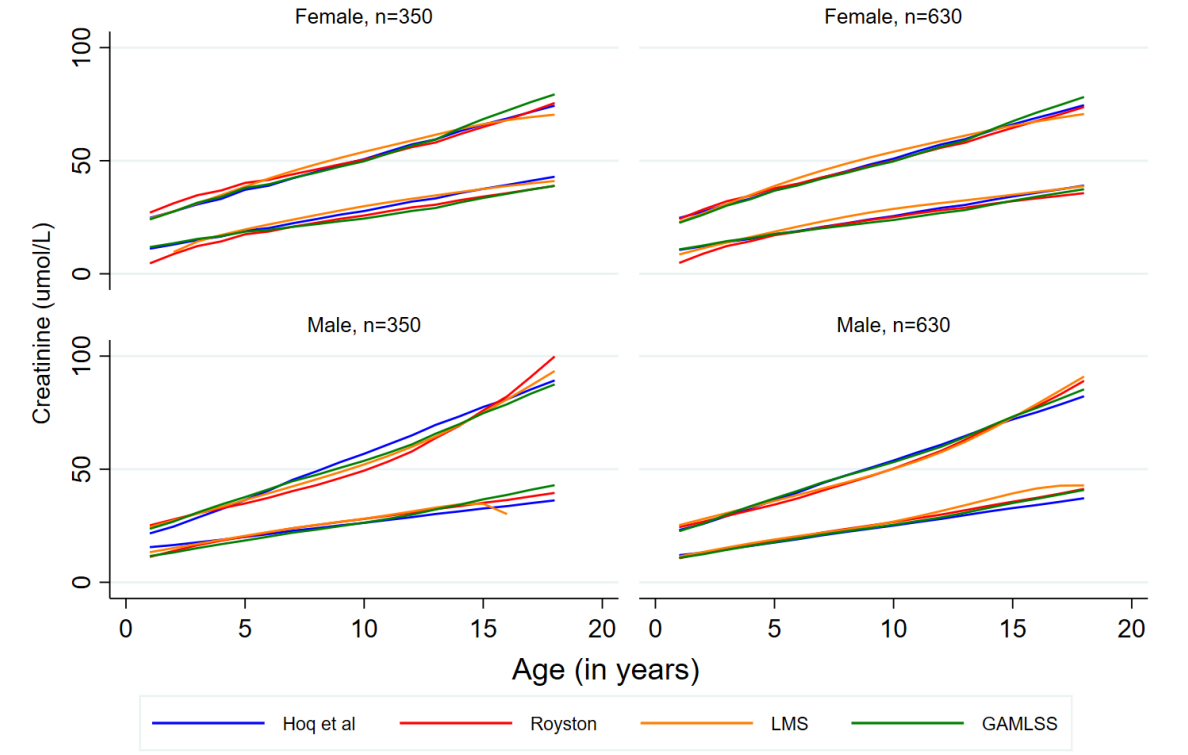


Figure 4: RIs for creatinine from the HAPPI Kids Study applying four methods for two sample sizes.

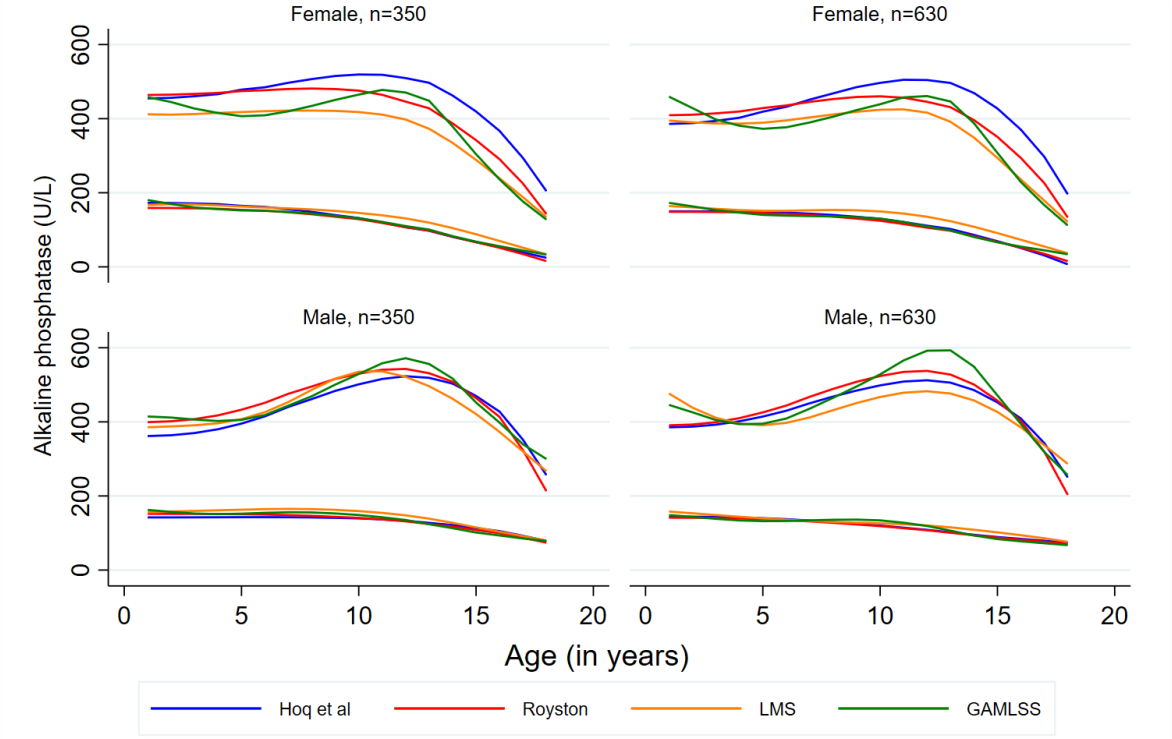


Figure 5: RIs for alkaline phosphatase form the HAPPI Kids Study applying four method for two sample sizes.

DISCUSSION

We have used a simulation study to compare the performance of four statistical methods for estimating continuous paediatric RIs in five scenarios for four different samples sizes. Regression-based parametric methods (Hoq *et al* [7] and Royston [10]) performed better in most scenarios and for most sample sizes. For the parametric methods and GAMLSS, the difference in RIs may not be clinically meaningful for sample size over 400. In contrast, a minimum of 1000 samples may be required to produce stable estimates using Cole's LMS methods. In a real data illustration, Hoq *et al*'s, Royston's and the GAMLSS method performed quite well and the results provided little basis to prefer one over the other in general.

The regression-based parametric methods performed well in most scenarios. For Royston's method, the exponential-normal and modulus-exponential-normal transformation allowed estimation of RIs for biomarkers with SDs associated with age including skewness and kurtosis [10]. Similarly, for Hoq *et al*'s method, the use of quantile regression (instead of estimating parameters of a complex parametric model) reduced RMSE, especially for biomarkers when the SD of the reference values was associated with age and curves differed by sex. In addition, the polynomial powers of age selected based on the mean values prevented estimates from being influenced by the reference values close to the limits. Compared to semi-parametric methods, the regression-based parametric methods may not be flexible enough to allow a dynamic variation for any age group based on sample reference values [15, 30].

The smoothness of the RI curves depends on identifying the polynomial powers of age for the regression-based methods or the EDF for each parameter of the transformation for the semi-

parametric methods. The selection procedure for the polynomial powers of age is well defined for the regression-based parametric methods [22]. In the absence of any clear recommendation for identifying EDF for the best fitting curves, we have applied a cross-validation approach based on minimising the error in prediction for Cole's LMS method [9] and used SBC i.e. log n instead of 2 as the penalty minimising the log likelihood for GAMLSS method [12].

Royston recommended 292 samples based on restricting the standard error (SE) of the limits of the estimated RIs to less than 10% of the constant SD [17]. Similarly, we found that 200 samples are sufficient to estimate age-specific RIs for a biomarker with reference values that are linearly associated with age with SD independent of age. However, Griffiths *et al*'s recommended more than 500 samples based on a simulation study for complex relationship between age and biomarker [13]. This recommendation is also similar to the findings of this study. Most studies that have used Cole's LMS and GAMLSS methods had large sample sizes [20, 31, 32]. This study confirms that these two methods perform well for large sample sizes.

The differences seen in the estimated RIs for biomarkers from the HAPPI Kids study using the four methods were consistent with the findings of our simulation study. The small differences in the estimated RIs from sample size 350 to 630 using Hoq *et al*'s and Royston's method may not be clinically significant [2, 30]. As already discussed, the flexibility of the GAMLSS methods suggested dynamic variations during pubertal ages for ALP. However, for large sample sizes the differences between methods may be minimal.

The regression-based parametric methods have some advantages over semi-parametric methods, such as expressing continuous RIs as an explicit function of age and sex, and estimation of SE of the estimated RIs [7, 10, 14]. These functions could easily be incorporated in the modern laboratory information system, a challenge highlighted by several studies in implementing continuous RIs [5, 7, 31, 33]. Estimation of a standard error enables assessment of the precision of the RI limits [7, 10, 14]. However, the CIs for the continuous RI limits were too narrow for large samples, resulting in poor coverage [10]. Unlike the parametric methods, Cole's LMS and GAMLSS methods do not provide estimates of variability of estimation [12,21]. For these methods, a standard error could be estimated by bootstrapping of subsamples [8, 15, 30].

In this study, we have only modelled age and sex-specific variation, where sex is a binary variable. Further research is needed to understand the performance of different methods based on the relationship of the biomarker with age and a second continuous variable such as height [16].

For the simulations, the data were generated using parametric functions that may have favoured the parametric methods. An alternative approach for data generation could be to take multiple sub-samples from a real dataset [16]. However, defining the true values, which is essential in assessing the performance, would have been difficult unless we had a very large sample.

CONCLUSIONS

In summary, we compared four statistical methods for estimating continuous paediatric RIs. Our findings indicate that the regression-based parametric methods described by Hoq *et al*

and Royston performed consistently well in most scenarios and for sample sizes 400 and over. The findings of the simulation study should provide evidence for future guidelines on the statistical method for estimating continuous RIs.

List of Abbreviations

Alkaline phosphatase (ALP), confidence interval (CI), equivalent degrees of freedom (EDF), exponential normal (EN), generalised additive model for location shape and scale (GAMLSS), generalised additive models (GAM), lambda-mu-sigma (LMS), root mean squared error (RMSE), and standard deviation (SD).

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The HAPPI Kids Study data analysed during the current study is not publicly available due to ethics requirements but are available from the corresponding author on reasonable request. The software code written for the simulation study is available as supplementary information or from the corresponding author upon reasonable request.

Competing Interest

The authors declare that they have no competing interests.

Funding

MH was support by the Postgraduate Scholarships Grant, The National Health and Medical Research Council, Australia, Grant Number: APP1168363, year 2019 and The Royal Children's Hospital Foundation, Melbourne, Australia, Grant Number 233, year 2014. The funding bodies have no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors Contributions

MH and JC designed the study with input from PM and SD. MH then designed the simulation study, analysed and interpreted all the data and drafted the manuscript. PM provided critical review and input to the manuscript. SD and JC provided advice on the design of the study, analysis and interpretation of the data, and contributed in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the HAPPI Kids study for sharing the data for the purpose of data illustration. The authors extend their gratitude to Dr Margarita Moreno-Betancur, Rushani Wijesuriya and Diana Zannino for their technical support in implementing the statistical simulation study.

REFERENCE

1. CLSI: CLSI Releases Guidelines for Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory. In.: COMTEX News Network, Inc.; 2008.

2. Daly CH, Higgins V, Adeli K, Grey VL, Hamid JS: Reference interval estimation: Methodological comparison using extensive simulations and empirical data. Clin Biochem 2017, 50(18):1145-1158.
3. Rodriguez-Capote K, Tovell K, Holmes D, Dayton J, Higgins TN: Analytical evaluation of the Diazyme glycated serum protein assay on the siemens ADVIA 1800: comparison of results against HbA1c for diagnosis and management of diabetes. J Diabetes Sci Technol 2015, 9(2):192-199.
4. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N: The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. Critical Reviews in Clinical Laboratory Sciences 2017, 54(6):358-413.
5. Loh TP, Antoniou G, Baghurst P, Metz MP: Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. Pathology 2014, 46(4):336-343.
6. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, Rauh M: Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clinical Chemistry 2015, 61(7):964-973.
7. Hoq M, Matthews S, Karlaftis V, Burgess J, Cowley J, Donath S, Carlin J, Yen T, Ignjatovic V, Monagle P: Reference Values for 30 Common Biochemistry Analytes Across 5 Different Analyzers in Neonates and Children 30 Days to 18 Years of Age. Clinical Chemistry 2019:clinchem.2019.306431.
8. Asgari S, Higgins V, McCudden C, Adeli K: Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals. Clin Biochem 2019, 73:82-89.
9. Cole TJ, Green PJ: Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992, 11(10):1305-1319.

10. Royston P, Wright EM: A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1998, 161(1):79-101.
11. Gannoun A, Girard S, Guinot C, Saracco J: Reference curves based on non-parametric quantile regression. Stat Med 2002, 21(20):3119-3135.
12. Rigby RA, Stasinopoulos DM: Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics) 2005, 54(3):507-554.
13. Griffiths JK, Iles TC, Koduah M, Nix AB: Centile charts II: alternative nonparametric approach for establishing time-specific reference centiles and assessment of the sample size required. Clin Chem 2004, 50(5):907-914.
14. Kroon FPB, Ramiro S, Royston P, Le Cessie S, Rosendaal FR, Kloppenburg M: Reference curves for the Australian/Canadian Hand Osteoarthritis Index in the middle-aged Dutch population. Rheumatology (Oxford) 2017, 56(5):745-752.
15. Wright EM, Royston P: A Comparison of Statistical Methods for Age-related Reference Intervals. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1997, 160(1):47-69.
16. Li K, Hu L, Peng Y, Yan R, Li Q, Peng X, Song W, Ni X: Comparison of four algorithms on establishing continuous reference intervals for pediatric analytes with age-dependent trend. BMC Medical Research Methodology 2020, 20(1):136.
17. Royston P: Constructing time-specific reference ranges. Statistics in Medicine 1991, 10(5):675-690.
18. Hoq M, Karlaftis V, Matthews S, Burgess J, Donath S, Carlin J, Monagle P, Ignjatovic V: A prospective, cross-sectional study to establish age-specific reference

intervals for neonates and children in the settings of clinical biochemistry, immunology and haematology: The HAPPI Kids study. BMJ Open 2019.

19. Morris TP, White IR, Crowther MJ: Using simulation studies to evaluate statistical methods. Statistics in Medicine 2019, 38(11):2074-2102.

20. Hoq M, Canterford L, Matthews S, Khanom G, Ignjatovic V, Monagle P, Donath S, Carlin J: Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review. Clin Biochem 2020.

21. Cole TJ: The LMS method for constructing normalized growth standards. Eur J Clin Nutr 1990, 44(1):45-60.

22. Royston P, Altman DG: Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. Journal of the Royal Statistical Society Series C (Applied Statistics) 1994, 43(3):429-467.

23. Royston P, Ambler G: Multivariable fractional polynomials. Stata Technical Bulletin 1998, 43:24-32.

24. Hastie T, Tibshirani R: Generalized additive models for medical research. Stat Methods Med Res 1995, 4(3):187-196.

25. Rigby RA, Stasinopoulos DM: Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. Statistical Modelling 2006, 6(3):209-229.

26. Stasinopoulos M, Rigby B, Akantziliotou C: Instructions on how to use the gamlss package in R Second Edition. In.; 2008.

27. StataCrop: Stata Statistical Software : Release 15. In. College Station, TX: StataCrop LP; 2017.

28. Royston P, Sauerbrei W: Two techniques for investigating interactions between treatment and continuous covariates in clinical trials. Stata Journal 2009, 9(2):230-251.

29. Zhiqiang Wang MSP: COLELMS: Stata module to calculate Cole's LMS values for growth data. In: Statistical Software Components. Boston College Department of Economics; 1998.

30. Wright EM, Royston P: Calculating reference intervals for laboratory measurements. Stat Methods Med Res 1999, 8(2):93-112.

31. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, Rauh M: Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clin Chem 2015, 61(7):964-973.

32. Bussler S, Vogel M, Pietzner D, Harms K, Buzek T, Penke M, Händel N, Körner A, Baumann U, Kiess W et al: New pediatric percentiles of liver enzyme serum levels (ALT, AST, GGT): Effects of age, sex, BMI and pubertal stage. Hepatology (Baltimore, Md) 2017.

33. Higgins V, Adeli K: Advances in pediatric reference intervals: from discrete to continuous. Journal of Laboratory and Precision Medicine 2018, 3(1).

6.3 Discussion

RIs are used as an assessment tool. It is important for the estimated RIs to be accurate. In our comparison in performance between the four methods, we have prioritised the method with minimum overall error of estimation of reference limits. Hence, we have considered RMSE that combines bias and variance, as the main performance measure since minimum error in estimation could be achieved by reducing bias or variance. RMSE was useful in aggregating the measure of estimation error across the age range 1 to 18 years into a single measure.

RMSE is dependent on age as failure to capture the change in biomarker with age (i.e. fractional polynomial powers of age or other parameters of transformation) would introduce bias specific to the age. However, as the sample size increase the model specification improve resulting in RMSE being less influenced by age. The RMSEs are lowest in middle ages most specifically at age 9.

Our proposed novel method performed consistently and similar if not better compared to other statistical methods in most scenarios [75]. The flexibility of the quantile regression allowed us to model complex shapes like ALP and phosphate with lowest average RMSE for the upper limit for sample size 400 and over. For biomarkers with a linear relationship with age e.g. creatinine and potassium, the differences in the average RMSE between our proposed novel method and Royston's method are minimal, with lowest average RMSE. The identification of fractional power of age depending on the mean reference values ensured the shape of the intervals were not influenced by the values close to the limits. Our method is also applicable for any distribution and does not depend on the parameters of distributions. Expressing the continuous RIs as an explicit function of age and sex, and estimation of SE of the estimated RIs are other advantages of our proposed method.

The sample size used by the HAPPI Kids Study was consistent with Royston's recommendation and was adequate for estimating continuous age-specific RIs [57]. In addition, the findings of the simulation study provided evidence that the sample size close to 400 was adequate and the differences in biases were minimal when compared with larger sample sizes (i.e. n = 1000) including biomarkers like creatinine and ALP. A sample size of 200 may be enough for a biomarker with linear relationship with age and SDs of the reference values that are linearly associated with age e.g. potassium, total protein, which is the case for most biomarkers. However, most RIs studies include a panel of biomarkers which may include a biomarker with complex relationship with age. Hence, we would recommend that a sample size of 400 or more is used for our method and Royston's method, whereas close to 1000 samples are required for GAMLSS, Cole's LMS [75].

Chapter 7: Discussion & Conclusion

7.1 Summary of findings

The HAPPI Kids study is one of the very few studies that collected data prospectively from pre-defined healthy children [64]. The study is also the first to utilise aliquots of the same blood sample from the same healthy participant and test the aliquots on different analysers, enabling head-to-head comparison between analysers in children [64]. However, the uniqueness of the study presented analytical challenges in terms of comparing the reference values between analysers taking the association with age into consideration, identifying the appropriate statistical methods for estimating continuous age-specific RIs with available data and validating the continuous age-specific RIs for use. In this thesis, I have addressed these analytical challenges and presented the findings of four studies.

Data from the HAPPI Kids study were used to estimate the continuous age-specific mean reference values on five analysers using mixed-effects regression analysis with fractional polynomial modelling of the means as a function of age and subsequently, to calculate coefficients of variation between the estimated mean values across age [66]. The findings revealed that the variations between analysers were clinically insignificant for 24 out of 30 biochemistry analytes tested since the maximum variation in mean reference values between five analysers for any age between 1 month and 18 years was less than the allowable limit of performance [66]. Consequently, common continuous age-specific paediatric RIs were reported as a function of age and sex for 24 biochemistry analytes applying a novel statistical method using 360 samples [66]. In addition, analyser-specific continuous paediatric RIs were reported for all the 30 biochemistry analytes [66].

To validate the continuous RIs for use in a routine laboratory, the existing validation guidelines were modified. We have proposed that at least 100 paediatric samples are tested, and samples are distributed uniformly across the range of the age and sex. The continuous RIs are considered valid for use in a routine laboratory if more than 90% of test results from the routine laboratory across all age and sex are within the intervals. Using the modified approach, the continuous RIs established using the HAPPI Kids data were validated for use in a routine laboratory [70].

There are a wide range of statistical methods available for estimating age-specific RIs with no consensus about the best method in practice. Hence, a systematic review was conducted examining the range of statistical methods used over the past 25 years for the estimation of age-specific RIs. The review revealed that statistical methods applied in previous research were not reported adequately and there was insufficient uptake of modern statistical methods for estimating age-specific paediatric RIs for analytes with age dependency [76].

Finally, we performed a simulation study to evaluate and compare statistical methods for constructing continuous age-specific RIs in children under different scenarios and for different sample sizes. The study findings demonstrated that our proposed novel statistical method performed similarly if not better in terms of differences between estimated and hypothetical true RIs in most scenarios compared to other commonly used statistical methods, and a sample size of 400 was adequate [75].

This research project also highlighted limitations in acquiring adequate samples from neonates and during the first month from birth, challenges in incorporating continuous age-

specific RIs into the existing LIS, and lack of guidelines to improve the uptake of modern statistical methods in estimating continuous age-specific RIs [77].

Some parts of the following sections are based on work published in the review paper highlting the current status, gaps, challenges and future considerations for paediatric RIs [77].

7.2 Translation of research findings into practice

7.2.1 Interpretation of pathology results obtained from multiple laboratories

Due to potential variation in pathology reference values derived from different analytical principles and/or analyser types (from the same or multiple laboratories), laboratories have a responsibility to report patient results with valid RIs for the method in use and the population they serve. Laboratories also play an important role in the education of physicians for the interpretation of pathology tests obtained from multiple laboratories [13, 66]. At present, it remains extremely difficult for physicians to identify, from the patient pathology report, biomarkers with adopted harmonised or common RIs (e.g. results comparable between laboratories), versus laboratory specific RIs that cannot be compared. This issue is of particular importance in health system models in which patients are diagnosed, managed and treated within both core and peripheral services and clinics.

CALIPER has taken the initiative in addressing the assessment of paediatric laboratory test results for physicians, healthcare workers, parents and patients worldwide, by the development and launch of a mobile application. The CALIPER App is an easy and transportable tool for the assessment of the latest reference value database (CALIPER) based on healthy children and adolescents [78]. The development of mobile Apps and complex LIS that enable the end user to access all RIs relevant to the patient's assessment would enable better clinical use of our understanding of inter-analyser differences.

7.2.2 Comparison of proportion of reference values outside the reference intervals using discrete versus continuous paediatric reference intervals

Zierk *et al* compared the flagging rate i.e. proportion of test results lying outside the RIs, in sick children and a reference population between discrete age-group RIs used in laboratories and published continuous age-specific RIs [12]. The authors reported that the flagging rate of test results outside the RIs for continuous age-specific RIs reduced substantially in comparison to discrete age-group RIs indicating improvement in clinical diagnosis for most blood count analytes and biochemical analytes, except for red cell distribution width, AST, ALT, potassium and lactate dehydrogenase (LDH) [12]. However, the proportion of samples and patients considered pathologic based on the continuous RIs was more than 5% for all analytes [12].

A recent CALIPER study reported that the flagging rate of test results outside the continuous RIs was approximately 2.5% below or above the lower or upper limit respectively for most analytes [36]. The flagging rate did not differ substantially when discrete age-group and continuous age-specific RIs using the same sample were compared [36]. However, the authors discussed that for continuous age-specific RIs the flagging rate of test results outside the RIs were consistent across age while discrete age-group RIs performed better near the centre of the age-group [36].

The difference in flagging rates between the study of Zierk *et al* and the CALIPER study is expected as one is an indirect study based on patient samples and the other is based on presumed healthy children [12, 36]. While the overall flagging rate can be made the same with both types of intervals, the continuous RIs are more likely to correctly classify each patient and should be used as a gold standard to compare the discrete age-group RIs.

7.2.3 Updating guidelines for statistical methods

As outlined in the systematic review study in Chapter 5, the statistical methods used in different steps of the RI estimation were not reported adequately in most publications [76]. This raises concern about the RIs estimated in different studies, and reproducing the RIs using the statistical methods will be challenging [76]. Hence, we recommend developing a checklist with a minimum requirement for reporting the statistical methods used for identifying outliers, assessing the distribution of the data and establishing RIs [76]. In addition, studies should report the 95% CI for the estimated RIs, which should be available in most statistical software.

The statistical methods applied in estimation of continuous age-specific RIs are very different from those utilised in the derivation of discrete age-group RIs [52]. Additionally, there is a variety of suitable methods applied for estimating continuous age-specific RIs [76], without any available consensus, guidelines or standards from peak professional bodies (e.g. IFCC Reference Intervals and Decision Limits Committee (C-RIDL)) to assist medical professionals and biostatisticians in this task. Hence, based on the findings from our simulation study reported in Chapter 6 [75] and other comparable studies [45, 79, 80], we recommend that regression based parametric methods should be used for estimating continuous age-specific RIs due to its minimal differences from the hypothetical true RIs. The regression based parametric methods are also useful in expressing the RIs as a function of age and allowing for estimaton of the SE (and thus 95% CI) for the estimated RIs.

7.2.4 Test results presented as a percentile

The representation of an individual's pathology test result outcome with appropriate RIs at discrete time-points (e.g. elevated biomarker at day x) can complicate clinical interpretation by masking of the trend in longitudinal test results that are not associated with the intra-

individual variation. This has been depicted in Figure 7-1, where 5, 10 and 15-year creatinine results for hypothetical patients, Eden and Ella, fell within the HAPPI Kids RIs and the "Australasian Harmonised Reference Intervals for Paediatrics" (AHRIP), but longitudinal trends in their respective creatinine levels in comparison with the trend observed for the HAPPI Kids (e.g. increasing for Eden and decreasing for Ella) in comparison with the trend observed for the HAPPI Kids continuous RI may be of clinical concern. In contrast, Eleanor's creatinine results of 30, 45 and 60 µmol/L (at 5, 10 and 15 years respectively) demonstrate her creatinine value remained stable on the 50th percentile. Future consideration should be given to the representation of paediatric pathology test results as percentiles, as commonly used for other child growth parameters (e.g. weight, height, and head circumference) and the potential advantages that this may have in assisting physicians in the clinical interpretation of individual data and thus monitoring of physiological conditions**.**

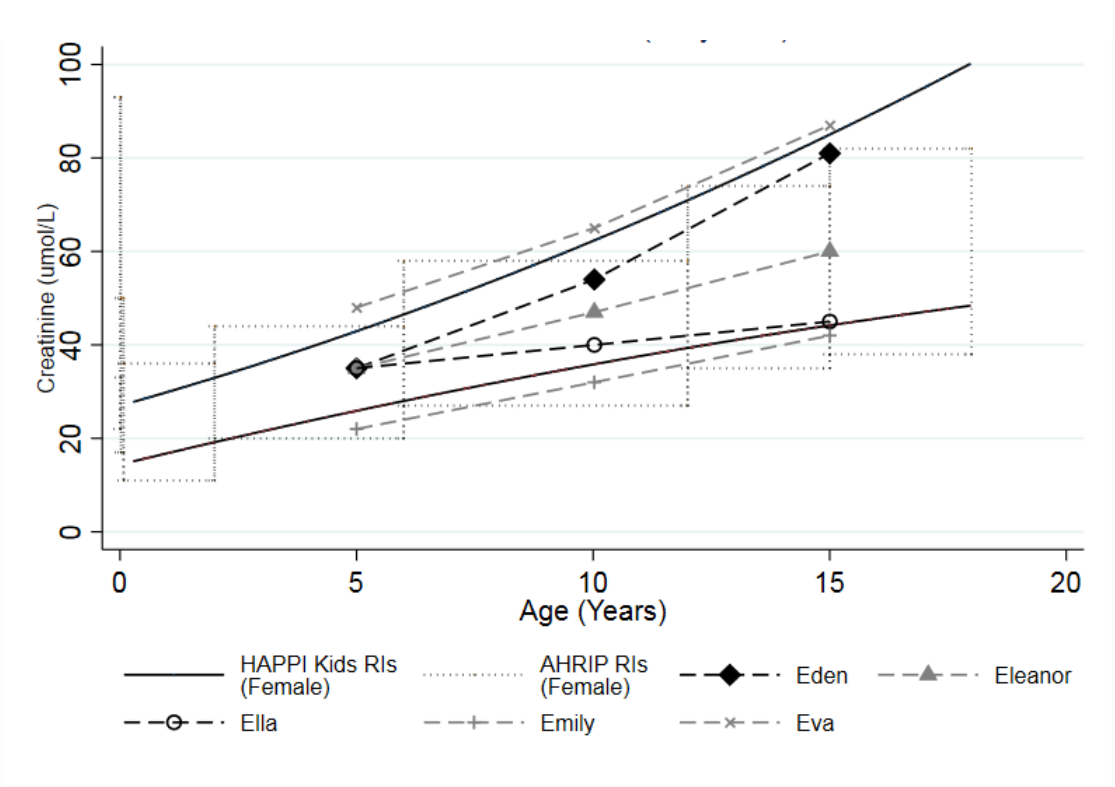


Figure 7-1: Visual comparison of trends in reference values over time between discrete age-group and continuous age-specific RIs.

7.3 Further research addressing the gaps and challenges

7.3.1 Targeted studies to establish neonatal and infant RIs

Physiological change in neonates is very dynamic, with the first few days of life being noted as "the most dramatic period of change that occurs in human life" [81]. As such, concentrations of respiratory and cardiovascular biomarkers can change within minutes after birth, whilst other biomarkers change more slowly as organ systems evolve to transition the neonate from the intrauterine environment to adult physiology [81]. These dramatic changes in biomarkers make the interpretation of pathological test results challenging, especially in the context of currently available paediatric RIs. As discussed previously in chapter 2, there is a lack of reliable and accurate neonatal and infantile RIs. Targeted studies of children aged 0, 1, 2, 3, 7, 14 and 28 days, and during the 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 9^{th}, and 12^{th} months of life are required. In an effort to address the limitations of available data, the CALIPER study has undertaken the recruitment of expectant mothers to facilitate the monitoring of circulating biomarkers at birth and throughout the first year of life [32]. Priority should be given to analytes with proven clinical utility during the neonatal and infancy periods.

7.3.2 Comparison of paediatric reference intervals between countries

Several countries have established continuous age-specific RIs and have harmonised RIs across laboratories or methods [31, 32, 66]. However, to date there has been no comparison of these RIs between countries. Considering CALIPER, Zierk *et al* and the HAPPI Kids study have published RIs for common biochemistry analytes, a unique opportunity now exists to compare the documented paediatric RIs between countries using an indirect approach [12, 36, 66]. While this type of comparative study will be confounded by pre- and post-analytical differences, and the statistical methods applied, it may still assist in understanding ethnic and geographic differences in certain biomarkers after adjusting for confounding factors. The

findings would significantly contribute to the establishment of global harmonisation guidelines and standards for paediatric RIs.

7.3.3 Incorporating results in an electronic database

The inability of existing LIS to incorporate advanced mathematical functions or graphical representation of patient pathology data has been identified and remains a key challenge to the incorporation of continuous age-specific RIs. Overcoming these post-analytical reporting issues may enable easier visual recognition of biomarker trends over time and thus improved longitudinal follow up of paediatric patients (See Figure 7-1).

Multi-national LIS and/or EMR developers have an active role to play in the future development and delivery of end-user requested functionality to improve clinical care in paediatric patient cohorts. From a commercial perspective, these types of LIS and EMR advancements would be directly transferrable to a wide variety of different patient cohort sub-sets, where physiological or treatment-based changes in biomarkers are evident (e.g. oxygen rates in newborn, therapeutic drug monitoring regimes, oncology patients etc.).

7.4 Conclusion

Paediatric RIs appropriate for the population being served by the servicing laboratory are essential for improving clinical interpretation and subsequent clinical decisions. As are the pre-analytical, analytical and post-analytical procedures utilised in the generation of test results. Despite the past and present challenges of establishing paediatric reference values and estimating paediatric RIs, the HAPPI Kids study has demonstrated that prospective collection of blood from healthy children is feasible. Addressing the analytical challenges, in this study, I have established continuous age-specific paediatric RIs for 30 biochemistry analytes on five

analysers as well as common RIs across analysers for 24 biochemistry analytes. I have provided further evidence by validating the established RIs and comparing our novel statistical method with other commonly used methods. I have made recommendation in terms of steps that should be applied in statistical analysis of the data and improvement in uptake of modern statistical methods in estimating age-specific RIs. I have also demonstrated that regression-based parametric methods perform better and a sample size of 400 is adequate for estimation of continuous age-specific RIs. These findings offer exciting improvements to pathology result interpretation, patient safety and ease of clinical management. However, with rapid advancement in technologies, both testing processes of the laboratory and the ways of reporting patient results are changing quickly. Hence, long-term planning to direct future research, software development and best practice clinical and statistical guidelines for paediatric RI development, implementation, review and accreditation will be critical for the sustainability of quality and acceptance into routine laboratory practise.

References

1. Rohr, U.-P., et al., *The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report.* PloS one, 2016. **11**(3): p. e0149856-e0149856.
2. The Royal College of Pathologists of Australasia, *Pathology The Facts*. 2019, RCPA.
3. Royston, P., *Estimation, reference ranges and goodness of fit for the three-parameter log-normal distribution.* Statistics in Medicine, 1992. **11**(7): p. 897-912.
4. Soldin, S.J., C. Brugnara, and E.C. Wong, *Pediatric References Intervals*. Jocelyn M. Hicks,editor emeritus ed, ed. 5. 2006, Washington, DC: AACC Press.
5. Whyte, M.B. and P. Kelly, *The normal range: it is not normal and it is not a range.* Postgraduate Medical Journal, 2018. **94**(1117): p. 613-616.
6. Ozarda, Y., et al., *Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits.* Crit Rev Clin Lab Sci, 2018. **55**(6): p. 420-431.
7. ISO, *15189:2012, I. (2017). ISO 15189:2012 - Medical laboratories -- Requirements for quality and competence. .* 2012.
8. Fuentes-Arderiu, X., *Biological reference intervals and ISO 15189.* Clin Chim Acta, 2006. **364**(1-2): p. 365-6.
9. Tahmasebi, H., et al., *Pediatric Reference Intervals for Biochemical Markers: Gaps and Challenges, Recent National Initiatives and Future Perspectives.* Ejifcc, 2017. **28**(1): p. 43-63.
10. Bloem, M., *The 2006 WHO child growth standards*. 2007, British Medical Journal Publishing Group.
11. Loh, T.P., et al., *Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals.* Pathology, 2014. **46**(4): p. 336-43.

12. Zierk, J., et al., *Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence.* Clin Chem, 2015. **61**(7): p. 964-73.

13. Adeli, K., et al., *Important considerations for interpreting biochemical tests in children.* BMJ, 2018. **361**: p. k1950.

14. Verras, P. and R. Greaves, *Interpreting paediatric biochemistry results.* 2005.

15. Colantonio, D.A., et al., *Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children.* Clin Chem, 2012. **58**(5): p. 854-68.

16. CLSI, *CLSI Releases Guidelines for Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory*. 2008, COMTEX News Network, Inc.

17. Ridefelt, P., et al., *Estimating reliable paediatric reference intervals in clinical chemistry and haematology.* Acta Paediatr, 2014. **103**(1): p. 10-5.

18. Coffin, C.M., et al., *Pediatric laboratory medicine: current challenges and future opportunities.* American Journal Of Clinical Pathology, 2002. **117**(5): p. 683-690.

19. Ozarda, Y., *Reference intervals: current status, recent developments and future considerations.* Biochem Med (Zagreb), 2016. **26**(1): p. 5-16.

20. Harris, E.K. and J.C. Boyd, *Statistical bases of reference values in laboratory medicine*. 1995: CRC Press.

21. Katayev, A., C. Balciza, and D.W. Seccombe, *Establishing Reference Intervals for Clinical Laboratory Test ResultsIs There a Better Way?* American Journal of Clinical Pathology, 2010. **133**(2): p. 180-186.

22. Howie, S., *Blood sample volumes in child health research: review of safe limits.* Bulletin of the World Health Organization, 2011. **89**.

23. Jones, G.R.D., et al., *Indirect methods for reference interval determination - review and recommendations.* Clin Chem Lab Med, 2018. **57**(1): p. 20-29.

24. Farrell, C.-J.L. and L. Nguyen, *Indirect Reference Intervals: Harnessing the Power of Stored Laboratory Data.* The Clinical biochemist. Reviews, 2019. **40**(2): p. 99-111.

25. Bailey, D., et al., *Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort.* Clin Chem, 2014. **60**(3): p. 518-29.

26. *Desirable Biological Variation Database Specificiations*. [cited 2019 Ferburuary]; Available from: https://www.westgard.com/biodatabase1.htm.

27. Jones, G.R.D., K. Sikaris, and J. Gill, *'Allowable Limits of Performance' for External Quality Assurance Programs – an Approach to Application of the Stockholm Criteria by the RCPA Quality Assurance Programs.* The Clinical Biochemist Reviews, 2012. **33**(4): p. 133-139.

28. Loh, T.P. and M.P. Metz, *Indirect estimation of pediatric between-individual biological variation data for 22 common serum biochemistries.* Am J Clin Pathol, 2015. **143**(5): p. 683-93.

29. Delanghe, J.R., et al., *Trueness verification of actual creatinine assays in the European market demonstrates a disappointing variability that needs substantial improvement. An international study in the framework of the EC4 creatinine standardization working group.* Clin Chem Lab Med, 2008. **46**(9): p. 1319-25.

30. Steinmetz, J., et al., *Standardization of gamma-glutamyltransferase assays by intermethod calibration. Effect on determining common reference limits.* Clin Chem Lab Med, 2007. **45**(10): p. 1373-80.

31. Tate, J.R., et al., *Harmonising Adult and Paediatric Reference Intervals in Australia and New Zealand: An Evidence-Based Approach for Establishing a First Panel of Chemistry Analytes.* The Clinical Biochemist Reviews, 2014. **35**(4): p. 213-235.

32. Adeli, K., et al., *The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper.* Crit Rev Clin Lab Sci, 2017. **54**(6): p. 358-413.

33. Pathology Harmony Group. *Harmonisation of Reference Intervals*. 2011; Available from: http://www.acb.org.uk/docs/default-source/committees/scientific/guidelines/acb/pathology-harmony-ii-biochemistry.pdf?sfvrsn=2.

34. Bhattacharya, C., *A simple method of resolution of a distribution into Gaussian components.* Biometrics, 1967: p. 115-135.

35. Tate, J.R., G. Koerbin, and K. Adeli, *Opinion Paper: Deriving Harmonised Reference Intervals - Global Activities.* Ejifcc, 2016. **27**(1): p. 48-65.

36. Asgari, S., et al., *Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals.* Clin Biochem, 2019. **73**: p. 82-89.

37. IFCC. *Paediatric Reference Range Initiatives*. 2019; Available from: https://www.ifcc.org/c-etplm/paediatric-reference-range-initiatives/.

38. Kohse, K.P., *KiGGS — The German survey on children's health as data base for reference intervals and beyond.* Clinical Biochemistry, 2014. **47**(9): p. 742-743.

39. Flanders, M.M., et al., *Pediatric Reference Intervals for Seven Common Coagulation Assays.* Clinical Chemistry, 2005. **51**(9): p. 1738-1742.

40. Rustad, P., et al., *The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties.* Scand J Clin Lab Invest, 2004. **64**(4): p. 271-84.

41. Daly, C.H., et al., *A systematic review of statistical methods used in constructing pediatric reference intervals.* Clin Biochem, 2013. **46**(13-14): p. 1220-7.

42. Daly, C.H., et al., *Reference interval estimation: Methodological comparison using extensive simulations and empirical data.* Clin Biochem, 2017. **50**(18): p. 1145-1158.

43. Estey, M.P., et al., *CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens Clinical Chemistry Assays: Direct validation using reference samples from the CALIPER cohort.* Clinical Biochemistry, 2013. **46**(13–14): p. 1197-1219.

44. Ni, X., et al., *Pediatric reference intervals in China (PRINCE): design and rationale for a large, multicenter collaborative cross-sectional study.* Science Bulletin, 2018. **63**(24): p. 1626-1634.

45. Li, K., et al., *Comparison of four algorithms on establishing continuous reference intervals for pediatric analytes with age-dependent trend.* BMC Medical Research Methodology, 2020. **20**(1): p. 136.

46. Southcott, E.K., et al., *Establishment of pediatric reference intervals on a large cohort of healthy children.* Clinica Chimica Acta, 2010. **411**(19–20): p. 1421-1427.

47. Bussler, S., et al., *New pediatric percentiles of liver enzyme serum levels (ALT, AST, GGT): Effects of age, sex, BMI and pubertal stage.* Hepatology (Baltimore, Md.), 2017.

48. Kohse, K.P., *National and international initiatives and approaches for the establishment of reference intervals in pediatric laboratory medicine.* LaboratoriumsMedizin, 2015. **39**(4): p. 197-212.

49. Hodge, V.J. and J. Austin, *A Survey of Outlier Detection Methodologies.* Artificial Intelligence Review, 2004. **22**(2): p. 85-126.

50. Horn, P.S., et al., *Effect of Outliers and Nonhealthy Individuals on Reference Interval Estimation.* Clinical Chemistry, 2001. **47**(12): p. 2137.

51. Yap, B.W. and C.H. Sim, *Comparisons of various types of normality tests.* Journal of Statistical Computation and Simulation, 2011. **81**(12): p. 2141-2155.

52. Wright, E.M. and P. Royston, *Calculating reference intervals for laboratory measurements.* Stat Methods Med Res, 1999. **8**(2): p. 93-112.

53. Cole, T.J., *The LMS method for constructing normalized growth standards.* Eur J Clin Nutr, 1990. **44**(1): p. 45-60.

54. Cole, T.J. and P.J. Green, *Smoothing reference centile curves: The lms method and penalized likelihood.* Statistics in Medicine, 1992. **11**(10): p. 1305-1319.

55. Stasinopoulos, D.M. and R.A. Rigby, *Generalized Additive Models for Location Scale and Shape (GAMLSS) in R.* 2007, 2007. **23**(7): p. 46.

56. Rigby, R.A. and D.M. Stasinopoulos, *Generalized additive models for location, scale and shape.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 2005. **54**(3): p. 507-554.

57. Royston, P., *Constructing time-specific reference ranges.* Statistics in Medicine, 1991. **10**(5): p. 675-690.

58. Royston, P. and E.M. Wright, *A method for estimating age-specific reference intervals ('normal ranges') based on fractional polynomials and exponential transformation.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 1998. **161**(1): p. 79-101.

59. Gannoun, A., et al., *Reference curves based on non-parametric quantile regression.* Stat Med, 2002. **21**(20): p. 3119-35.

60. Wei, Y., et al., *Quantile regression methods for reference growth charts.* Statistics in Medicine, 2006. **25**(8): p. 1369-1382.

61. Koenker, R., *Quantile regression cambridge univ*. 2005, Press.

62. Australasian Association of Clinical Biochemists, *Harmonisstion of adult and paediatric reference intervals for general checmistry analytes across Australian pathology laboratories*, in *Final performance report*. 2014, Australian Government: Department of Health.

63. Cole, T.J. and P.J. Green, *Smoothing reference centile curves: the LMS method and penalized likelihood.* Stat Med, 1992. **11**(10): p. 1305-19.

64. Hoq, M., et al., *A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the settings of clinical biochemistry, immunology and haematology: The HAPPI Kids study.* BMJ Open, 2019.

65. Ridefelt, P., et al., *Pediatric reference intervals for general clinical chemistry components - merging of studies from Denmark and Sweden.* Scand J Clin Lab Invest, 2018. **78**(5): p. 365-372.

66. Hoq, M., et al., *Reference Values for 30 Common Biochemistry Analytes Across 5 Different Analyzers in Neonates and Children 30 Days to 18 Years of Age.* Clinical Chemistry, 2019: p. clinchem.2019.306431.

67. Loh, T.P., E. Ranieri, and M.P. Metz, *Derivation of pediatric within-individual biological variation by indirect sampling method: an LMS approach.* Am J Clin Pathol, 2014. **142**(5): p. 657-63.

68. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* International Journal of Nursing Studies, 2010. **47**(8): p. 931-936.

69. Higgins, V. and K. Adeli, *Advances in pediatric reference intervals: from discrete to continuous.* Journal of Laboratory and Precision Medicine, 2018. **3**(1).

70. Hoq, M., et al., *Validation of the HAPPI Kids continuous age-specific paediatric reference intervals.* Journal of Applied Laboratory Medicine, 2020.

71. Hoq, M., et al. *Statistical methods applied in the estimation of age-specific paediatric reference intervals for a laboraotry blood test: A systematic review*. in *Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference*. 2019. Adelaide, SA: Clinical Biochemist Reviews Supplement.

72. Cole, T.J., *Fitting Smoothed Centile Curves to Reference Data.* Journal of the Royal Statistical Society. Series A (Statistics in Society), 1988. **151**(3): p. 385-418.

73. Royston, P., *Constructing time-specific reference ranges.* Stat Med, 1991. **10**(5): p. 675-90.

74. Arseneau, E. and C.M. Balion, *Statistical methods used in the calculation of geriatric reference intervals: a systematic review.* Clin Chem Lab Med, 2016. **54**(3): p. 377-88.

75. Hoq, M., et al. *Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study*. in *Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference*. 2019. Adelaide SA: Clinical Biochemist Review Supplement.

76. Hoq, M., et al., *Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review*. Clin Biochem, 2020.

77. Hoq, M., et al., *Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations.* Clinical Biochemist Reviews, 2020. **41**(2): p. 43-52.

78. CALIPER. *CALIPER Pediatric Reference Interval Database*. Available from: https://caliper.research.sickkids.ca/#/.

79. Wright, E.M. and P. Royston, *A Comparison of Statistical Methods for Age-related Reference Intervals.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 1997. **160**(1): p. 47-69.

80. Griffiths, J.K., et al., *Centile charts II: alternative nonparametric approach for establishing time-specific reference centiles and assessment of the sample size required.* Clin Chem, 2004. **50**(5): p. 907-14.

81. Doherty, T. and I. Salik, *Physiology, Neonatal*. 2019, Treasure Island (FL): StatPearls Publishing.

Appendices

Appendix A: Supplementary documents for the publication included in Chapter 2 (Hoq *et al*. A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: the HAPPI Kids study protocol. *BMJ Open*. 2019; 9(4):e025897)

Supplementary Table 1: Questionnaire for assessing inclusion and exclusion criteria.

Participant/Category	Questions / Data fields
All Sample	Study ID
	Date of Recruitment
	Site of Recruitment
	Date Sample Collected
	Hospital
	UR
	First Name
	Surname
	Date of Birth
	Sex
	Address
	Ethnicity of Mother
	Ethnicity of Father
	Tube Type Collected and Volume of blood collected
Neonates	Gestation
	Time of Birth
	Day of Bleed
	Nature of Delivery
	Weight
	APGAR Score at 5 minutes
	Feeding Method
	Date/Time of Vitamin K
	Current Medications
Paediatric	Procedure performed
	Current Medications
	Clinical Information
Haematology	Dose the participant have a presence of coagulation disorders?
	Does the participant have a family history of coagulation disorders?

	Is there participant currently on anticoagulation medication?
Immunology	Does the participant have a presence of an immune system disorder or immune deficiency syndrome?
	Does the participant have a presence of a genetic disorder?
	Does the participant have a presence of a rheumatologic disorder?
	Has any family member or relative of your child ever been diagnosed with an immune system disorder or immune deficiency syndrome?
	Has the participant had an infection or a febrile illness within the last 7 days?
	Has the participant had a hospital admission for intravenous (IV) antibiotics to clear an infection on more than 2 occasions in life?
	Has the participant needed 2 or more months of oral antibiotics more than 2 occasions in their life
	Does the participant have a history of a failure to thrive?
	Has the participant been a recipient of blood products in the last 3 months?
	Has a doctor ever diagnosed your child with food allergy, asthma, eczema or hayfever?
	Has anyone in your child's immediate family (that is, your child's parents and/or siblings) ever been diagnosed with food allergy, asthma, eczema or hayfever?
Biochemistry	Does the participant have a history of liver and renal disease?
	Does the participant have a presence of endocrine diseases?
	Does the participant have a presence of metabolic disease?
	Does the participant have a presence of renal disease?
	Does the participant have a presence of hepatic disease?
	Does the participant have a history of failure to thrive?

Appendix B: Supplementary documents for the publication included in Chapter 3 (Hoq *et al*. Reference values for 30 common biochemistry analytes across five different analyzers in neonates and children 30 days to 18 year olds, Clinical Chemistry. 2019; 65(10): 1317 - 1326.)

Supplementary Table 1: HKI Contributors

Role	**Person**	
Principle Investigator		
	Prof Paul Monagle	
	A/Prof Vera Ignjatovic	
Study Coordinators		
	Ms Vicky Karlaftis	
	Ms Chantal Attard	
	Ms Wincy Au-Yang	
	Ms Janet Burgess	
	Ms Jessica Cowley	
Participating Laboratories for sample testing		**Testing Scientist**
	The Royal Children's Hospital	Angela Chiriano
		Erik Haworth
		Chuong Tran
	The Royal Melbourne Hospital	Tina Nhan
	Melbourne Pathology	Mrs Daniella Angeleski
		Mr Warren Louey
	Monash Health	Dr Julie Newman
		Peter Edwards
	Dorevitch	Alan Webster
	Australian Clinical Laboratories	Daniel Cochrane
		Michelle Webber
Collaborating Clinicians for sample collection		**Site**
	Dr Ian McKenzie	Royal Children's Hospital
	Ms Jenny Ryan	Royal Women's Hospital
	Ms Anna Bourke	Royal Women's Hospital
	A/Prof Christine Rodda	Western Hospital
	Dr Martin Wright	Western Hospital
	A/Prof Wei Qi Fan	Northern Hospital
	Dr Simon Hauser	Northern Hospital

Pathology collectors		
	Ms Chauvy Burgess	
	Ms Jody Hand	
	Ms Kyra Hijniakoff	
	Ms Kathryn Bowers	
	Ms Lisa Howieson	
	Ms Aimee Chua	
Steering Committee		
	Prof Paul Monagle	Co-Chair
	A/Prof Vera Ignjatoivc	Co-Chair
	Dr Bill Renwick	Australian Clinical Laboratories
	Dr Sharon Choo	Royal Children's Hospital
	Dr Ellen Maxwell	Melbourne Pathology
	Cathy Cockshott	Dorevitch Pathology
	Dr Sue Matthews	Royal Children's Hospital
	Dr Kuang-Chih Hsiao	Royal Children's Hospital
	Dr Tina Yen	Royal Children's Hospital
	Mr Ray Czajko	Royal Melbourne Hospital
	Mr Peter Edwards	Biochemistry - Principal Scientist
	Mr Nick Drossos	Dorevitch
	Mr Daniel Cochrane	Australian Clinical Laboratories
	Ms Donna McNally	Australian Clinical Laboratories
	Ms Vicky Karlaftis	Study Coordinator
	Chantal Attard	Study Coordinator
	Jessica Cowley	Study Coordinator
	Mr Monsurul Hoq	Statistician
Biochemistry advisory panel		
	Dr Sue Matthews	Royal Children's Hospital
	Dr Tina Yen	Royal Children's Hospital
	A/Prof Ken Sikaris	Melbourne Pathology
	Dr Jim Doery	Monash Health
	Mr Ray Czajko	Royal Melbourne Hospital
Statistical analysis		
	Mr Monsurul Hoq	
	Prof John Carlin	
	A/Prof Susan Donath	

Supplementary Table 2: Analytical methods

Measurand	Analyser	ORTHO Vitros 5600	ABBOTT Architect (c4000, c8000, c16000)	Roche Cobas c701/c702	Siemens ADVIA 1650/1800	Beckman Coulter AU5800
Sodium	**Analytical principle**	Ion Selective Electrode - Direct (Undiluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 919. Flame Atomic Emission Spectroscopy	NIST Reference Material - Flame Photometry	The ISE Compensator is traceable against Flame Photometry (Sodium, Potassium) and Coulometry (Chloride).	CDC Flame Photometry Reference Method NIST SRM 909b	NIST SRM 919
Potassium (Serum)	**Analytical principle**	Ion Selective Electrode - Direct (Undiluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 918. Flame Atomic Emission Spectroscopy	NIST Reference Material - Flame Photometry	The ISE Compensator is traceable against Flame Photometry (Sodium, Potassium) and Coulometry (Chloride).	CDC Flame Photometry Reference Method NIST SRM 909b	NIST SRM 918
Chloride	**Analytical principle**	Ion Selective Electrode - Direct (Undiluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)	Ion Selective Electrode Indirect (Diluted)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 919. Coulometric-amperometric titration method	NIST Reference Material - Flame Photometry	The ISE Compensator is traceable against Flame Photometry (Sodium, Potassium) and Coulometry (Chloride).	NIST Coulometric Reference Method NIST SRM 919	NIST SRM 919
Bicarbonate	**Analytical principle**	Enzymatic (Phosphoenolpyruvate carboxylase/Malate dehydrogeanse) Endpoint	Enzymatic	Enzymatic	Enzymatic	Enzymatic
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 351. Corning model 965 Analyser with thermal conductivity detection	This method has been standardized against a primary standard.	This method has been standardized against a primary standard.	Corning 965 Reference Method NIST SRM 192	NIST SRM 351
Urea	**Analytical principle**	Urease. End Point with Ammonia generation	UREASE, Kinetic	UREASE, Kinetic	UREASE, Kinetic	UREASE, Kinetic
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 912. CDC Urease/GLDH method.	NIST SRM 912 - Differential Scanning Calorimetry	ID/MS	CDC Reference Method NIST SRM 912a, NIST SRM 909b	NIST SRM 909b, Level 1

Creatinine	**Analytical principle 1**	Creatinine Amido Hydrolase (enzymatic)	Alkaline Picrate (Jaffe), Kinetic	Enzymatic, Creatininase	Alkaline Picrate (Jaffe), Kinetic	Alkaline Picrate (Jaffe), Rate compensated
	Analytical principle 2	NA	NA	Alkaline Picrate (Jaffe), Kinetic	NA	NA
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Standardised to SRM 914 by GC/IDMS	Standardised to SRM 914/967 by IDMS	Standardised to SRM 914/967 by IDMS	IDMS Reference Method NIST SRM 967	Standardised to SRM 914/967 by IDMS
Bilirubin, Total	**Analytical principle**	Caffeine Sodium Benzoate, End Point	Diazo reaction, Diazonium Salt	Diazo reaction, Diazonium Salt	Chemical Oxidation (Vanadate)	Diazo reaction with Caffeine Benzoate
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Internal Manufactuer Master Lots to HPLC (Lauff et. al.)	The Doumas method/NIST SRM 916	The Doumas method.	AACC Reference Method NIST SRM 916	NIST SRM 916a
Alkaline Phosphatase	**Analytical principle**	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Multiple-point rate	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Other Rate Reaction	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Rate Reaction IFCC Recommended	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Rate Reaction IFCC Recommended	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Rate Reaction IFCC Recommended
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Internal manufacturer Calibrator 3. ALP IFCC method adapted to a centrifugal analyser at 37C.	IFCC method	IFCC method (2011).	IFCC Reference Method	Beckman Coulter master calibrator
Aspartate Amino Transferase	**Analytical principle 1**	Aspartate with Pyridoxal-5-phosphate multiple-point rate.	TRIS Buffer - Aspartate Approx. 240 mmolar with Pyridoxal-5-Phosphate (IFCC Recommended method)	TRIS Buffer - Aspartate Approx. 240 mmolar with Pyridoxal-5-Phosphate (IFCC Recommended method)	IFCC "Modified" without Pyridoxal-5-Phosphate	TRIS Buffer - Aspartate Approx. 240 mmolar with Pyridoxal-5-Phosphate (IFCC Recommended method)
	Analytical principle 2	NA	NA	IFCC "Modified" without Pyridoxal-5-phosphate	NA	NA
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Internal manufacturer Calibrator 3. AST IFCC method adapted to a centrifugal analyser at 37C.	IFCC method	The original IFCC formulation using calibrated pipettes together with a manual photometer providing absolute values and the substrate-specific absorptivity, ε.	IFCC Reference Method	IFCC method
Alanine Amino Transferase	**Analytical principle 1**	Alanine with Pyridoxal-5-Phosphate multiple-point rate	TRIS Buffer - Alanine approx. 500 mmolar with Pyridoxal-5-Phosphate (IFCC recommended method)	TRIS Buffer - Alanine approx. 500 mmolar with Pyridoxal-5-Phosphate (IFCC recommended method)	IFCC "modified" without Pyridoxal-5- phosphate	TRIS Buffer - Alanine approx. 500 mmolar with Pyridoxal-5-Phosphate (IFCC recommended method)
	Analytical principle 2	NA	NA	D - IFCC "modified" without Pyridoxal-5-phosphate	NA	NA
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)

	Traceability of Calibrator	Internal manufacturer Calibrator 3. ALT IFCC method adapted to a centrifugal analyser at 37C.	IFCC method	The original IFCC formulation using calibrated pipettes together with a manual photometer providing absolute values and the substrate-specific absorptivity, ε.	Standardised against IFCC	IFCC method and IRMM/IFCC-454
Gamma-Glutamyl Transferase	**Analytical principle**	L-Gamma-Glutamyl-p-nitroanilide and glycylglycine substrate multple-point rate.	Other Gamma-Glutamyl-3-carboxy-4-nitroanilide Substrate	Gamma-Glutamyl-3-Carboxy-4-nitroanilide Substrate - 6.-mmolar (IFCC recommended method)	Other Gamma-Glutamyl-3-carboxy-4-nitroanilide Substrate	Gamma-Glutamyl-3-Carboxy-4-nitroanilide Substrate - 6.-mmolar (IFCC recommended method)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Internal manufacturer Calibrator 3. GGT IFCC method adapted to a centrifugal analyser at 37C.	IFCC method	The original IFCC formulation (2002) and against the GGT method published by Persijn and van der Slik (1976), respectively.	Standardised against IFCC	IFCC method and IRMM/IFCC-452
Protein, total	**Analytical principle**	Biruet End Point - No Blank	Biruet End Point - No Blank	End Point with Blank	End Point with Blank	End Point with Blank
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 927. NCCLS credentialed Biruet method	NIST SRM 927 - Biuret	NIST SRM 927	Biuret Reference Method NIST SRM 927a	NIST SRM 927c
Albumin	**Analytical principle 1**	Bromocresol Green	Bromocresol Purple	Bromocresol Green - Citrate Buffer	Bromocresol Purple	Bromocresol Purple
	Analytical principle 2	NA	NA	Bromocresol Purple	NA	NA
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 927 assigned values to working human albumin standards prepared from human albumin. BCG Method.	ERM-DA470 - Visible Spectrophotomtery	ERM-DA470k / IFCC reference preparation	ERM-DA470k / IFCC Reference Material	Albumin reference material DA470 (IFCC)
Calcium	**Analytical principle**	Arsenazo III Dye	Arsenazo Dye	MN-BAPTA (5-nitro-5'methyl-BAPTA)	Arsenazo Dye	Arsenazo Dye
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 915. Flame atomic absorption spectroscopy.	NIST SRM 956 - ID-ICP-MS	SRM 956 c level 2 reference material	Inductively Coupled Atomic Emission Method NIST SRM 915, NIST SRM 909b	NIST SRM 909b, Level 1
Magnesium	**Analytical principle**	Formazan Dye	Enzymatic	Xylidyl Blue	Xylidyl Blue	Xylidyl Blue
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 929. Flame atomic absorption spectroscopy.	NIST SRM 956 - ID-ICP-MS	Atomic absorption spectrometry	Atomic Absorption Method NIST SRM 929	NIST SRM 909b, Level 1

Phosphate	**Analytical principle**	Phosphomolybdate Formation (and reduction).	Phosphomolybdate Formation	Phosphomolybdate Formation	Phosphomolybdate Formation	Phosphomolybdate Formation
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 200. Phosphomolybdate/phenylenediame method	NIST SRM 2186-I - Gravimetric	NERL primary Reference Material	Ammonium Molybdate Reference Method NIST SRM 937	Beckman Coulter master calibrator
Lactate dehydrogenase	**Analytical principle**	Pyruvate to Lactate - multiple-point rate Reaction - Pyruvate 17ug	Lactate to Pyruvate - IFCC recommended method	Lactate to Pyruvate - IFCC recommended method	Lactate to Pyruvate - Rate Reaction	Lactate to Pyruvate - IFCC recommended method
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Pyruvate to lactate total LDH method adapted to a centrifugal analyser.	Standardised against IFCC	The original IFCC6 formulation using calibrated pipettes together with a manual photometer providing absolute values and the substrate-specific absorptivity, ε.	Standardised against IFCC	IFCC method and IRMM/IFCC-453
Creatine Kinase	**Analytical principle**	Multiple-point Rate Reaction - NAC Activator	Rate Reaction - NAC Activator - IFCC recommended method - 2 part Imidazole Buffer	Rate Reaction - NAC Activator - IFCC recommended method - 2 part Imidazole Buffer	NAC Activator/Other Buffer	Rate Reaction - NAC Activator - IFCC recommended method - 2 part Imidazole Buffer
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Modified Scandinavian Committee on enzymes method for CK determination at 37C.	IFCC method	IFCC method	Standardised against IFCC	IFCC method
Lipase	**Analytical principle**	Spectrophotometric - 1 - oleoyl-2,3-diaectylglycerol substrate	Spectrophotometric - 1,2 - diglyceride substrate	Spectrophotometric - glutaric acid - (6 - methylresorufin) ester	Spectrophotometric - glutaric acid - (6 - methylresorufin) ester	Spectrophotometric - 1,2 - diglyceride substrate
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	Lipase activity in a standard triolein emulsion with a pH-Stat analyser.	Standard manufactured from purified human pancreatic lipase. Calibrator is matched to this standard.	Roche reagent using calibrated pipettes together with a manual photometer providing absolute values and the substrate-specific absorptivity, ε.	Internal Standard	Beckman Coulter master calibrator
Amylase	**Analytical principle**	Starch (amylopectin) - Dye (2 point Rate)	2-Chloro-PNP-alpha-maltotrioside	IFCC based ethylidene-G7-PNP hydrolysis to PNP	IFCC based ethylidene-G7-PNP hydrolysis to PNP	IFCC based ethylidene-G7-PNP hydrolysis to PNP
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)

	Traceability of Calibrator	Amylase paranitrophenol maltopentaoside method at 37C	Extinction Coefficient	Roche system reagent using calibrated pipettes together with a manual photometer providing absolute values and substrate-specific absorptivity, ε.	Standardised against IFCC	IFCC method and IRMM/IFCC-456
Uric Acid	Analytical principle	Uricase/peroxidase	Uricase/peroxidase	Uricase/peroxidase	Uricase/peroxidase	Uricase/peroxidase
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 913 using the CDC uricase method.	NIST SRM 913 - Gravimetric	ID/MS	CDC Candidate Reference Method (uricase) NIST SRM 913a, NIST SRM 909b	IDMS
Triglyceride	**Analytical principle**	Lipase / Glycerol Kinase/Glycerol-phophate oxidase (GPO)/Peroxidase	Lipase / Glycerol Kinase/ GPO-PAP (4-aminophenazone or 4-aminoantipyrine)	Lipase / Glycerol Kinase/ GPO-PAP (4-aminophenazone or 4-aminoantipyrine)	Lipase / Glycerol Kinase/ GPO-PAP (4-aminophenazone or 4-aminoantipyrine)	Lipase / Glycerol Kinase/ GPO-PAP (4-aminophenazone or 4-aminoantipyrine)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 1951. GPO triglyceride spectrophometric method	ACS Grade Glycerol - Gas Chrotatography.	ID/MS	Reference Method NIST SRM 909c	IDMS
Cholesterol	**Analytical principle**	Cholesterol ester hydrolase/Cholesterol Oxidase/Peroxidase	Cholesterol Oxidase/Peroxidase	Cholesterol Oxidase/Peroxidase	Cholesterol Oxidase/Peroxidase	Cholesterol Oxidase/Peroxidase
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 911. Centers for disease control (CDC) modified Abell-Kendall method.	SPRL Total Cholesterol Verification set - Abell-Kendall	Abell/Kendall and also by isotope dilution/mass spectrometry	CDC Reference Method (modified Abell-kendall) NIST SRM 909b	NIST SRM 909b, Level 1
HDL Cholesterol	**Analytical principle**	Separation method - Phosphotungstic acid (PTA)/Mg2+/Emulgen B-66 Cholesterol oxidase/peroxidase	Non Separation Method - Cholesterol esterase/Oxidase - Accelerator/Selective Detergent	Non Separation Method - Cholesterol esterase/Oxidase - Cyclodextrin Sulphate/Peg modified enzymes	Elimination	Separation methods - Cholesterol oxidase/peroxidase - DxSO4/Mg2+
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)

	Traceability of Calibrator	NIST SRM 911. Cholesterol Reference Method Laboratory Network (CRMLN) Designated Comparison method.	CDC reference method	The designated CDC reference method (ultracentrifugation method). The standardization meets the requirements of the "HDL Cholesterol Method Evaluation Protocol for Manufacturers" of the US National Reference System for Cholesterol, CRMLN (Cholesterol Reference Method Laboratory Network), November 1994.	NCEP Designated Comparison Method	CDC reference method
Iron	**Analyser**	**Roche Cobas c501/c502**	**ABBOTT Architect (c4000, c8000, c16000)**	**Roche Cobas c701/c702**	**Siemens ADVIA 1650/1800**	**Beckman Coulter AU5800**
	Analytical principle	Ferrozine Method without Deproteinisation	Colorimetric No Protein Removal - Ferrozine or Ferene Colour Reagent	Colorimetric No Protein Removal - Ferrozine or Ferene Colour Reagent	Colorimetric No Protein Removal - Ferrozine or Ferene Colour Reagent	Colorimetric No protein removal - Other colour Reagent
	Reagent	Roche Diagnostic	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)
	Traceability of Calibrator	NIST SRM 937.	NIST SRM 3126 - Gravimetric	SRM 937	Candidate AACC Reference Method (Ferrozine) NIST SRM 937	Beckman Coulter master calibrator
Ferritin	**Analyser**	**Roche cobas e601/e602**	**ABBOTT Architect (c4000, c8000, c16000)**	**Roche Cobas c701/c702**	**Simens ADVIA Centaur/XP/XPT**	**Beckman Counter Unicel Dxl 600/800**
	Analytical principle	Electrochemiluminescence (ECLIA) immunoassay (monoclonal ferritin-specific antibody)	Immunochemiluminometric Assay (ICMA)	Immunoturbidimetric	Immunochemiluminometric Assay (ICMA)	Immunochemiluminometric Assay (ICMA)
	Reagent	Roche Diagnostic	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Counter
	Traceability of Calibrator	National Institute for Biological Standards and Control (NIBSC) 80/602 with the enzymun-test ferritin method.	WHO 1st IS 80/602	WHO 1st IS 80/602	WHO 1st IS 80/602	WHO 2nd IS 80/578
Transferrin	**Analyser**	**Roche Cobas c501/c502**	**ABBOTT Architect (c4000, c8000, c16000)**	**Roche Cobas c701/c702**	**Siemens ADVIA 1650/1800**	**Beckman Coulter AU5800**
	Analytical principle	Immunoturbidimetric (Anti-rabbit Ab)	Immunoturbidimetric	Immunoturbidimetric	Immunoturbidimetric	Immunoturbidimetric
	Reagent	Roche Diagnostic	Abbott Diagnostic	Roche Diagnostic	Siemens (Bayer)	Beckman Coulter (AU)

	Traceability of Calibrator	Institute for Reference Materials and Measurements (IRMM) BCR470/CRM470 reference preparation for proteins in human serum (RPPHS)	Calibrators referenced against the International Protein Standard ERM - DA470	Calibrators referenced against the International Protein Standard ERM - DA470	Calibrators referenced against the International Protein Standard ERM - DA470	Calibrators referenced against the International Protein Standard ERM - DA470

	Analyser	**ORTHO Vitros 5600**	**ABBOTT Architect (i1000SR, i2000SR, i4000SR)**	**Rohe Cobas e602**	**Simens ADVIA Centaur/XP/XPT**	**Beckman Counter Unicel Dxl 600/800**
Thyroid Stimulating Hormone	**Analytical principle**	Immunometric Immunoassay	Immunochemiluminometric Assay (ICMA - Sandwich Asay)	Electrochemluminescence	Immunochemiluminometric Assay (ICMA - Sandwich Asay)	Immunochemiluminometric Assay (ICMA - Sandwich Asay)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Roche Diagnostics (Hitachi)	Siemens (Bayer)	Beckman Counter
	Traceability of Calibrator	2nd International Reference Preparation (IRP) 80/558	WHO 2nd IRP 80/558	WHO 2nd IRP 80/558	WHO 3rd IS 81/565	Traceable to WHO 3rd IS 81/565. Based on EN ISO 17511
Free Thyroxine	**Analytical principle**	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Abbott Diagnostic	Siemens (Bayer)	Beckman Counter
	Traceability of Calibrator	In-house reference calibrators correlated to another commercially available test	Calibrators are matched to an abbott internal reference standard. Standard is manufactured by Gravimetric method.	The Elecsys FT4 method. The Elecsys FT4 assay is traceable to the Enzymun-Test which was standardized using equilibrium dialysis.	Traceable to an internal standard manufactured using U.S.P.(United States Pharmacopeia) material	Beckman Coulter working calibrators based on EN ISO 17511
Free Triiodothyronine	**Analytical principle**	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)	Luminescent Immunoassay (LIA - Competitive Assay)
	Reagent	Ortho-Clinical Diagnostics	Abbott Diagnostic	Abbott Diagnostic	Siemens (Bayer)	Beckman Counter
	Traceability of Calibrator	In-house reference calibrators correlated to another commercially available test	Calibrators are matched to an abbott internal reference standard. Standard is manufactured by Gravimetric method.	The FT3 assay (REF 03051986). The FT3 assay (REF 03051986) is traceable to the FT3 assay (REF 11731386) which was standardized using equilibrium dialysis.	Traceable to an internal standard manufactured using U.S.P.(United States Pharmacopeia) material	Beckman Coulter working calibrators based on EN ISO 17511

Supplementary Table 3: Number of outliers excluded based on statistical criteria

	Ortho Vitros		Abbott Archiotect		Roche Cobas				Siemens ADVIA		Beckman Coulter	
Analyte	N	n	N	n	N	n	N	n	N	n	N	n
Sodium	358	1	240	3	359	2			357	1	359	1
Potassium	358	0	359	2	359	0			357	0	359	2
Chloride	358	1	359	9	359	0			358	0	359	0
Bicarbonate	357	2	359	0	359	0			353	3	352	0
Creatinine	Enzymatic		Jaffe		Enzymatic		Jaffe		Jaffe		Jaffe	
	357	1	358	1	358	1	333	2	359	2	348	2
Urea	358	0	359	0	359	0			341	0	348	1
Calcium	357	1	334	0	359	0			342	1	325	1
Magnesium	357	0	360	0	359	1			342	0	332	0
Phosphate	357	0	360	0	359	0			342	0	350	0
Triglyceride	359	3	357	7	357	6			359	5	358	5
Cholesterol	359	0	358	0	357	0			359	0	359	0
HDL cholesterol	359	0	358	0	357	0			359	0	358	0
ALP	358	2	359	1	359	1			349	1	356	1
AST	With P5P		With P5P		With P5P		Without P5P		Without P5P		With P5P	
	360	3	361	3	361	3	327	3	361	3	348	3
ALT	With P5P		With P5P		With P5P		Without P5P		Without P5P		With P5P	
	358	8	334	7	359	10	359	8	359	8	358	21
GGT	339	3	361	3	361	3			348	3	354	3
Total protein	358	0	333	0	312	1			354	0	178	0
Total bilirubin	271	86	244	0	340	0			342	0	277	0
Albumin	BCG		BCP		BCP		BCG		BCP		BCP	
	358	0	359	0	359	2	339	0	359	0	346	0
Ldh	359	2	357	3	357	2			357	3	359	3
CK	357	11	357	11	356	9			358	9	358	10
Lipa	361	7	360	4	359	6			361	5	360	4
Amy	359	1	358	1	357	1			358	1	346	1
Uric Acid	359	0	358	0	358	0			358	0	358	0
Iron	356	0	358	0	358	0			359	0	353	0
Ferritin	359	8	358	9	357	8			359	11	359	8
Transferrin	358	0	358	2	358	0			359	0	358	0
TSH	359	5	358	4	359	4			324	4	359	4
FT3	358	0	358	0	359	0			348	0	312	0
FT4	357	2	356	0	359	3			356	2	312	1

Supplementary Table 4: Parametric equations for lower and upper limits of 30 biochemistry analytes by five analyser types for children aged 30 days to 18 years.

Analytes/Analyser	Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)
Sodium (mmol/L)		
Ortho	135.77 + 0.08 Age[a]	141.37 + 0.23 Age
Abbott	135.01 + 0.12 Age	140.90+0.18 Age
Roche	138	143.17 + 0.12 Age
Siemens	138	145
Beckman	135.93 + 0.07 Age	142.97+0.07 Age
Potassium (mmol/L)		
Ortho	4.14 – 0.03 Age	5.10 – 0.01 Age
Abbott	3.86 – 0.02 Age	5.00 – 0.02 Age
Roche	3.96 – 0.02 Age	5.01 – 0.01 Age
Siemens	3.98 – 0.02 Age	5.10 – 0.02 Age
Beckman	3.90 – 0.02 Age	4.92 – 0.02 Age
Chloride (mmol/L)		
Ortho	102.35 – 0.10 Age	119
Abbott	102	110
Roche	99.75 – 0.20 Age	113.72 – 0.13 Age
Siemens	104.45– 0.14 Age	111.0
Beckman	103.30 - 0.11 Age	110.0 – 0.08 Age
Bicarbonate (mmol/L)		
Ortho	17.67 + 0.13 Age	23.68 + 0.22*Age
Abbott	15.33 + 0.19 Age	26.71 + 0.19*Age
Roche	17.05+ 0.13 Age	41
Siemens	18.37 + 0.17 Age	37
Beckmann	11.62 + 0.60 $Age^{0.5}$	16.35 + 1.18 $Age^{0.5}$
Urea (mmol/L)		
Ortho	2.30 + 0.71 Male[b] – 0.02 Age*Male	5.93 + 0.04 Age + 0.14 Male - 0.01 Age *Male
Abbott	2.45 – 0.02 Age + 0.61 Male – 0.02 Age*Male	5.91 + 0.04Age + 0.02 Male - 0.01 Age*Male
Roche	2.20 + 0.72 Male – 0.03 Age*Male	6.30 – 0.18 Male + 0.01 Age*Male
Siemens	2.81 – 0.02 Age + 0.49 Male – $2.00*10^{-3}$ Age*Male	6.70 + 0.61 Male – 0.03 Age*Male
Beckman	2.40 – 0.23 Male + 0.03 Age*Male	6.45 – 0.02 Age + 0.58 Male – 0.01 Age*Male
Creatinine (µmol/L)		
Ortho (Enzymatic)	14.89 + 1.66 Age – 1.55 Male + 0.16 Age*Male	25.98 + 2.65 Age – 0.96 Male + 0.95 Age*Male
Roche (Jaffe)	13.43 + 1.79 Age – 2.42 Male + 0.37 Male*Age	26.13 + 2.95 Age – 6.17 Male + 1.28 Male*Age
Roche (Enzymatic)	16.22 + 1.93 Age –1.79 Male - $0.70*10^{-2}$ Male*Age	27.03 + 3.32 Age – 0.52 Male + 0.78 Age*Male
Abbott (Jaffe)	24.72 + 1.36 Age + 2.28 Male - 0.24 Male*Age	37.23 + 1.98 Age – 2.97 Male + 0.79 Male*Age

Analytes/Analyser	Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)
Creatinine (µmol/L)		
Siemens (Jaffe)	12.34 + 1.70 Age + 2.27 Male – 0.46 Male*Age	22.97 + 2.96 Age – 4.72 Male +1.12 Male*Age
Beckman (Jaffe)	7.77 + 1.96 Age + 3.46 Male – 0.53 Male*Age	21.87 + 2.55 Age – 3.13 Male + 0.92 Male*Age
Calcium (mmol/L)		
Ortho	2.22 + 0.09 $Age^{-0.5}$	2.57 + 0.03 $Age^{-0.5}$
Abbott	2.16 + 0.08 $Age^{-0.5}$	2.55
Roche	2.25 + 0.03 Age^{-1}	2.57 + 0.04 Age^{-1}
Siemens	2.24 + 0.03 Age^{-1}	2.59 + 0.02 Age^{-1}
Beckman	2.18 + 0.02 Age^{-1}	2.52 + 0.03 Age^{-1}
Magnesium (mmol/L)		
Ortho	0.75 – 0.03 ln(Age) + $1.72*10^{-3}$ Male	0.95 – 0.03 ln(Age) + 0.03 Male
Abbott	0.77 – 0.03 $Age^{0.5}$ + 0.04 Male	1.03 – 0.04 $Age^{0.5}$ + 0.01 Male
Roche	0.80 – 0.03 ln(Age) + 0.01 Male	1.02 – 0.03 ln(Age) + $3.31*10^{-3}$ Male
Siemens	0.84 – 0.03 ln(Age) + 0.02 Male	1.10 – 0.04 ln(Age) + 0.02 Male
Beckman	0.77– 0.03 ln(Age) + 0.01 Male	0.98– 0.03 ln(Age) + 0.02 Male
Phosphate (mmol/L)		
Ortho	1.44 + 0.12 Age^{-1} – $7.38*10^{-5}$ Age^3	2.08 + 0.15 Age^{-1} - $6.12*10^{-5}$ Age^3
Abbott	1.25 + 0.19 Age^{-1} – $8.46*10^{-5}$ Age^3	1.98 + 0.19 Age^{-1} - $6.71*10^{-5}$ Age^3
Roche	1.35 + 0.16 Age^{-1} – $9.10*10^{-5}$ Age^3	1.94 + 0.18 Age^{-1} – $5.70*10^{-5}$ Age^3
Siemens	1.37 + 0.27 Age^{-1} – $9.16*10^{-5}$ Age^3	2.07 + 0.14 Age^{-1} – $8.67*10^{-5}$ Age^3
Beckman	1.33 + 0.15 Age^{-1} – $9.03*10^{-5}$ Age^3	1.98 + 0.14 Age^{-1} – $7.36*10^{-5}$ Age^3
Triglyceride (mmol/L)		
Ortho	0.4	1.51 + $4.8*10^{-2}$ Age
Abbott	0.4	1.38 + $6.7*10^{-2}$ Age
Roche	0.37 + $8.50*10^{-3}$Age	1.60 + $5.7*10^{-2}$ Age
Siemens	0.4	1.57 + $5.25*10^{-2}$ Age
Beckman	0.4	1.48 + $6.90*10^{-2}$ Age
Cholesterol (mmol/L)		
Ortho	2.49 + $2.70*10^{-2}$ Age^2 – $9.40*10^{-3}$ Age^2ln(Age)	4.97 + $2.40*10^{-2}$ Age^2 – $7.50*10^{-3}$ Age^2ln(Age)
Abbott	2.58 + $3.4*10^{-2}$ Age^2 – $1.20*10^{-2}$ Age^2ln(Age)	5.13 + $2.69*10^{-2}$ Age^2 – $8.67*10^{-3}$ Age^2ln(Age)
Roche	2.49 + $1.24*10^{-2}$ Age^2 – $0.74*10^{-3}$ Age^3	5.16 + $1.00*10^{-2}$ Age^2 – $0.50*10^{-3}$ Age^3
Analytes/Analyser	**Lower Limit ($Y_{2.5}$)**	**Upper limit ($Y_{97.5}$)**

Cholesterol (mmol/L)		
Siemens	$2.59 + 1.41*10^{-2}$ Age^2 – $8.40*10^{-4}$ Age^3	$5.36 + 1.39*10^{-2}$ Age^2 – $7.11*10^{-4}$ Age^3
Beckman	$2.68 + 3.01*10^{-2}$ Age^2 – $1.08*10^{-2}$ Age^{2}ln(Age)	$5.39 + 2.21*10^{-2}$ Age^2 – $6.68*10^{-3}$ Age^{2}ln(Age)
HDL Cholesterol (mmol/L)		
Ortho	0.43 + 0.14 Age – 0.04 Age* ln(Age)	1.41 + 0.21 Age – 0.06 Age*ln(Age)
Abbott	0.39 + 0.17 Age – 0.05 Age*ln(Age)	1.44 + 0.13 Age – 0.04 Age*ln(Age)
Roche	0.28 + 0.22 Age – 0.07 Age*ln(Age)	1.58 + 0.13 Age – 0.04 Age*ln(Age)
Siemens	0.38 + 0.22 Age – 0.07 Age*ln(Age)	1.65 + 0.12 Age – 0.04 Age*ln(Age)
Beckman	0.38 + 0.23 Age – 0.07 Age*ln(Age)	1.45 + 0.15 Age – 0.05 Age*ln(Age)
Alkaline Phosphatase (U/L)		
Ortho	134.840 - 0.138 Age^3 + 0.043 Age^{3}ln(Age) + 0.160 Male*Age^3 – 0.055 Male*Age^{3}ln(Age)	324.311+ 0.171 Age^3 – 0.070 Age^{3}ln(Age) + 0.328 Male*Age^3 – 0.112 Male*Age^{3}ln(Age)
Abbott	130.834 - 0.081 Age^3 + 0.020 Age^{3}ln(Age) + 0.005 Male*Age^3 + 0.002 Male*Age^{3}ln(Age)	338.568+ 0.378 Age^3 – 0.145 Age^{3}ln(Age) – 0.028 Male*Age^3 + 0.014 Male*Age^{3}ln(Age)
Roche	146.728 - 0.118 Age^3 + 0.033 Age^{3}ln(Age) + 0.091 Male*Age^3 – 0.029 Male*Age^{3}ln(Age)	361.003+ 0.484 Age^3 – 0.182 Age^{3}ln(Age) + 0.224 Male*Age^3 – 0.075 Male*Age^{3}ln(Age)
Siemens	172.579 - 0.136 Age^3 + 0.038 Age^{3}ln(Age) + 0.091 Male*Age^3 – 0.028 Male*Age^{3}ln(Age)	430.929 + 0.610 Age^3 – 0.229 Age^{3}ln(Age) + 0.155 Male*Age^3 – 0.050 Male*Age^{3}ln(Age)
Beckman	165.013 - 0.1655 Age^3 + 0.050 Age^{3}ln(Age) + 0.149 Male*Age^3 – 0.049 Male*Age^{3}ln(Age)	440.898 + 0.208 Age^3 – 0.088 Age^{3}ln(Age) + 0.469 Male*Age^3 – 0.161 Male*Age^{3}ln(Age)
Aspartate Amino Transferase (U/L)		
Ortho (P5P)	30.43 – 0.66 ln(Age) – 1.59 (ln(Age))^2 + 0.75 Male	59.13 – 10.71 ln(Age)+ 1.78 (ln(Age))^2 – 0.17 Male
Abbott (P5P)	28.47 – 1.69 ln(Age) – 1.18 (ln(Age))^2 + 1.15 Male	56.74 – 6.90 ln(Age) -0.52 (ln(Age))^2 + 7.84 Male
Roche (P5P)	30.65 – 0.77 ln(Age) – 1.72 (ln(Age))^2 + $1.41*10^{-1}$ Male	59.07 – 10.67 ln(Age) + 1.73 (ln(Age))^2 – $6.6*10^{-2}$ Male
Beckman (P5P)	25.36 – 0.33 ln(Age) – 1.37 (ln(Age))^2 + 1.54 Male	54.62 – 10.83 ln(Age) + 1.47 (ln(Age))^2 + 2.439 Male
Siemens (no P5P)	29.52 – 1.45 ln(Age) – 1.28 (ln(Age))^2 + 2.07 Male	57.05 – 8.30 ln(Age) + 0.44 (ln(Age))^2 + 5.15 Male
Analytes/Analyser	**Lower Limit ($Y_{2.5}$)**	**Upper limit ($Y_{97.5}$)**

Aspartate Amino Transferase (U/L)		
Roche (no P5P)	26.23 – 0.53 ln(Age)– 1.57 $(\ln(Age))^2$ + 0.66 Male	50.58 – 8.62 ln(Age) + 0.97 $(\ln(Age))^2$ + 1.15 Male
Alanine Amino Transferase (U/L)		
Ortho (P5P)	15.71 + 2.16 Age^{-1} + 0.50 Male	40.82 + 2.56 Age^{-1} + 3.89 Male
Abbott (P5P)	2.81 + 1.83 Age^{-1} + 3.04*10^{-2} Male	23.22 + 5.07 Age^{-1} + 5.47 Male
Roche (P5P)	9.88 + 1.81 Age^{-1} + 0.66 Male	37.04 – 0.52 Age^{-1} + 1.01 Male
Beckman (P5P)	7.87 + 2.11 Age^{-1} + 0.61 Male	35.03 – 0.51 Age^{-1} + 0.52*10^{-2} Male
Siemens (no P5P)	9.44 + 2.53 Age^{-1} + 1.40 Male	36.65 – 0.82 Age^{-1} + 0.44 Male
Roche (no P5P)	7.88 + 2.11 Age^{-1} + 0.69 Male	29.96 + 0.52 Age^{-1} + 2.00 Male
Gamma Glutamyl Transferase (U/L)		
Ortho	5	17.32 + 0.89 Age
Abbott	5.69 + 7.65*10^{-2} Age	14.77 + 0.82 Age
Roche	4.88 + 1.32*10^{-1} Age	15.53 + 0.88 Age
Siemens	4.81 + 1.23*10^{-1} Age	15.47 + 0.83 Age
Beckman	3	17.08 + 0.85 Age
Total Protein (g/L)		
Ortho	54.94 + 2.87 ln(Age)	66.18 + 5.75 ln(Age)
Abbott	53.37 + 4.10 ln(Age)	67.10 + 5.43 ln(Age)
Roche	55.45 + 3.29 ln(Age)	67.60 + 5.15 ln(Age)
Siemens	57.87 + 2.12 ln(Age)	66.94 + 5.61 ln(Age)
Beckman	52.80 + 2.67 ln(Age)	65.83 + 4.82 ln(Age)
Total Bilirubin (µmol/L)		
Ortho	1	5.3 + 1.47 Age
Abbott	5	10.53 + 1.23 Age
Roche	1.62 + 7.9*10^{-2} Age	6.29 + 1.32 Age
Siemens	1.68 + 0.19 Age	9.28 + 1.35 Age
Beckman	2.29 + 0.11 Age	6.91 + 1.34 Age
Albumin (g/L)		
Ortho (BCG)	35.92 + 0.15 Age	43.57 + 0.48 Age
Abbott (BCP)	30	41.99 + 0.30 Age
Roche (BCP)	35	41.94 + 0.31 Age
Roche (BCG)	38.30 + 0.18 Age	47.15 + 0.27 Age
Siemens (BCP)	34.58 + 0.08 Age	41.94 + 0.32 Age
Beckman (BCP)	33.88 + 0.09 Age	41.88 + 0.23 Age
Lactate dehydrogenase (U/L)		
Ortho (P to L)	586.73 - 54.35 $Age^{0.5}$	1193.72 – 130.30 $Age^{0.5}$
Abbott (L to P)	145.91– 23.78 $Age^{0.5}$	520.79 – 62.58 $Age^{0.5}$
Roche (L to P)	263.50 – 31.10 $Age^{0.5}$	510.39 – 63.37 $Age^{0.5}$
Siemens (L to P)	275.74 – 31.45 $Age^{0.5}$	475.55 – 48.34 $Age^{0.5}$
Beckman (L to P)	280.44 – 34.34 $Age^{0.5}$	484.06 – 49.84 $Age^{0.5}$
Analytes/Analyser	**Lower Limit ($Y_{2.5}$)**	**Upper limit ($Y_{97.5}$)**

Creatinine Kinase (U/L)		
Ortho	65.12 – 3.24 Age – 8.61 Male + 2.20 Age*Male	251.56 – 4.19 Age + 16.98 Male + 3.29 Age*Male
Creatinine Kinase (U/L)		
Abbott	68.24 – 2.84 Age – 1.56 Male + 1.77 Age*Male	250.04 – 3.58 Age + 28.35 Male + 2.49 Age*Male
Roche	66.81 – 2.65 Age + 1.38 Male + 1.18 Age*Male	250.57 – 4.65 Age – 8.72 Male + 5.09 Age*Male
Siemens	73.54 – 2.98 Age + 1.90 Male + 1.63 Age*Male	300.15 – 5.70 Age + 26.11 Male + 2.11 Age*Male
Beckman	71.24 – 2.84 Age + 2.60 Male + 1.59 Age*Male	263.47 – 4.09 Age + 25.26 Male + 2.85 Age*Male
Lipase (U/L)		
Ortho	21.39 + 0.48 Age – 6.18 Male	151.06 – 1.09 Age – 35.60 Male
Abbott	4.12 + 0.18 Age – 2.15 Male	53.96 – 0.11 Age + 1.10 Male
Roche	13.00 – 2.00 Male	31.68 + 0.30 Age + 3.55 Male
Siemens	17.84 + 0.08 Age – 1.97 Male	33.78 + 0.43 Age – 0.80 Male
Beckman	4.96 + 0.06 Age – 1.70 Male	34.00 – 1.0 Male
Amylase (U/L)		
Ortho	15	68.14 + 22.93 ln(Age) – 0.13 Age^2
Abbott	11.56 + 8.90 ln(Age)– $2.69*10^{-2}$ Age^2	74.07 + 30.61 ln(Age) – 0.22 Age^2
Roche	12.35 + 10.84 ln(Age) – $5.11*10^{-2}$ Age^2	83.54 + 30.18 ln(Age) – 0.22 Age^2
Siemens	14.14 + 10.34 ln(Age) – $4.14*10^{-2}$ Age^2	83.51 + 37.77 ln(Age) – 0.30 Age^2
Beckman	10.02 + 12.54 ln(Age) – $4.15*10^{-2}$ Age^2	89.1 + 39.2 ln(Age) – 0.32 Age^2
Uric Acid (µmol/L)		
Ortho	144.81 + 0.12 Age^2 – 23.11 Male + 0.14 Age^2*Male	308.82 + 0.20 Age^2 + 14.07 Male + 0.43 Age^2*Male
Abbott	143.58 + 0.13 Age^2 – 25.47 Male + 0.24 Age^2*Male	319.62 + 0.22 Age^2 + 18.70 Male + 0.35 Age^2*Male
Roche	137.51 + 0.19 Age^2 – 29.39 Male + 0.19 Age^2*Male	311.74 + 0.21 Age^2 – 3.90 Male + 0.52 Age^2*Male
Siemens	153.10 + 0.18 Age^2 – 27.62 Male + 0.22 Age^2*Male	345.77 + 0.17 Age^2 + 13.84 Male + 0.61 Age^2*Male
Beckman	146.81 + 0.16 Age^2 – 30.2 Male + 0.23 Age^2*Male	322.65 + 0.22 Age^2 + 30.31 Male + 0.45Age^2*Male
Iron (µmol/L)		
Ortho	3.45 + 0.07 Age	21.62 + 0.50 Age
Abbott	2.21 + 0.18 Age	23.29 + 0.45 Age
Roche	2.57 + 0.13 Age	21.72 + 0.45 Age
Siemens	2.33 + 0.14 Age	23.19 + 0.45 Age

Analytes/Analyser	Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)
Iron (µmol/L)		
Beckman	1.95 + 0.12 Age	21.88 + 0.42 Age
Ferritin (µg/L)		
Ortho	33.14 -17.27 $Age^{0.5}$ -36.76 Male + 25.26 Male*$Age^{0.5}$ + 3.91 $Age^{0.5}$ln(Age) – 4.81 Male* $Age^{0.5}$ln(Age)	65.60 + 27.98 $Age^{0.5}$ – 4.86 Male + 0.93 Male*$Age^{0.5}$ – 4.73 $Age^{0.5}$ln(Age) -2.21Male* $Age^{0.5}$ln(Age)
Abbott	33.19 – 19.52 $Age^{0.5}$ – 3.11 Male – 2.59Male*$Age^{0.5}$ + 4.56 $Age^{0.5}$ln(Age) + 2.32 Male*$Age^{0.5}$ln(Age)	35.54 + 36.58 $Age^{0.5}$ + 75.55 Male – 64.40 Male* $Age^{0.5}$ – 8.07 $Age^{0.5}$ln(Age) + 17.12 Male*$Age^{0.5}$*ln(Age)
Roche	24.90 – 13.63 $Age^{0.5}$ – 36.26 Male + 25.23 Male*$Age^{0.5}$ +2.90 $Age^{0.5}$ln(Age) – 4.73 Male* $Age^{0.5}$ln(Age)	66.31 +24.78 $Age^{0.5}$ – 12.32 Male + 8.44 Male*$Age^{0.5}$ – 4.61 $Age^{0.5}$ln(Age) – 3.65 Male* $Age^{0.5}$ln(Age)
Siemens	27.96 – 15.90 $Age^{0.5}$ – 26.53Male +18.64 Male*$Age^{0.5}$ +3.6 $Age^{0.5}$ln(Age) – 3.6 Male* $Age^{0.5}$ln(Age)	169.75 – 59.49 $Age^{0.5}$ – 144.27 Male + 106.56 Male*$Age^{0.5}$ + 15.05 $Age^{0.5}$ln(Age) – 27.90 Male* $Age^{0.5}$ln(Age)
Beckman	27.87 -15.80 $Age^{0.5}$ – 20.44 Male + 13.25 Male*$Age^{0.5}$ + 3.56 $Age^{0.5}$ln(Age) – 2.27 Male* $Age^{0.5}$ln(Age)	33.37 + 22.08 $Age^{0.5}$ + 11.24 Male – 11.17Male*$Age^{0.5}$ – 3.75 $Age^{0.5}$ln(Age) + 1.90 Male* $Age^{0.5}$ln(Age)
Transferrin (g/L)		
Ortho	2.16 – 3.80*10^{-3}Age – 4.84*10^{-2} Male + 2.44*10^{-2} Age*Male	3.60 + 4.84*10^{-2}Age + 0.27 Male – 6.76*10^{-2} Age*Male
Abbott	0.70 + 7.80*10^{-2}Age + 0.4 Male – 3.8*10^{-2} Age*Male	2.75 + 8.32*10^{-2}Age + 0.78 Male – 0.10 Age*Male
Roche	2.10 – 2.80*10^{-2} Male + 1.30*10^{-2} Age*Male	3.50 + 3.10*10^{-2}Age + 0.30 Male – 4.40*10^{-2} Age*Male
Siemens	2.10 + 1.40*10^{-3}Age + 4.80*10^{-2} Male + 1.20*10^{-2} Age*Male	3.50 + 4.40*10^{-2}Age + 6.60*10^{-2} Male – 4.80*10^{-2} Age*Male
Beckman	2.20 – 1.20*10^{-3}Age – 6.20*10^{-2} Male + 1.40*10^{-2} Age*Male	3.60 + 3.50*10^{-2}Age + 0.10 Male – 5.30*10^{-2} Age*Male
Thyroid Stimulating Hormone (mlU/L)		
Ortho	0.91	5.41
Abbott	0.76	4.37
Roche	1	6
Siemens	0.81	4.62
Beckman	0.89	5.28

Analytes/Analyser	Lower Limit ($Y_{2.5}$)	Upper limit ($Y_{97.5}$)
Free Thyroxine (ρmol/L)		
Ortho	$12.50 - 1.21*10^{-3}Age^3 + 5.46*10^{-3}$ Male* $Age^3 + 1.25*10^{-4}$ $Age^3ln(Age) - 1.81*10^{-3}$ $Age^3ln(Age)$ Male - 1.25 Male	$22.94 - 1.12*10^{-2}Age^3 - 1.00*10^{-2}$ Male *$Age^3 + 3.61*10^{-3}Age^3ln(Age) + 3.67*10^{-3}$ $Age^3ln(Age)$ Male + 0.55 Male
Free Thyroxine (ρmol/L)		
Abbott	$11.87 - 5.52*10^{-3}Age^3 - 0.47$ Male $+ 1.76*10^{-3}$ Male*$Age^3 + 1.8510^{-3}$ Age^3 ln(Age) – $0.61*10^{-3}$ Age^3 ln(Age)*Male	$17.86 - 5.04*10^{-3}Age^3 - 1.23$ Male $- 5.33*10^{-3}$ Male *$Age^3 + 1.82*10^{-3}$ Age^3 ln(Age) + $1.94*10^{-3}$ Age^3 ln(Age)*Male
Roche	$15.28 - 7.6*10^{-3}Age^3 - 0.78$ Male $+ 5.46*10^{-4}$ Male*$Age^3 + 2.47*10^{-3}$ Age^3 ln(Age) – $5.10*10^{-5}$ Age^3 ln(Age)*Male	$23.18 - 3.00*10^{-3}Age^3 - 1.55$ Male $- 5.00*10^{-3}$ Male*$Age^3 + 0.84*10^{-3}$ Age^3 ln(Age) + $2.1*10^{-3}$ Age^3 ln(Age)*Male
Siemens	$13.13 - 8.20*10^{-3}Age^3 - 0.72$ Male $+ 4.43*10^{-3}$ Male*$Age^3 + 2.71*10^{-3}$ Age^3 ln(Age) - $1.42*10^{-3}$ Age^3 ln(Age)*Male	$20.02 - 6.91*10^{-3}Age^3 - 2.00$ Male $+ 8.06*10^{-3}$ Male $Age^3 + 2.25*10^{-3}$ Age^3 ln(Age) – $2.42*10^{-3}$ Age^3 ln(Age) Male
Beckman	$9.90 - 5.10*10^{-3}Age^3 - 0.78$ Male + $2.28*10^{-3}$ Male $Age^3 + 1.70*10^{-3}$ Age^3 ln(Age) - $7.21*10^{-4}$ Age^3 ln(Age) Male	$17.12 - 8.89*10^{-3}Age^3 - 2.40$ Male $+ 4.50*10^{-4}$ Male *$Age^3 + 2.90*10^{-3}$ Age^3 ln(Age) + $1.11*10^{-4}$ Age^3 ln(Age)*Male
Free Triiodothyronine (ρmol/L)		
Ortho	6.30 – 0.08Age – 0.80 Male + 0.10 Age *Male	10.32 – 0.11 Age – 1.16 Male + 0.17 Age *Male
Abbott	4.1 – 0.03 Age+ 0.32 Male – 0.005 Age*Male	6.8 – 0.05Age – 0.06 Male –0.02 Age *Male
Roche	5.15 – 0.02 Age - 0.05 Male + 0.02 Age*Male	8.69 – 0.05 Age+ 0.67 Male – 0.02 Age*Male
Siemens	5.87 – 0.07 Age – 0.97 Male + 0.09 Age*Male	9.43 –0.10 Age– 0.9 Male + 0.07 Age*Male
Beckman	5.92 – 0.09 Age – 0.82 Male + 0.09 Age*Male	9.35 – 0.06 Age – 0.85 Male + 0.06 Age*Male

a. In years

b. Replace male with "1", if subject is male. Otherwise replace male with "0" if subject is female.

Supplementary Table 5A: Summary data of neonates on Ortho Vitros

Analytes	n	Median	Min	Max
Sodium (mmol/L)	20	144	134	148
Potassium (mmol/L)	20	4	4	6
Chloride (mmol/L)	20	106	101	111
Bicarbonate (mmol/L)	20	22.5	18.0	27.0
Urea (mmol/L)	20	3.8	1.5	10.3
Creatinine (**µmol**/L)	20	59.0	40.0	157.0
Calcium (mmol/L)	20	2.39	2.07	2.59
Magnesium (mmol/L)	20	0.77	0.65	0.88
Phosphate (mmol/L)	20	2.31	1.90	2.79
Triglyceride (mmol/L)	20	1.10	0.50	2.60
Total Cholesterol (mmol/L)	20	2.20	1.40	3.20
High-density lipoprotein Cholesterol (mmol/L)	20	0.64	0.37	1.03
Alkaline Phosphatase (U/L)	20	125	79	384
Aspartate Amino Transferase (U/L)	20	60	38	106
Alanine Amino Transferase (U/L)	20	29	17	40
Gamma-Glutamyl Transferase (U/L)	20	102.0	49.0	190.0
Total Protein (g/L)	20	60	52	67
Total Bilirubin (µmol/L)	20	110	56	198
Albumin (g/L)	20	36.0	31.0	41.0
Lactate dehydrogenase (U/L)	20	1505	1012	2150
Creatine Kinase (U/L)	20	488	244	1208
Lipase (U/L)	20	41	19	165
Amylase (U/L)	20	15	15	34
Uric Acid (µmol/L)	20	0.294	0.191	0.389
Iron (**µ**mol/L)	20	8.6	4.6	11.5
Ferritin (**µ**g/L)	20	342	140	584
Transferrin (µg/L)	20	1.77	1.42	2.29
Thyroid Stimulating Hormone (mlU/L)	20	10.2	2.9	18.2
Free Thyroxine (**ρ**mol/L)	20	51.1	36.6	68.6
Free Triiodothyronine (ρmol/L)	20	11.2	5.9	18.6

Supplementary Table 5B: Summary data of neonates on Abbott Architect

	n	Median	Min	Max
Sodium (mmol/L)	20	140	134	148
Potassium (mmol/L)	20	4	4	6
Chloride (mmol/L)	20	107	100	112
Bicarbonate (mmol/L)	20	23.0	16.0	27.0
Urea (mmol/L)	20	3.8	1.6	10.2
Creatinine (µmol/L)	20	53.0	39.0	168.0
Calcium (mmol/L)	18	2.03	1.70	2.36
Magnesium (mmol/L)	18	0.74	0.60	0.90
Phosphate (mmol/L)	18	2.09	1.58	2.74
Triglyceride (mmol/L)	16	1.25	0.70	2.20
Total Cholesterol (mmol/L)	5	2.30	1.20	2.60
High-density lipoprotein Cholesterol (mmol/L)	20	0.60	0.30	1.00
Alkaline Phosphatase (U/L)	20	113	70	307
Aspartate Amino Transferase (U/L)	20	55	31	110
Alanine Amino Transferase (U/L)	5	6	6	11
Gamma-Glutamyl Transferase (U/L)	18	77.5	33.0	154.0
Total Protein (g/L)	20	53	38	65
Total Bilirubin (µmol/L)	20	114	58	218
Albumin (g/L)	20	31.0	24.0	35.0
Lactate dehydrogenase (U/L)	20	587	427	823
Creatine Kinase (U/L)	20	517	258	1304
Lipase (U/L)	16	14	4	52
Amylase (U/L)	5	15	8	30
Uric Acid (µmol/L)	16	0.315	0.200	0.410
Iron (µmol/L)	16	8.8	4.5	11.6
Ferritin (µg/L)	20	297	134	616
Transferrin (µg/L)	20	1.57	1.30	2.14
Thyroid Stimulating Hormone (mlU/L)	20	7.6	2.1	14.9
Free Thyroxine (ρmol/L)	20	29.4	23.0	39.6
Free Triiodothyronine (ρmol/L)	20	7.1	4.3	12.1

Supplementary Table 5C: Summary data of neonates on Roche Cobas

Analytes	n	Median	Min	Max
Sodium (mmol/L)	20	145	136	149
Potassium (mmol/L)	20	4	4	6
Chloride (mmol/L)	20	106	95	115
Bicarbonate (mmol/L)	20	27.5	17.0	36.0
Urea (mmol/L)	20	3.5	1.2	10.0
Creatinine (µmol/L) Enzymatic	20	60.5	42.0	175.0
Creatinine (U/L) Jaffe	20	55.0	34.0	181.0
Calcium (mmol/L)	20	2.39	2.11	2.65
Magnesium (mmol/L)	20	0.79	0.68	0.92
Phosphate (mmol/L)	20	2.19	1.76	2.64
Triglyceride (mmol/L)	20	1.20	0.70	2.50
Total Cholesterol (mmol/L)	20	1.95	1.30	3.00
High-density lipoprotein Cholesterol (mmol/L)	20	0.48	0.14	0.85
Alkaline Phosphatase (U/L)	20	129	84	434
Aspartate Amino Transferase (U/L) – P5P	20	57	39	109
Aspartate Amino Transferase (U/L) – No P5P	20	49	36	93
Alanine Amino Transferase (U/L) – P5P	20	18	10	35
Alanine Amino Transferase (U/L) – No P5P	20	16	8	30
Gamma-Glutamyl Transferase (U/L)	20	99.0	44.0	211.0
Total Protein (g/L)	18	57	48	68
Total Bilirubin (µmol/L)	20	113	66	193
Albumin (g/L) - BCP	20	39.0	30.0	38.0
Albumin (g/L) - BCG	20	33.0	35.0	44.0
Lactate dehydrogenase (U/L)	0	562	396	763
Creatine Kinase (U/L)	20	455	235	1157
Lipase (U/L)	20	13	8	31
Amylase (U/L)	20	14	5	35
Uric Acid (µmol/L)	20	0.290	0.190	0.380
Iron (µmol/L)	20	8.0	3.0	11.0
Ferritin (µg/L)	20	310	134	605
Transferrin (g/L)	20	1.70	1.40	2.20
Thyroid Stimulating Hormone (mlU/L)	20	11.5	3.2	20.0
Free Thyroxine (ρmol/L)	20	41.0	29.8	63.8
Free Triiodothyronine (ρmol/L)	20	9.1	5.0	15.8

Supplementary Table 5D: Summary data of neonates on Siemens ADVIA

Analytes	n	Median	Min	Max
Sodium (mmol/L)	21	145	136	149
Potassium (mmol/L)	21	4	4	6
Chloride (mmol/L)	21	107	102	113
Bicarbonate (mmol/L)	21	25.0	17.0	37.0
Urea (mmol/L)	21	4.2	1.8	11.3
Creatinine (µmol/L)	21	56.0	19.0	160.0
Calcium (mmol/L)	21	2.41	2.10	2.64
Magnesium (mmol/L)	21	0.86	0.73	0.94
Phosphate (mmol/L)	21	2.28	1.73	2.72
Triglyceride (mmol/L)	20	1.20	0.60	2.90
Total Cholesterol (mmol/L)	20	2.35	1.50	3.60
High-density lipoprotein Cholesterol (mmol/L)	20	0.65	0.40	1.10
Alkaline Phosphatase (U/L)	21	144	92	478
Aspartate Amino Transferase (U/L)	21	53	33	111
Alanine Amino Transferase (U/L)	21	20	10	38
Gamma-Glutamyl Transferase (U/L)	21	94.0	11.0	201.0
Total Protein (g/L)	21	62	55	70
Total Bilirubin (µmol/L)	21	121	4	205
Albumin (g/L)	21	33.0	30.0	37.0
Lactate dehydrogenase (U/L)	20	604	437	880
Creatine Kinase (U/L)	20	540	260	1287
Lipase (U/L)	20	18	13	27
Amylase (U/L)	20	16	7	37
Uric Acid (µmol/L)	20	0.330	0.210	0.430
Iron (µmol/L)	20	7.6	2.7	11.1
Ferritin (µg/L)	20	316	114	679
Transferrin (g/L)	20	1.71	1.43	2.31
Thyroid Stimulating Hormone (mlU/L)	16	9.6	2.5	17.1
Free Thyroxine (pmol/L)	21	8.6	5.5	14.4
Free Triiodothyronine (ρmol/L)	21	33.0	15.7	48.3

Supplementary Table 5E: Summary data of neonates on Beckman Coulter

Analytes	n	Median	Min	Max
Sodium (mmol/L)	20	147	136	152
Potassium (mmol/L)	20	4	4	6
Chloride (mmol/L)	20	108	99	112
Bicarbonate (mmol/L)	20	17.0	13.0	20.0
Urea (mmol/L)	20	3.8	1.5	11.2
Creatinine (µmol/L)	20	57.0	38.0	171.0
Calcium (mmol/L)	20	2.34	2.03	2.62
Magnesium (mmol/L)	20	0.81	0.69	0.93
Phosphate (mmol/L)	20	2.25	1.88	2.70
Triglyceride (mmol/L)	18	1.2	0.60	2.90
Total Cholesterol (mmol/L)	18	2.05	1.20	3.10
High-density lipoprotein Cholesterol (mmol/L)	18	0.74	0.46	1.18
Alkaline Phosphatase (U/L)	20	152	94	520
Aspartate Amino Transferase (U/L)	18	53	36	103
Alanine Amino Transferase (U/L)	20	16	6	32
Gamma-Glutamyl Transferase (U/L)	20	99.0	45.0	205.0
Total Protein (g/L)	15	57	50	66
Total Bilirubin (µmol/L)	20	111	66	191
Albumin (g/L)	20	32.0	29.0	36.0
Lactate dehydrogenase (U/L)	18	611	410	819
Creatine Kinase (U/L)	18	8	5	13
Lipase (U/L)	18	9	4	50
Amylase (U/L)	18	17	5	39
Uric Acid (µmol/L)	18	0.318	0.195	0.423
Iron (µmol/L)	18	8.0	3.0	11.0
Ferritin (µg/L)	18	223	87	463
Transferrin (g/L)	18	1.80	1.47	2.35
Thyroid Stimulating Hormone (mlU/L)	20	9.0	2.6	17.0
Free Thyroxine (ρmol/L)	18	34.4	24.0	52.6
Free Triiodothyronine (ρmol/L)	18	7.8	4.6	12.7

Supplementary Document 1: Estimation of age-specific reference intervals for alkaline phosphatase (ALP)

Initially we visually investigated the presence of outliers, normality assumptions and relations of biomarkers with age. Figure 1 suggests that there is an outlier with measurement > 800; ALP varies with age, and sex differences are evident after year 10. The distribution is approximately normal. However, the ranges are wide during puberty.

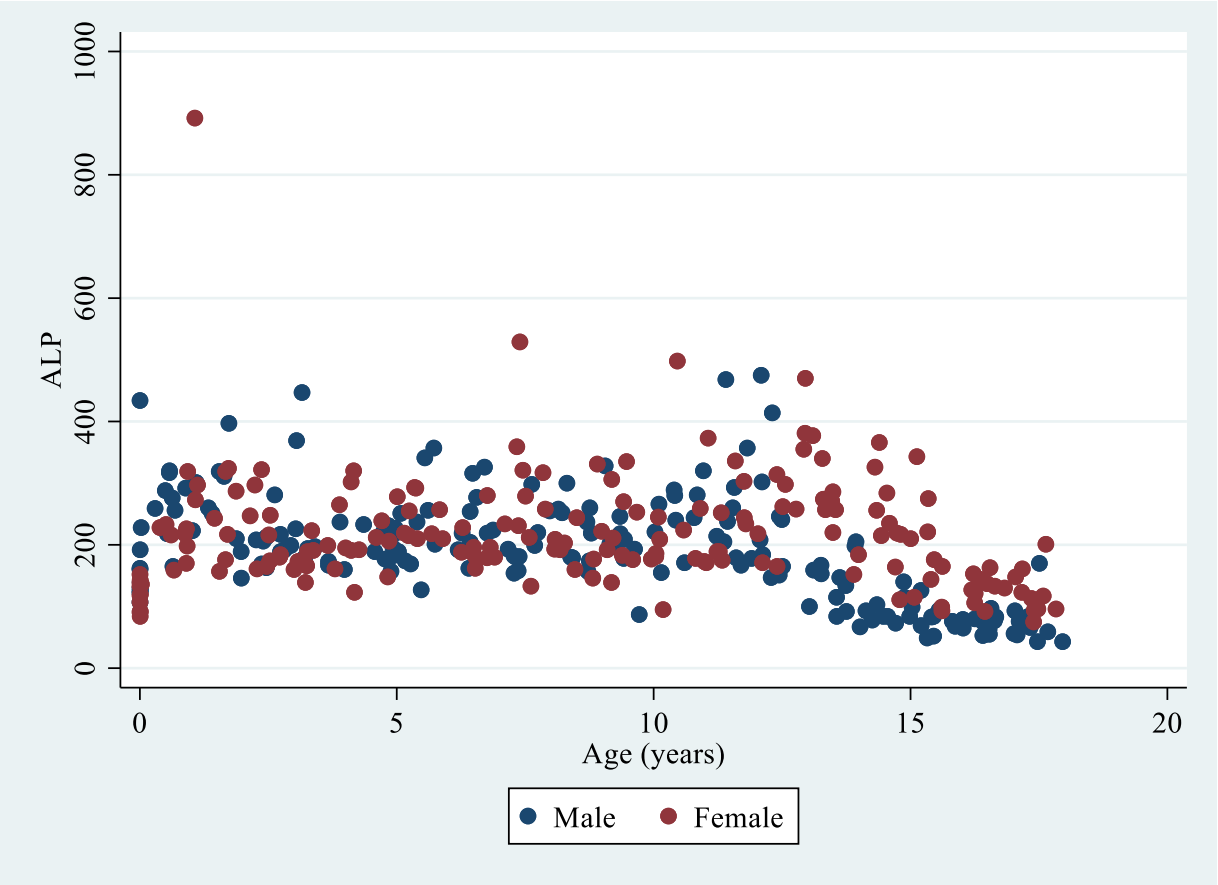


Figure 1: Scatter plot of ALP on Roche and age of the child

Following initial investigation we fitted a fractional polynomial regression model for the mean reference values, using the –**fp–** command in Stata.

For a degree-2 fractional polynomial model, a total of 44 models are fitted using different combination of powers from (-2, -1, -0.5, 0, 0.5, 1, 2, 3). The following table shows the best

model for each degree. The deviance as defined by twice the negative log likelihood is given for each model. If the difference in deviance between the model deviance and the model with lowest deviance is negligible, the model is selected. Otherwise the model with lowest deviance is selected. For example the difference in deviance between the model m =1 and m =2 is 13.9, is significant. Hence there is enough evidence that the model m = 2 fits the data better than the mode m = 1.

```
fp <age>, replace all: reg alp_mp sex <age>
(fitting 44 models)
(....10%....20%....30%....40%....50%....60%....70%....80%....90%....100%)

Fractional polynomial comparisons:
------------------------------------------------------------------------------
         age |   df    Deviance  Res. s.d.   Dev. dif.   P(*)   Powers
-------------+----------------------------------------------------------------
     omitted |    0   4245.388     89.716    103.464    0.000
      linear |    1   4185.969     82.706     44.045    0.000   1
       m = 1 |    2   4155.831     79.307     13.908    0.001   3
       m = 2 |    4   4141.924     77.895      0.000       --   3 3
------------------------------------------------------------------------------
(*) P = sig. level of model with m = 2 based on F with 353 denominator dof.

      Source |       SS           df       MS      Number of obs   =       359
-------------+----------------------------------   F(3, 355)       =     44.04
       Model |  801709.188         3  267236.396   Prob > F        =    0.0000
    Residual |  2154000.81       355   6067.6079   R-squared       =    0.2712
-------------+----------------------------------   Adj R-squared   =    0.2651
       Total |  2955709.99       358  8256.17317   Root MSE        =    77.895

------------------------------------------------------------------------------
      alp_mp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
        Male |   30.48784   8.222574     3.71   0.000     14.31676    46.65892
       age_1 |   .1634509   .0506174     3.23   0.001     .0639032    .2629987
       age_2 |  -.0670206   .0178976    -3.74   0.000    -.1022194   -.0318219
       _cons |    211.854   8.469113    25.01   0.000     195.1981      228.51
------------------------------------------------------------------------------
```

In order to exclude serious outliers, we used the above model to predict the mean values and calculate residuals. Since all the samples were collected from healthy children, we defined upper and lower bounds using 3*IQR of the residuals following the same principle as Tukey's test. The following box plot (figure 2) confirms that there is an outlying observation and the corresponding sample value was excluded from further analysis. The histogram in Figure 3 also suggests that the distribution is sufficiently close to normal.

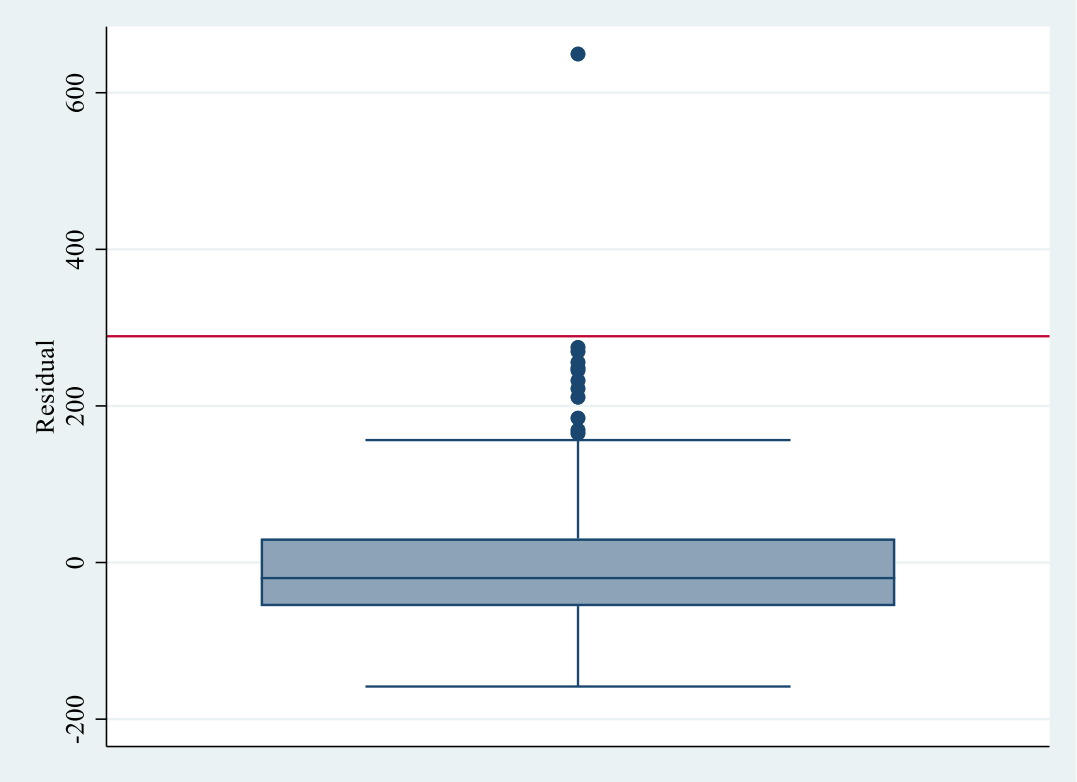


Figure 2: Box plot of residual

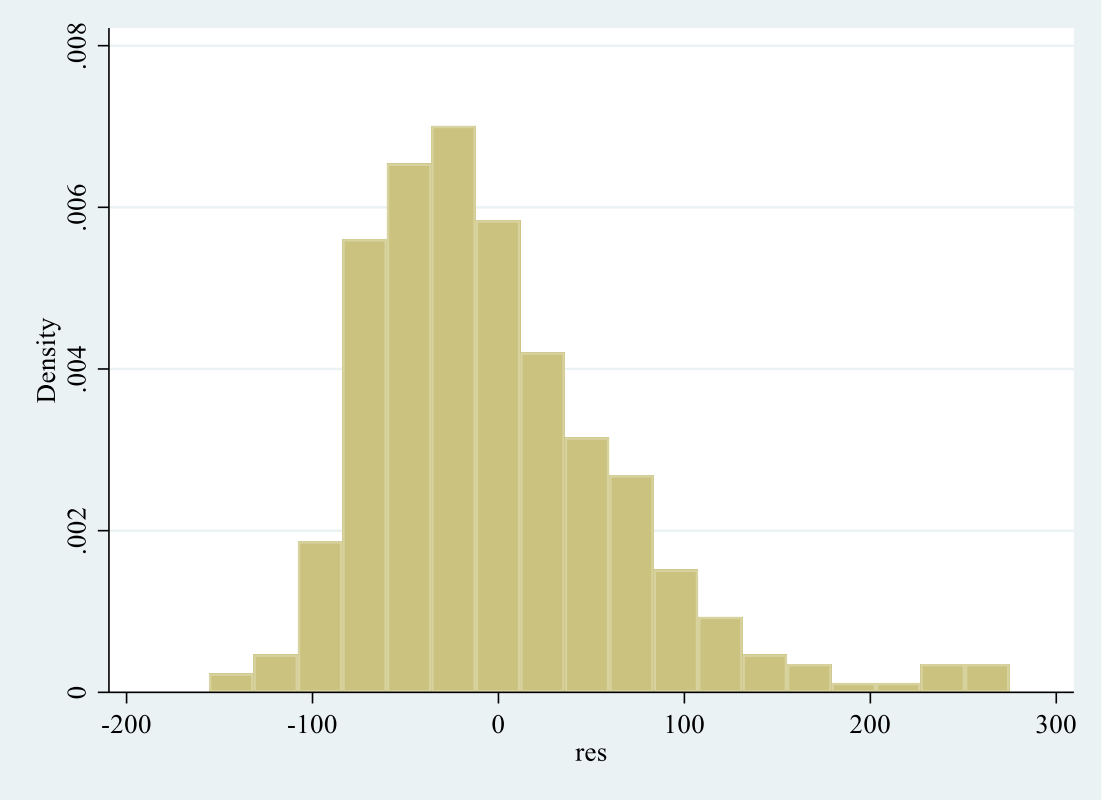


Figure 3: Distribution of residual

The observed mean of the neonatal data does not fall within the 95% confidence interval of the predicted values at age 0 using the model. Hence we decided to exclude neonatal data from the model.

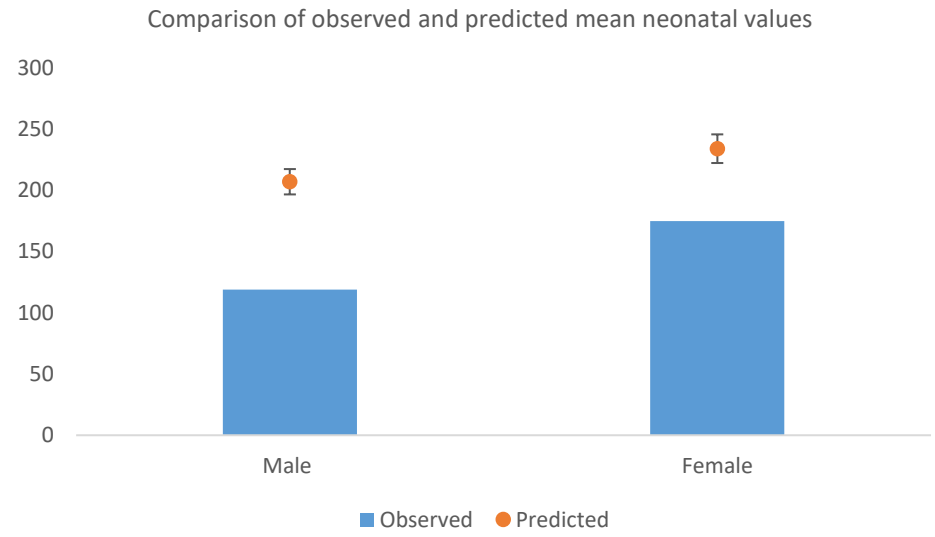


Figure 4: Comparison of observed mean for neonates and predicted mean values (95% CI) at birth by sex

A separate Stata command mfpi was used to assess the interaction between age and sex.

```
mfpi regress alp_mp age, showmodel with(sex) linear(age) fp1(age) fp2(age)

Variables in adjustment model
-----------------------------
      age: power(s) = 3 3

Interactions with sex (359 observations). Flex-1 model (least flexible)

-------------------------------------------------------------------------------
Var         Main        Interact     idf   Chi2     P     Deviance tdf   AIC
-------------------------------------------------------------------------------
age         Linear      Linear         1   13.63   0.0002  4172.338  3  4178.338
age         FP1(3)      FP1(3)         1   14.98   0.0001  4140.851  4  4148.851
age         FP2(3 3)    FP2(3 3)       2   17.78   0.0001  4124.147  7  4138.147
-------------------------------------------------------------------------------
idf = interaction degrees of freedom; tdf = total model degrees of freedom
```

Based on the initial fractional polynomial model for the mean we selected Age^3 and $Age^3 ln(Age)$ as the two power variables of age for the quantile regression models. Even though sex differences of 30 unit at birth was statistically significant in the model fitted for

the mean, we decided to exclude sex as a main effect from the models since the difference at birth is not clinically significant based on other studies. However, we included the interaction with age and sex in the model as it was both statistically and clinically significant. Hence we fitted the following models for 2.5th and 97.5th percentile using −**qreg-** command in Stata and presented the fitted values in figure 5.

```
qreg alp_mp age_1 age_2 c.age_1#i.sex c.age_2#i.sex, quantile(2.5)

Iteration  1:  WLS sum of weighted deviations =  6603.1196
Iteration  1: sum of abs. weighted deviations =  5392.0996
Iteration  2: sum of abs. weighted deviations =  5351.2363
Iteration  3: sum of abs. weighted deviations =  3176.5897
Iteration  4: sum of abs. weighted deviations =  1730.7407
Iteration  5: sum of abs. weighted deviations =  1191.3461
Iteration  6: sum of abs. weighted deviations =  1151.0573
Iteration  7: sum of abs. weighted deviations =  1089.6686
Iteration  8: sum of abs. weighted deviations =  979.60176
Iteration  9: sum of abs. weighted deviations =  946.63298
Iteration 10: sum of abs. weighted deviations =  933.72091
Iteration 11: sum of abs. weighted deviations =  925.47102
Iteration 12: sum of abs. weighted deviations =  924.44862
Iteration 13: sum of abs. weighted deviations =  923.15481
Iteration 14: sum of abs. weighted deviations =  922.37864

.025 Quantile regression                              Number of obs =        358
  Raw sum of deviations 1385.825 (about 56)
  Min sum of deviations 922.3786                      Pseudo R2     =     0.3344

------------------------------------------------------------------------------
         alp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       age_1 |  -.1173602    .101026    -1.16   0.246    -.3160486    .0813283
       age_2 |    .033487   .0361058     0.93   0.354    -.0375227    .1044966
             |
 sex#c.age_1 |
       Male  |   .0906227   .1182441     0.77   0.444    -.1419288    .3231741
             |
 sex#c.age_2 |
       Male  |  -.0288957   .0430876    -0.67   0.503    -.1136364    .0558449
             |
       _cons |   146.7276   11.97321    12.25   0.000     123.1798    170.2754
------------------------------------------------------------------------------


qreg alp_mp age_1 age_2 c.age_1#i.sex c.age_2#i.sex, quantile(97.5)

Iteration  1:  WLS sum of weighted deviations =  7997.3596
Iteration  1: sum of abs. weighted deviations =  7958.9515
Iteration  2: sum of abs. weighted deviations =    6172.32
Iteration  3: sum of abs. weighted deviations =  4632.2411
Iteration  4: sum of abs. weighted deviations =  3502.0649
Iteration  5: sum of abs. weighted deviations =  3324.7528
Iteration  6: sum of abs. weighted deviations =  2856.2231
Iteration  7: sum of abs. weighted deviations =  2437.2813
Iteration  8: sum of abs. weighted deviations =  2428.3601
Iteration  9: sum of abs. weighted deviations =  2232.8033
Iteration 10: sum of abs. weighted deviations =  2195.2005
Iteration 11: sum of abs. weighted deviations =  2099.2172
```

```
Iteration 12: sum of abs. weighted deviations =  1951.6933
Iteration 13: sum of abs. weighted deviations =  1945.7383

.975 Quantile regression                              Number of obs =        358
  Raw sum of deviations 2215.925 (about 381)
  Min sum of deviations 1945.738                      Pseudo R2     =     0.1219

------------------------------------------------------------------------------
         alp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       age_1 |    .4844109   .3342036     1.45   0.148    -.1728697    1.141691
       age_2 |   -.1816445   .1194416    -1.52   0.129    -.4165512    .0532622
             |
 sex#c.age_1 |
       Male  |    .2242867   .3911628     0.57   0.567     -.545016    .9935894
             |
 sex#c.age_2 |
       Male  |   -.0754658   .1425379    -0.53   0.597    -.3557961    .2048646
             |
       _cons |     361.003   39.60852     9.11   0.000     283.1046    438.9013
------------------------------------------------------------------------------
```

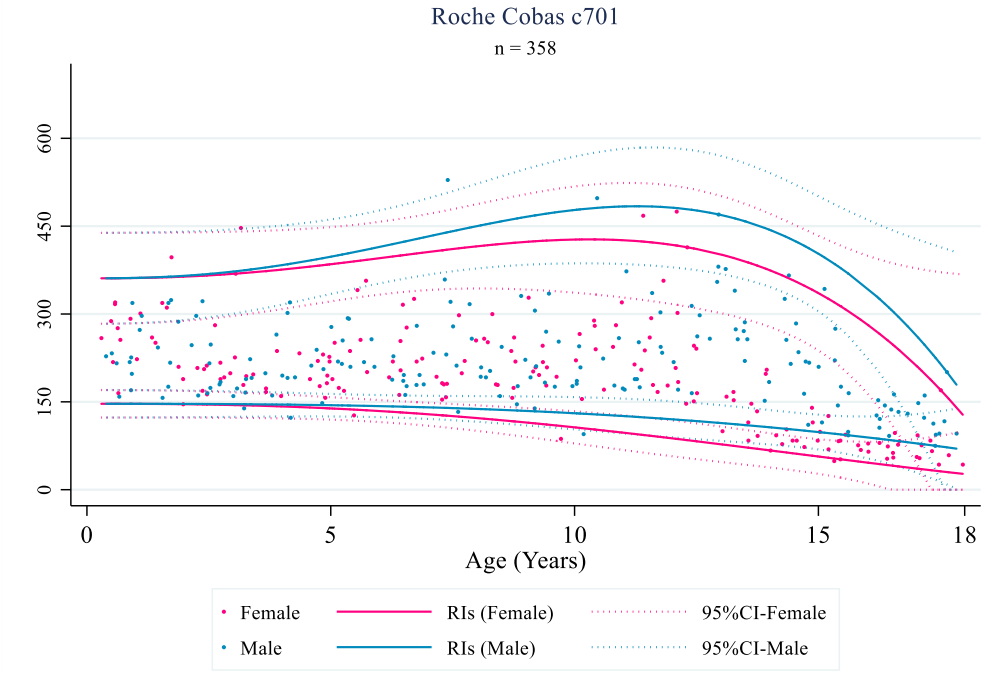


Figure 5: Reference intervals for ALP on Roche

Supplementary Document 2: Comparison of mean reference values between analysers

A. Creatinine:

Mixed-effect regression analysis was performed to investigate whether the mean reference values of creatinine vary by analysers (compared to Roche Cobas c701, Enzymatic).

Table 3: Findings of the mixed-effect regression analysis for creatinine.

creatinine	**Coefficients**	**95% Confidence intervals**		**P value**
age	2.76	2.62	2.90	0.00
Analysers				
Roche Cobas c701, enzymatic	1			
Ortho VITROS 5600,Enzymatic	-1.06	-1.83	-0.29	0.01
Roche Cobas c701, Jaffe	-3.13	-3.89	-2.36	0.00
Abbott Architect c800, Jaffe	11.67	10.91	12.44	0.00
Siemens ADVIA 1800,Jaffe	-3.80	-4.57	-3.02	0.00
Beckman Coulter AU5800, Jaffe	-6.62	-7.40	-5.83	0.00
Analysers*age				
Roche Cobas c701, enzymatic	1.00			
Ortho VITROS 5600,Enzymatic	-0.39	-0.46	-0.32	0.00
Roche Cobas c701, Jaffe	-0.05	-0.12	0.03	0.22
Abbott Architect c800, Jaffe	-1.04	-1.11	-0.97	0.00
Siemens ADVIA 1800,Jaffe	-0.21	-0.29	-0.14	0.00
Beckman Coulter AU5800, Jaffe	-0.26	-0.33	-0.18	0.00
Constant	19.94	18.47	21.41	0.00

Using the mixed-effect model, mean reference values were predicted for different analysers at specific ages (Table 2 & Figure 1).

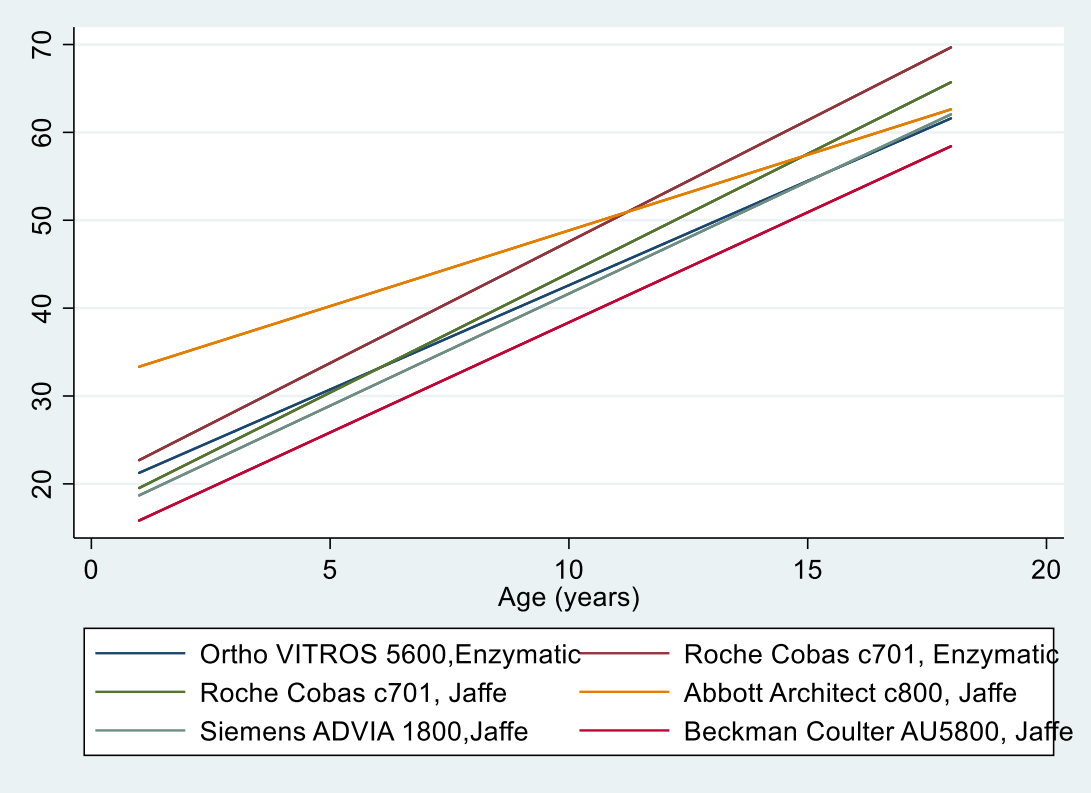


Figure 1: Age-specific mean reference values by analyser types for creatinine.

Table 4: Mean reference values for creatinine by analysers types at each interger ages

age	Ortho Vitros, Enz	Roche Cobas, Enz	Roche Cobas, Jaffe	Abbott Architect, Jaffe	Siemens ADVIA, Jaffe	Beckman Coulter, Jaffe
1	21.3	22.7	19.5	33.3	18.7	15.8
2	23.6	25.5	22.2	35.1	21.2	18.3
3	26.0	28.2	25.0	36.8	23.8	20.8
4	28.4	31.0	27.7	38.5	26.3	23.3
5	30.7	33.8	30.4	40.2	28.9	25.9
6	33.1	36.5	33.1	42.0	31.4	28.4
7	35.5	39.3	35.8	43.7	34.0	30.9
8	37.9	42.0	38.5	45.4	36.5	33.4
9	40.2	44.8	41.3	47.1	39.1	35.9
10	42.6	47.6	44.0	48.8	41.6	38.4
11	45.0	50.3	46.7	50.6	44.2	40.9
12	47.4	53.1	49.4	52.3	46.7	43.4
13	49.7	55.9	52.1	54.0	49.3	45.9
14	52.1	58.6	54.8	55.7	51.8	48.4
15	54.5	61.4	57.6	57.5	54.4	50.9
16	56.8	64.1	60.3	59.2	57.0	53.4
17	59.2	66.9	63.0	60.9	59.5	55.9
18	61.6	69.7	65.7	62.6	62.1	58.4

Subsequently coefficient of variation (CV) i.e. variations by analysers were calculated for all five analysers and between analysers for specific method at each integer age until 18 to investigate whether variation between analysers were within acceptable limit.

Table 3: Mean reference values for creatinine by analysers types for enzymatic method and coefficient of variation between analysers for enzymatic method at each interger ages

	Over all			Enzymatic method			Jaffe method		
age	**Mean**	**SD**	**CV**	**Mean**	**SD**	**CV**	**Mean**	**SD**	**CV**
1	21.9	6.1	**27.8**	22.0	1.0	**4.7**	21.7	7.8	**36.0**
2	24.3	5.8	**23.7**	24.6	1.3	**5.4**	24.1	7.4	**30.7**
3	26.8	5.5	**20.5**	27.1	1.6	**5.9**	26.5	7.0	**26.4**
4	29.2	5.2	**17.8**	29.7	1.9	**6.3**	28.9	6.6	**22.9**
5	31.6	4.9	**15.6**	32.3	2.1	**6.6**	31.3	6.2	**19.9**
6	34.1	4.7	**13.7**	34.8	2.4	**6.9**	33.6	5.8	**17.3**
7	36.5	4.4	**12.2**	37.4	2.7	**7.2**	36.0	5.4	**15.1**
8	39.0	4.2	**10.9**	40.0	3.0	**7.4**	38.4	5.1	**13.2**
9	41.4	4.0	**9.8**	42.5	3.2	**7.6**	40.8	4.7	**11.6**
10	43.8	3.9	**8.8**	45.1	3.5	**7.8**	43.2	4.4	**10.2**
11	46.3	3.7	**8.1**	47.7	3.8	**8.0**	45.6	4.1	**8.9**
12	48.7	3.6	**7.5**	50.2	4.1	**8.1**	48.0	3.8	**7.9**
13	51.2	3.6	**7.0**	52.8	4.3	**8.2**	50.3	3.5	**7.0**
14	53.6	3.6	**6.7**	55.4	4.6	**8.3**	52.7	3.3	**6.3**
15	56.0	3.6	**6.4**	57.9	4.9	**8.5**	55.1	3.1	**5.7**
16	58.5	3.6	**6.2**	60.5	5.2	**8.5**	57.5	3.0	**5.3**
17	60.9	3.7	**6.1**	63.1	5.4	**8.6**	59.9	3.0	**5.0**
18	63.3	3.9	**6.1**	65.6	5.7	**8.7**	62.3	3.0	**4.8**

Overall: Maximum CV= 8.7, Minimum CV = 4.7 and Average CV = 7.3

For enzymatic method: Maximum CV= 36.0, Minimum CV = 4.8 and Average CV = 9.3

For Jaffe method: Maximum CV= 27.8, Minimum CV = 6.1 and Average CV = 11.9

The maximum variation of 28% (or 36% for Jaffe method) at age 1 year was more than the desirable specification for allowable total error of 8.9% for creatinine. However, the maximum variation of 8.7% for enzymatic method at age 18 was less than the allowable total

error. Hence, there is enough evidence that the reference values of creatinine vary both statistically and clinically and would require analyser specific reference intervals for Jaffe method to identify children with elevated level.

B. GGT:

Similarly for GGT we initially investigated the difference in mean values by analysers and predicted the mean values at specific aged based on the model.

Table 3: Findings of the mixed-effect regression analysis for GGT.

	Coefficients	**95% Confidence intervals**		**P value**
Age	0.03	0.02	0.04	0.00
Analysers				
Ortho Vitros 5600	1			
Abbott Architect c800	0.08	0.00	0.17	0.05
Roche Cobas c701	0.07	-0.01	0.16	0.09
Siemens ADVIA 1800	0.08	0.00	0.17	0.05
Beckman Coulter AU5800	0.10	0.02	0.19	0.02
Analysers*age				
Ortho Vitros 5600	1			
Abbott Architect c800	-0.01	-0.02	0.00	0.01
Roche Cobas c701	-0.01	-0.02	0.00	0.07
Siemens ADVIA 1800	-0.01	-0.02	0.00	0.03
Beckman Coulter AU5800	-0.02	-0.02	-0.01	0.00
Constant	2.12	2.03	2.20	0.00

Using the mixed-effect model, mean reference values were predicted for different analysers at specific ages (Figure 2).

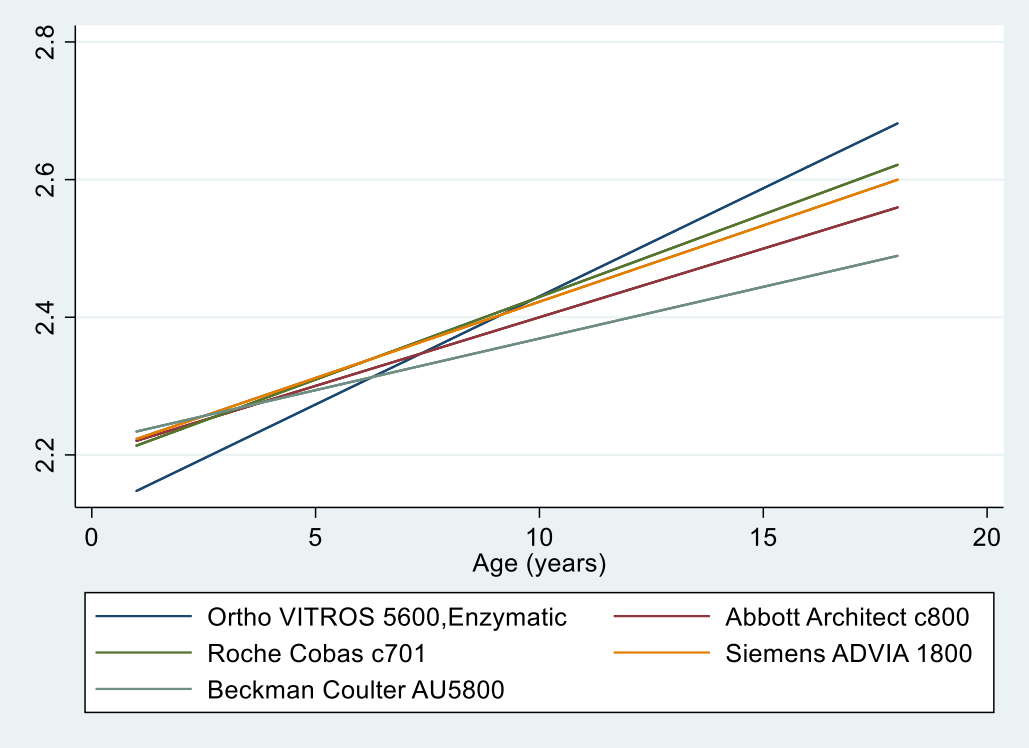


Figure 2: Age-specific mean reference values by analyser types for GGT.

Table 4: Mean reference values for GGT by analysers types and coefficient of variation between analysers at each interger ages

age	Ortho Vitros	Abbott Architect	Roche Cobas	Siemens ADVIA	Beckman Coulter	Average	Standard Deviation	Coefficient of Variation (%)
1	2.1	2.2	2.2	2.2	2.2	2.2	0.0	1.6
2	2.2	2.2	2.2	2.2	2.2	2.2	0.0	1.3
3	2.2	2.3	2.3	2.3	2.3	2.3	0.0	1.1
4	2.2	2.3	2.3	2.3	2.3	2.3	0.0	0.8
5	2.3	2.3	2.3	2.3	2.3	2.3	0.0	0.7
6	2.3	2.3	2.3	2.3	2.3	2.3	0.0	0.6
7	2.3	2.3	2.4	2.4	2.3	2.3	0.0	0.6
8	2.4	2.4	2.4	2.4	2.3	2.4	0.0	0.7
9	2.4	2.4	2.4	2.4	2.4	2.4	0.0	0.9
10	2.4	2.4	2.4	2.4	2.4	2.4	0.0	1.1
11	2.5	2.4	2.5	2.4	2.4	2.4	0.0	1.3
12	2.5	2.4	2.5	2.5	2.4	2.5	0.0	1.5
13	2.5	2.5	2.5	2.5	2.4	2.5	0.0	1.7
14	2.6	2.5	2.5	2.5	2.4	2.5	0.0	1.9
15	2.6	2.5	2.5	2.5	2.4	2.5	0.1	2.1
16	2.6	2.5	2.6	2.6	2.5	2.5	0.1	2.4
17	2.7	2.5	2.6	2.6	2.5	2.6	0.1	2.6
18	2.7	2.6	2.6	2.6	2.5	2.6	0.1	2.8

Maximum CV = 2.8, minimum CV = 0.6, mean CV = 1.4

The coefficient of variation between analyser is mostly consistent across age and maximum coefficient of variation is less than desirable specification for allowable total error of 22% for GGT.

Supplementary Document 3: Comparison of values between two methods for Creatinine, AST, ALT and Albumin

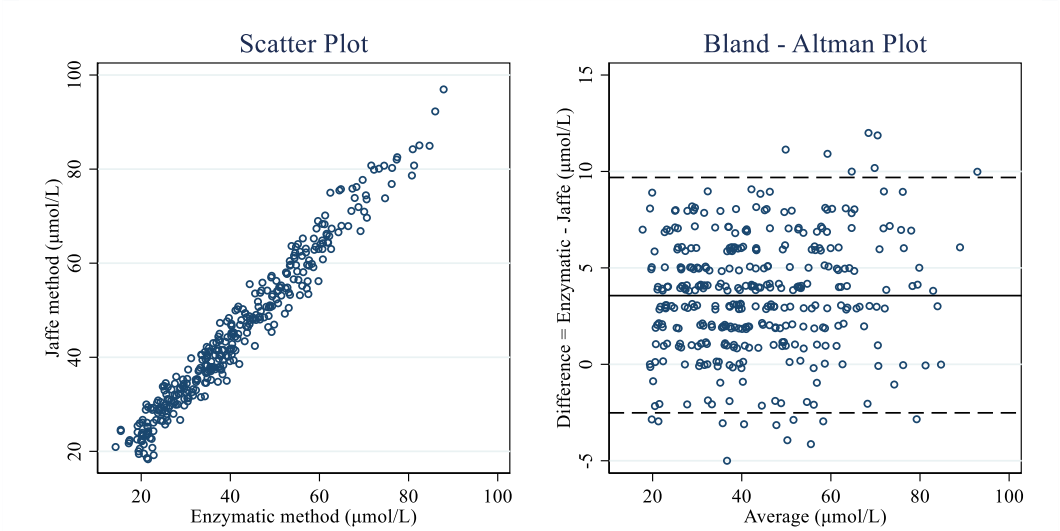


Figure 1: (Left plot) Creatinine measurement based on Jaffe versus enzymatic method on Roche Cobas and (right plot) Bland-Altman Plot.

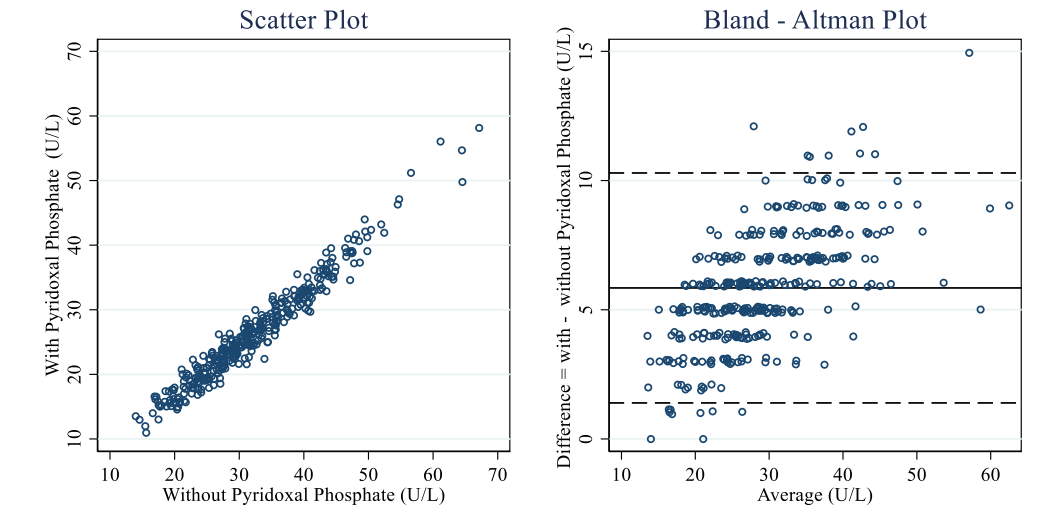


Figure 2: (Left plot) AST measurement based on TRIS Buffer - Aspartate Approx. 240 mmolar with Pyridoxal-5-Phosphate versus without Pyridoxal-5-Phosphate and (right plot) Bland-Altman Plot.

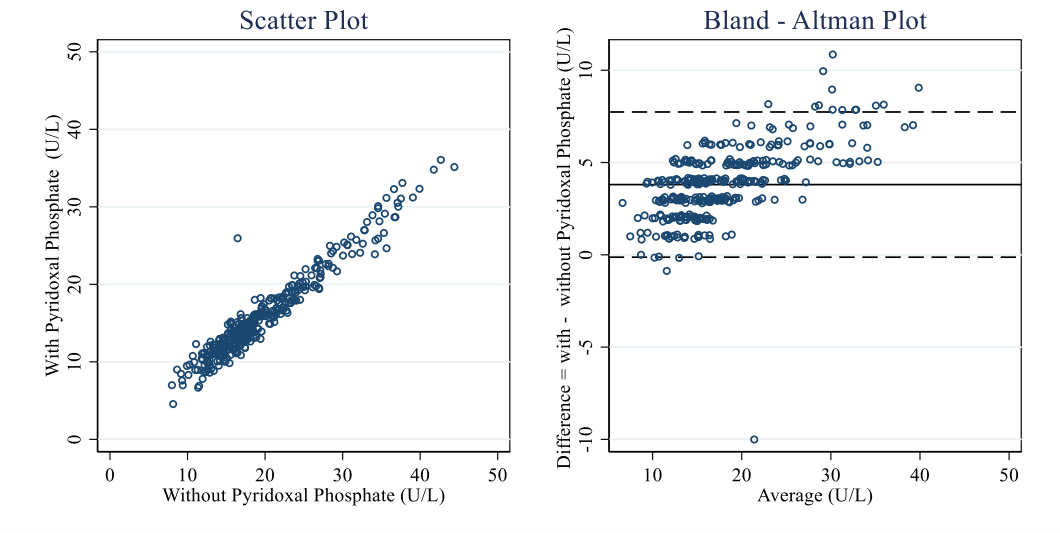


Figure 3: (Left plot) ALT measurement based on TRIS Buffer - Alanine approx. 500 mmolar with Pyridoxal-5-Phosphate versus without Pyridoxal-5-Phosphate and (right plot) Bland-Altman Plot.

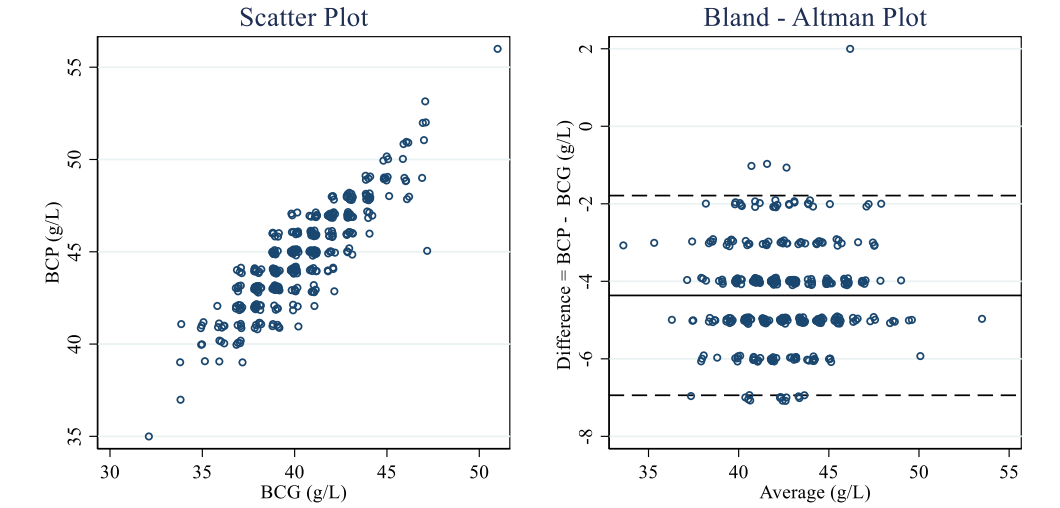
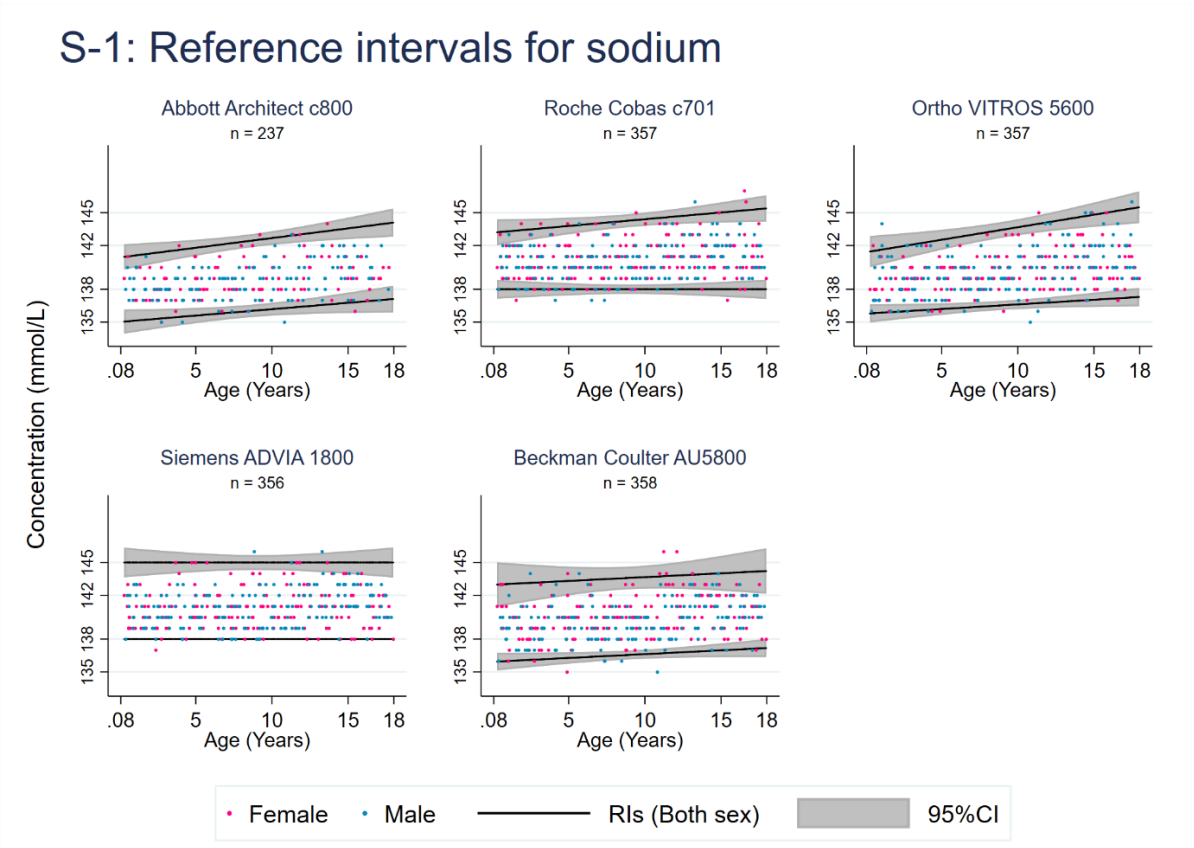
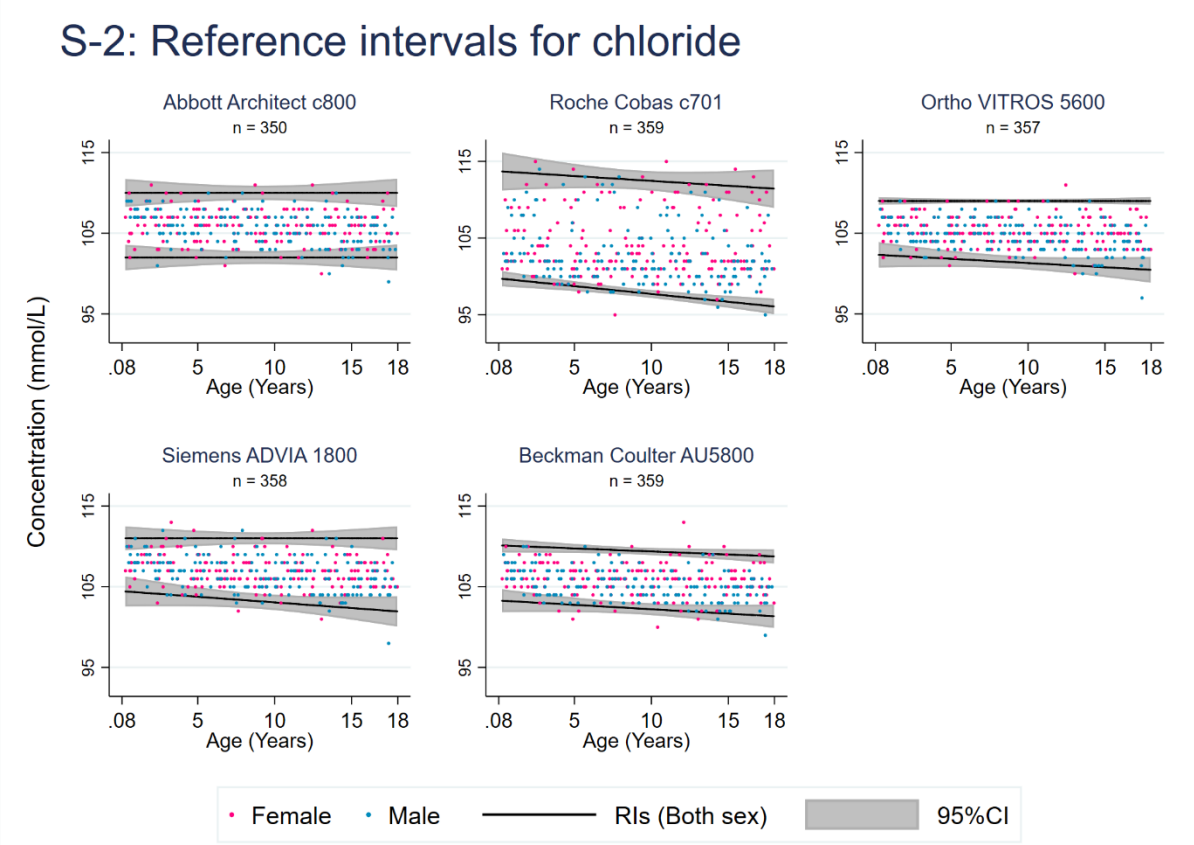


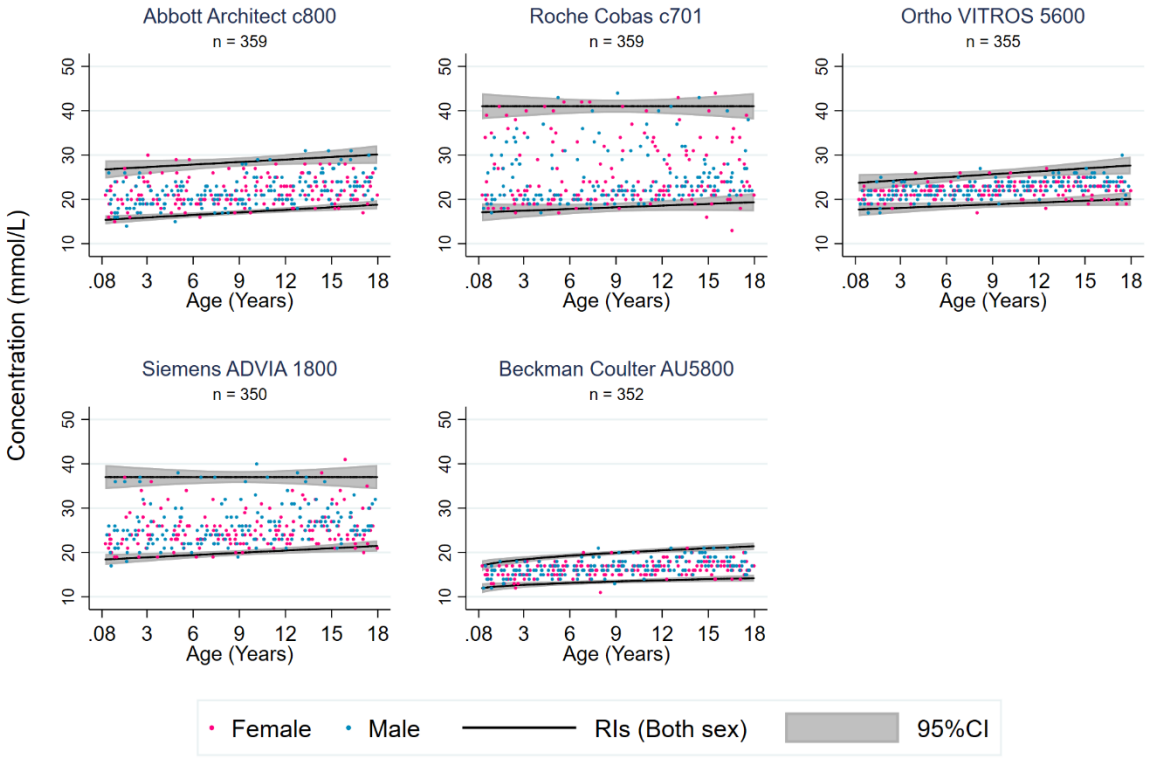
Figure 4: (Left plot) Albumin measurement based on Bromocresol Purple versus Bromocresol Green - Citrate Buffer phosphate and (right plot) Bland-Altman Plot.

Supplementary Figure 1 to 30: Age-specific RIs for 30 analytes on five analysers

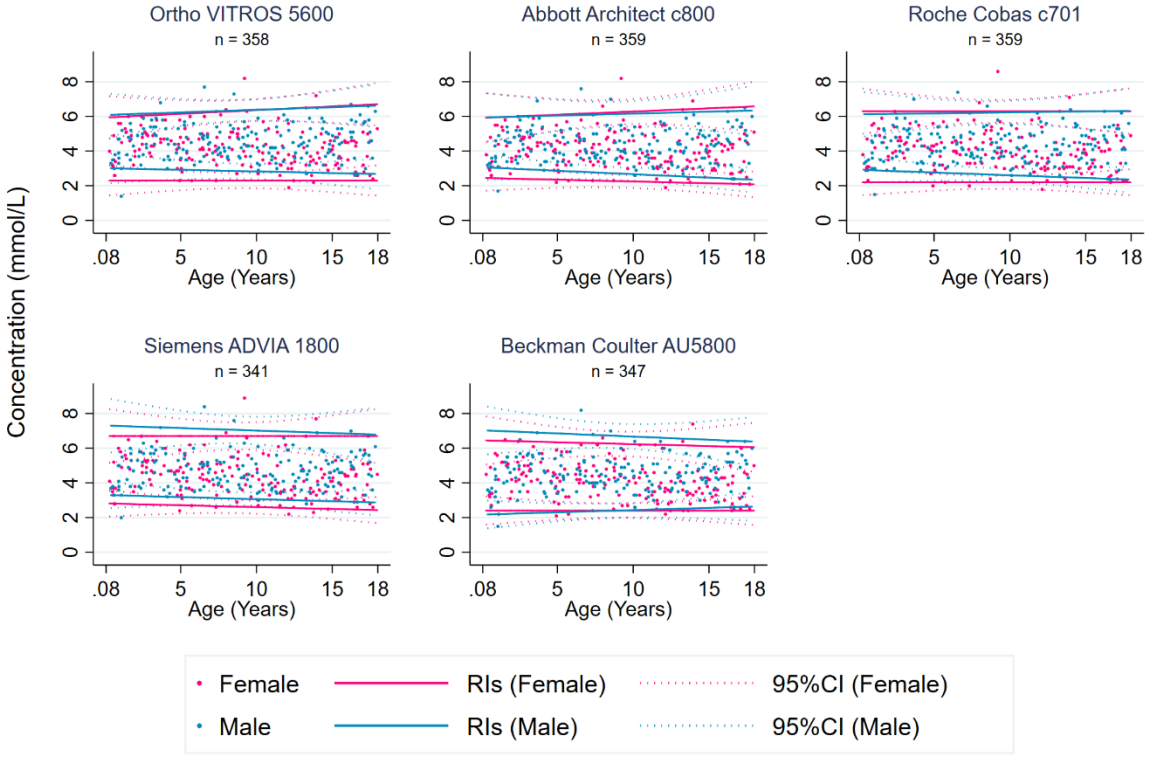




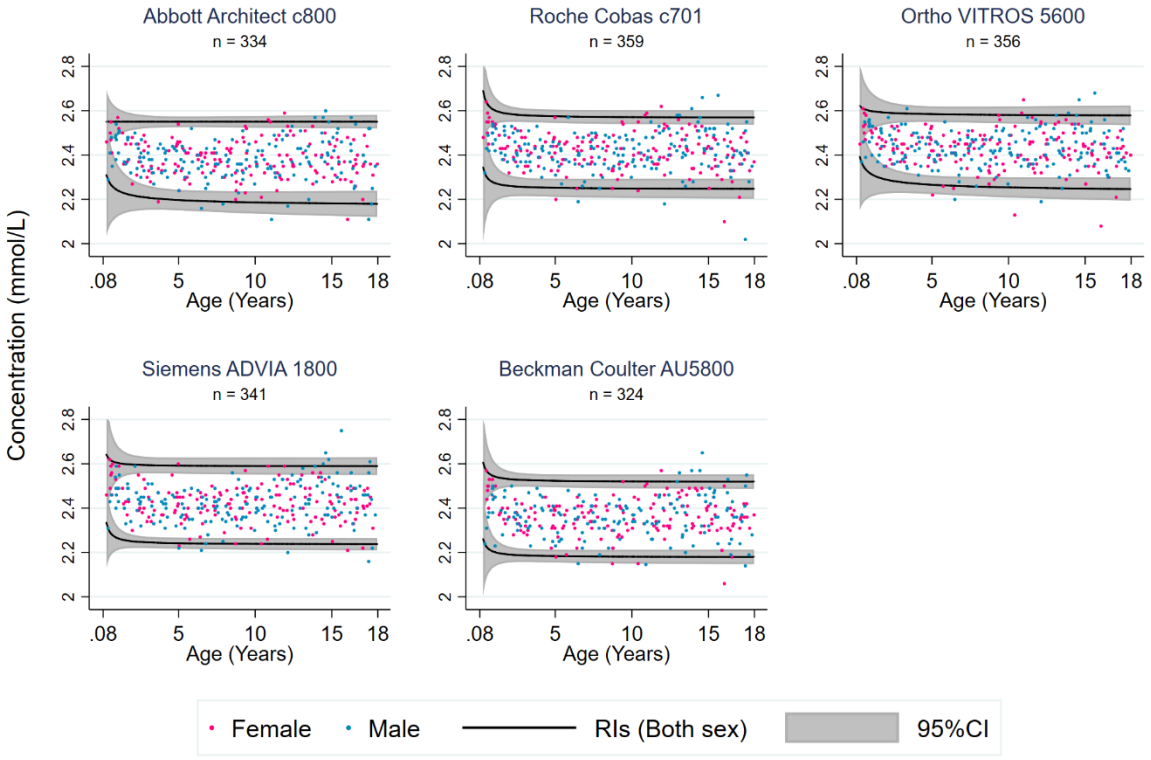
S-3: Reference intervals for bicarbonate



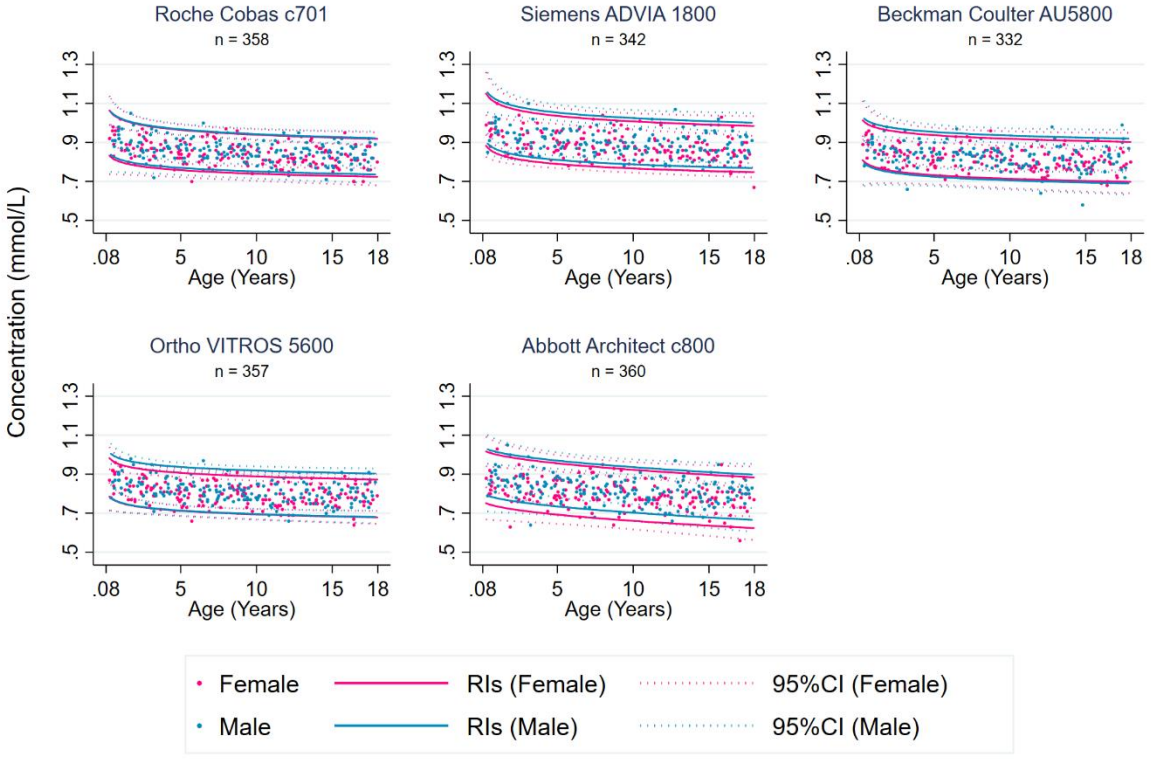
S-4: Reference intervals for urea



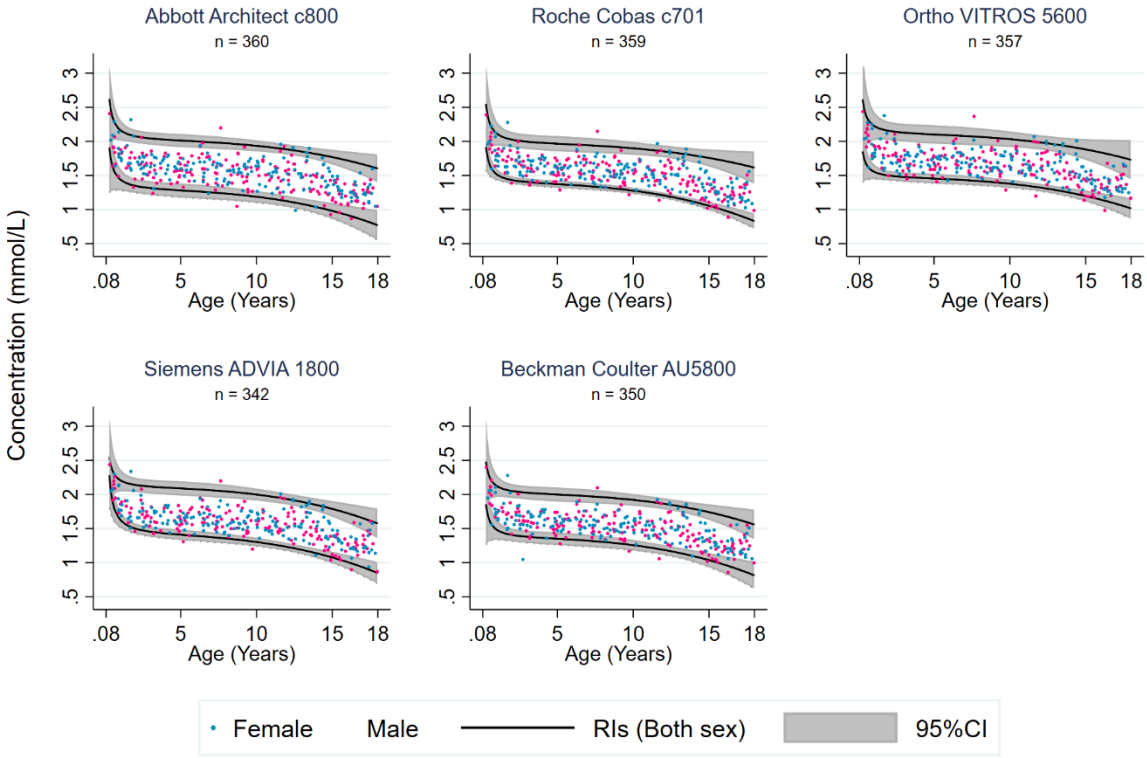
S-5: Reference intervals for calcium



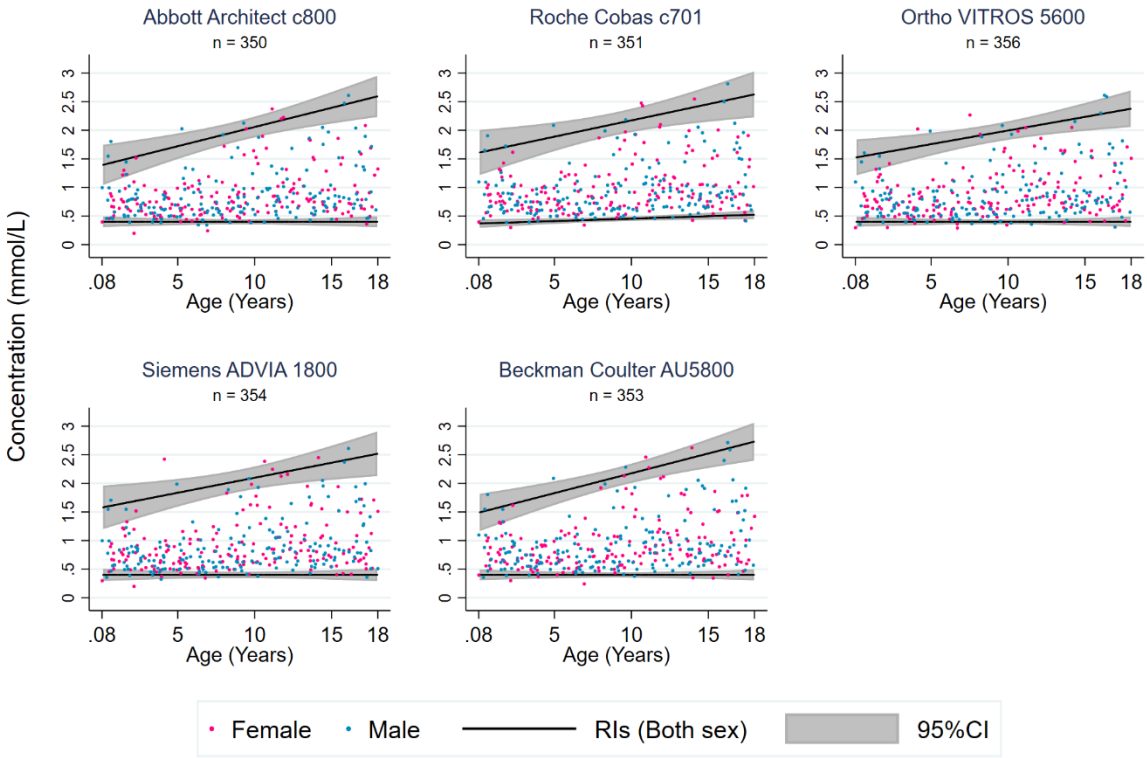
S-6: Reference intervals for magnesium



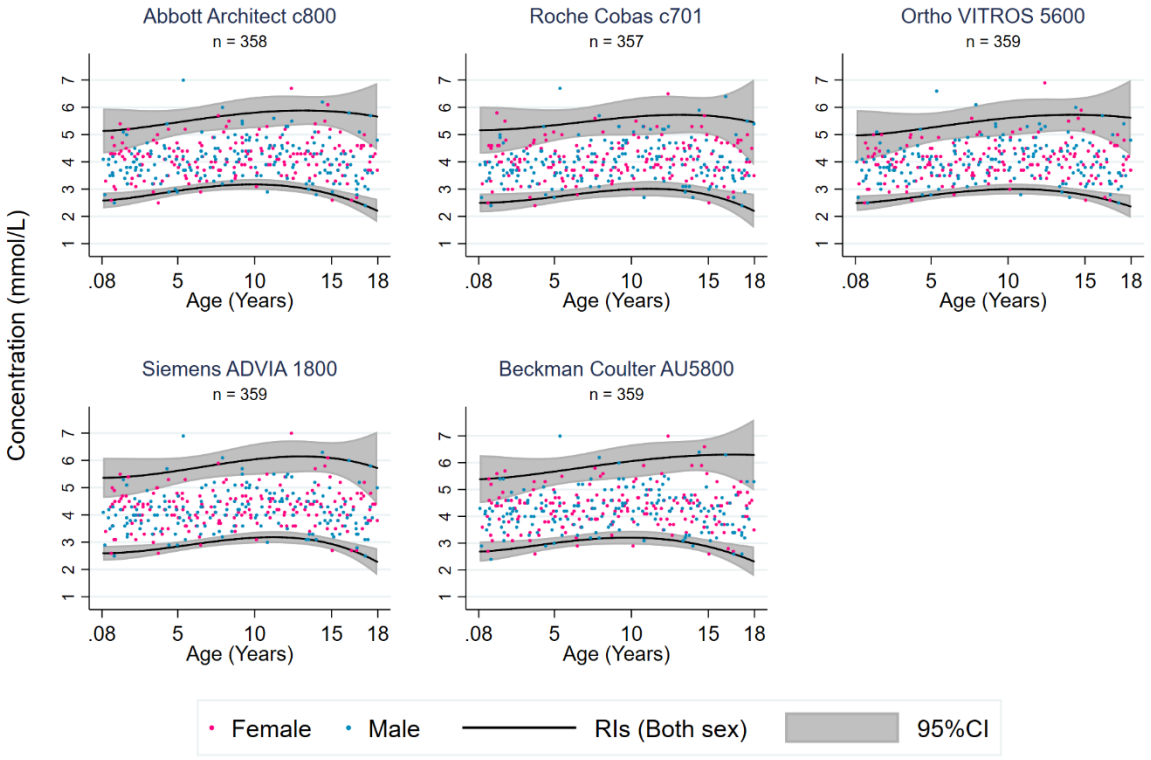
S-7: Reference intervals for phosphate*



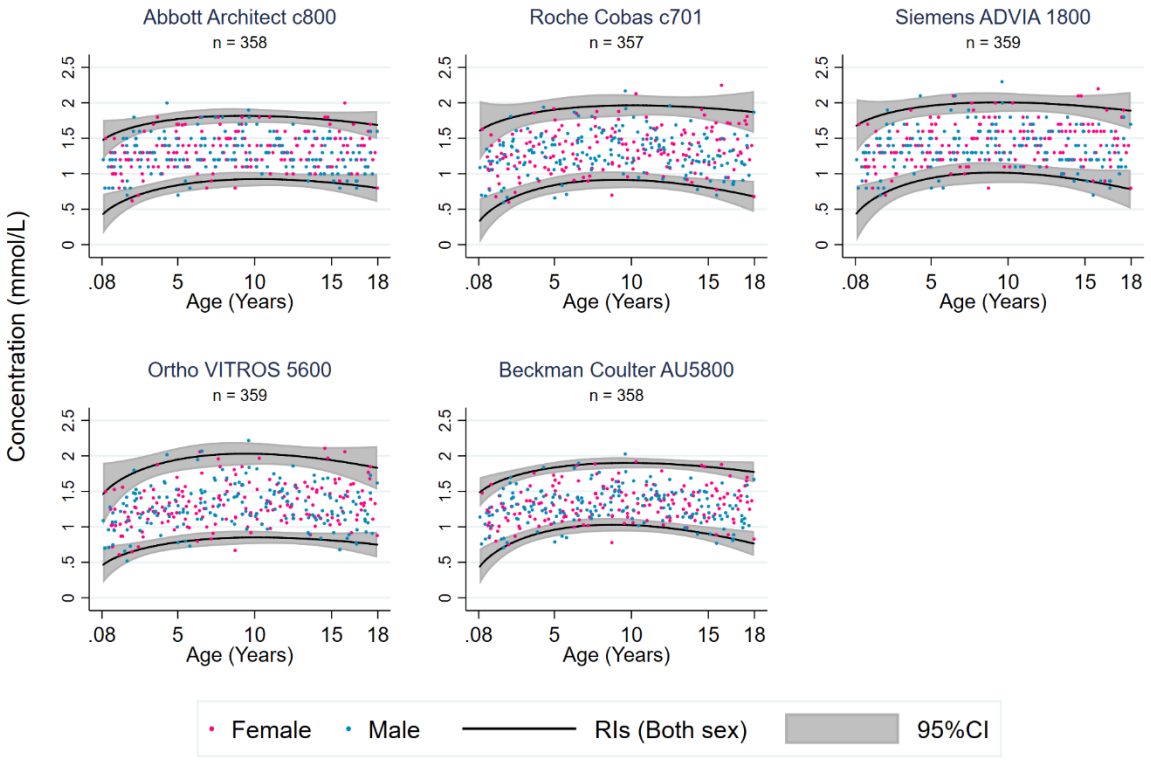
S-8: Reference intervals for triglyceride



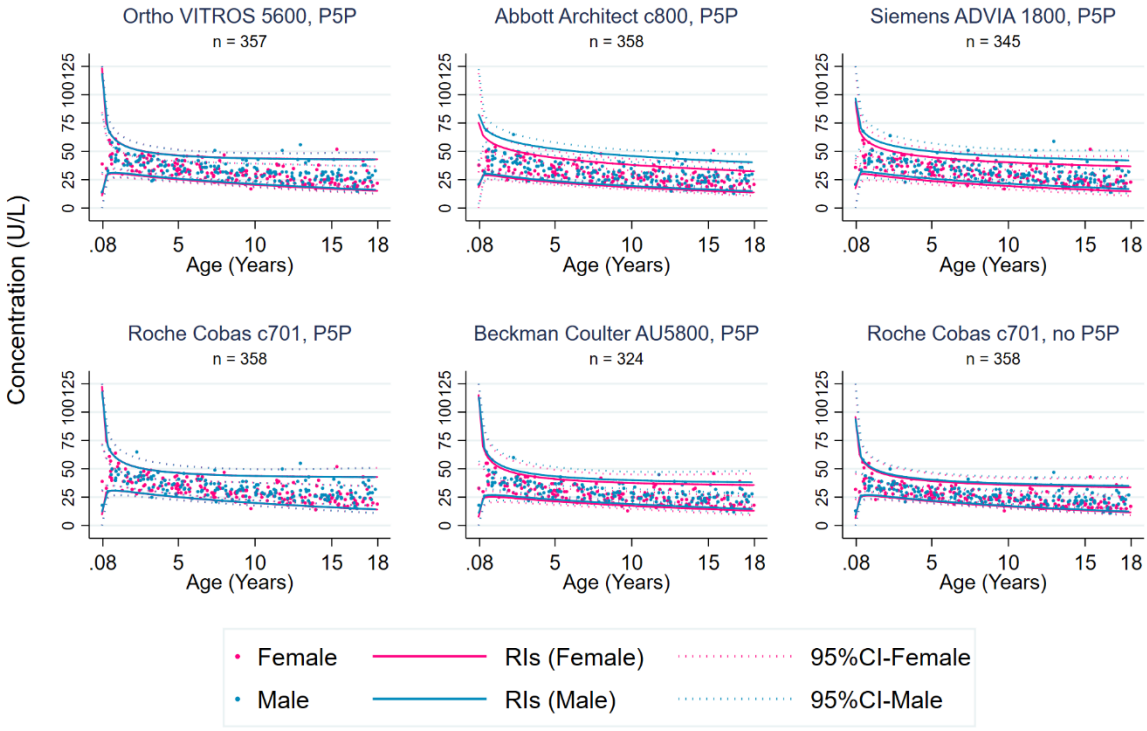
S-9: Reference intervals for cholesterol



S-10: Reference intervals for HDL cholesterol

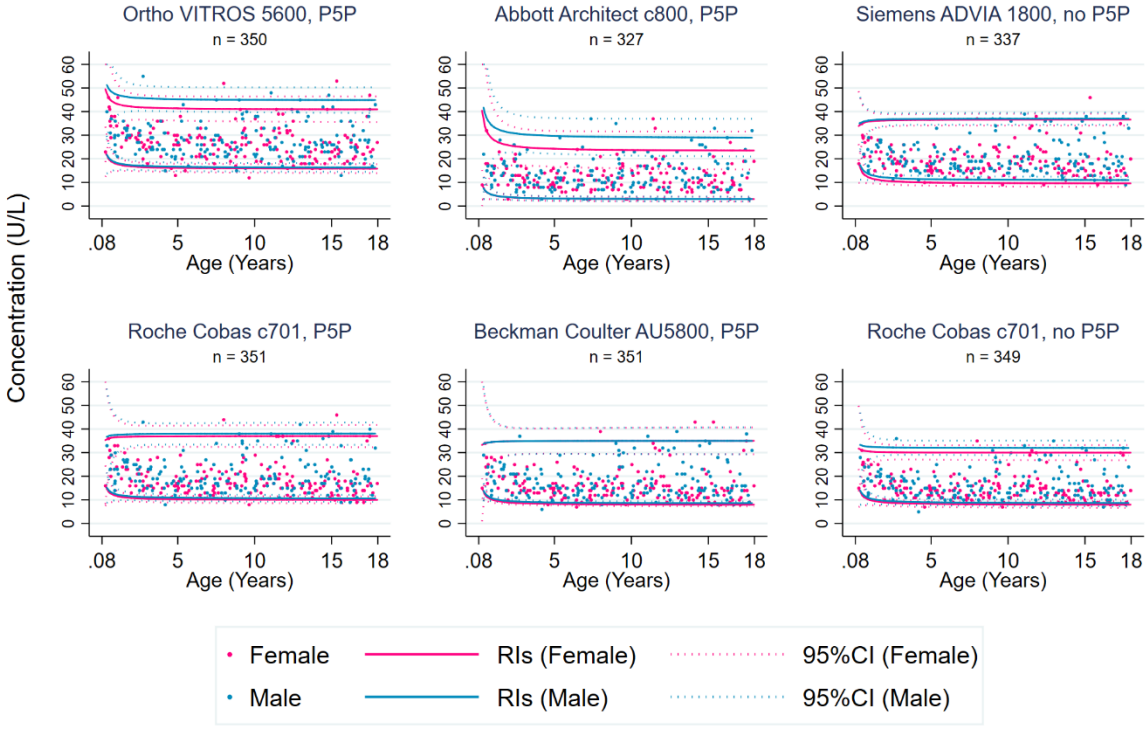


S-11: Reference intervals for aspartate amino transferase



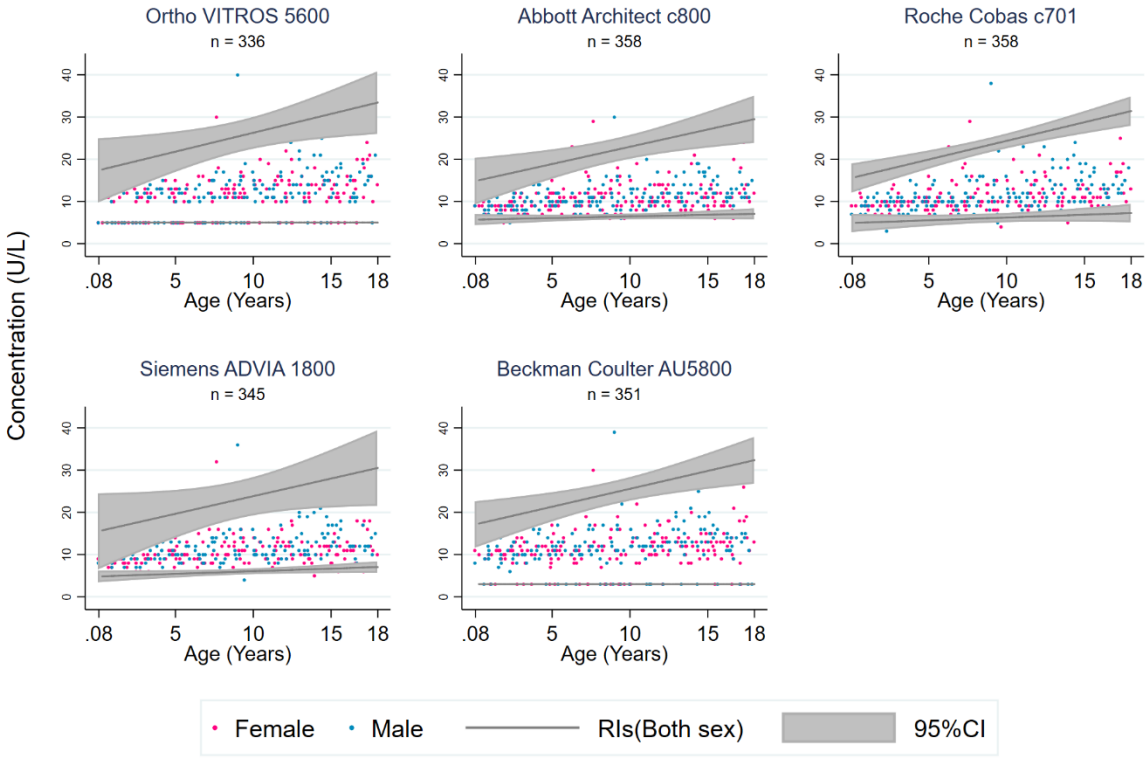
P5P stands for Pyridoxal-5-Phosphate.

S-12: Reference intervals for alanine amino transferase

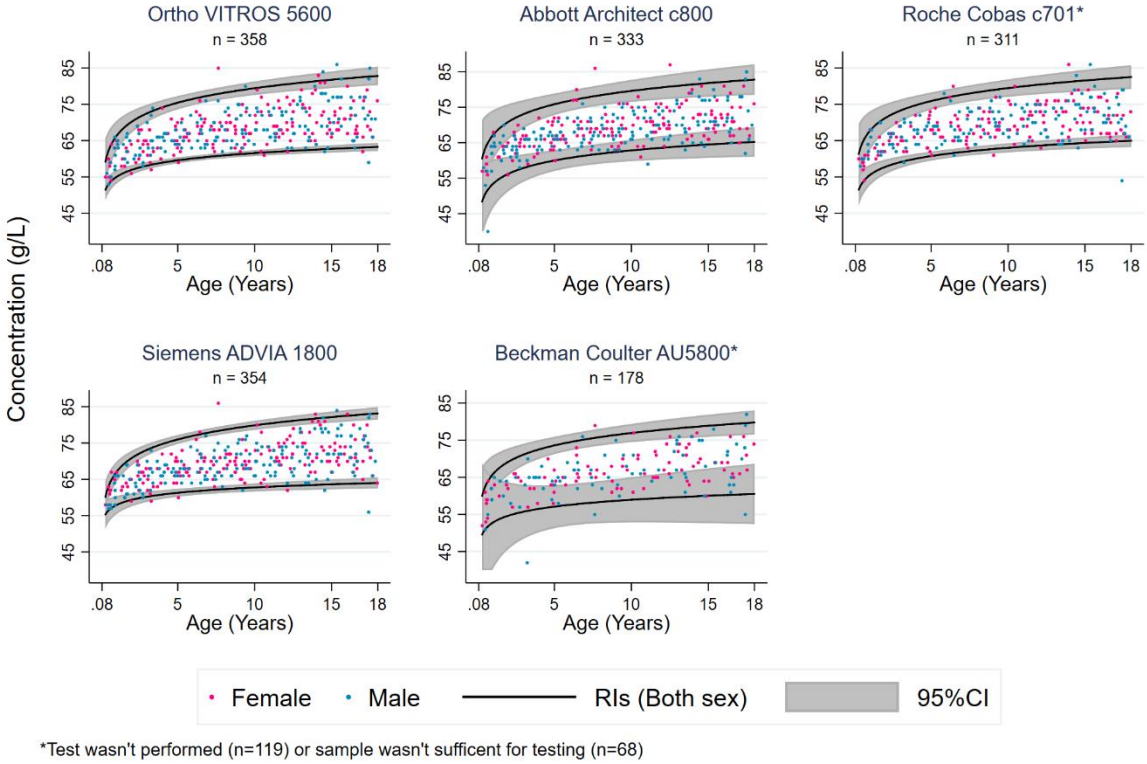


P5P stand for Pyridoxal-5-Phosphate

S-13: Reference intervals for gamma-glutamyl transferase

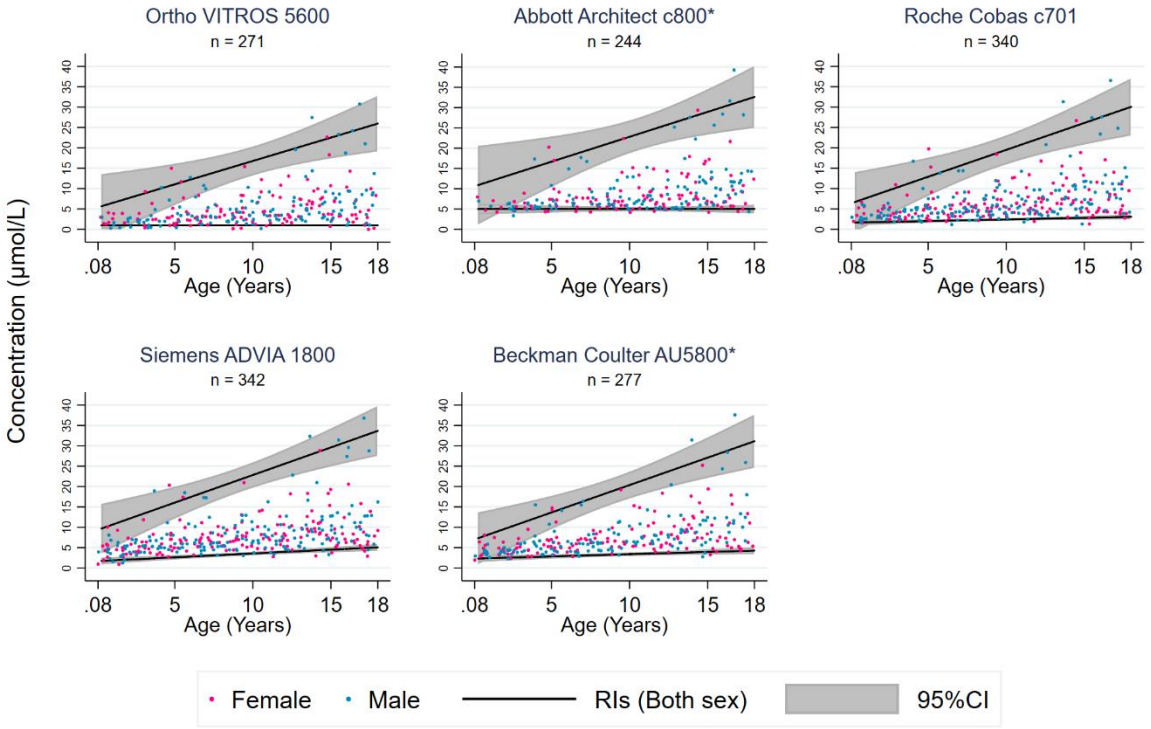


S-14: Reference intervals for total protein



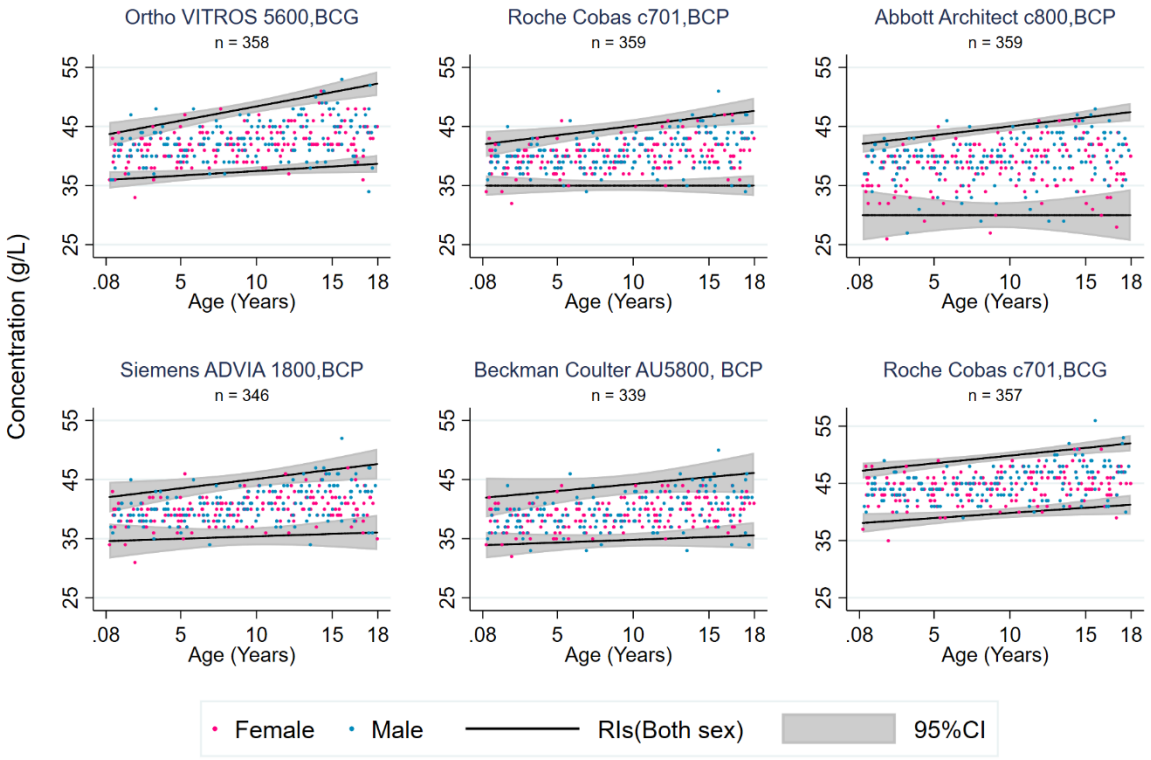
*Test wasn't performed (n=119) or sample wasn't sufficent for testing (n=68)

S-15: Reference intervals for total bilirubin

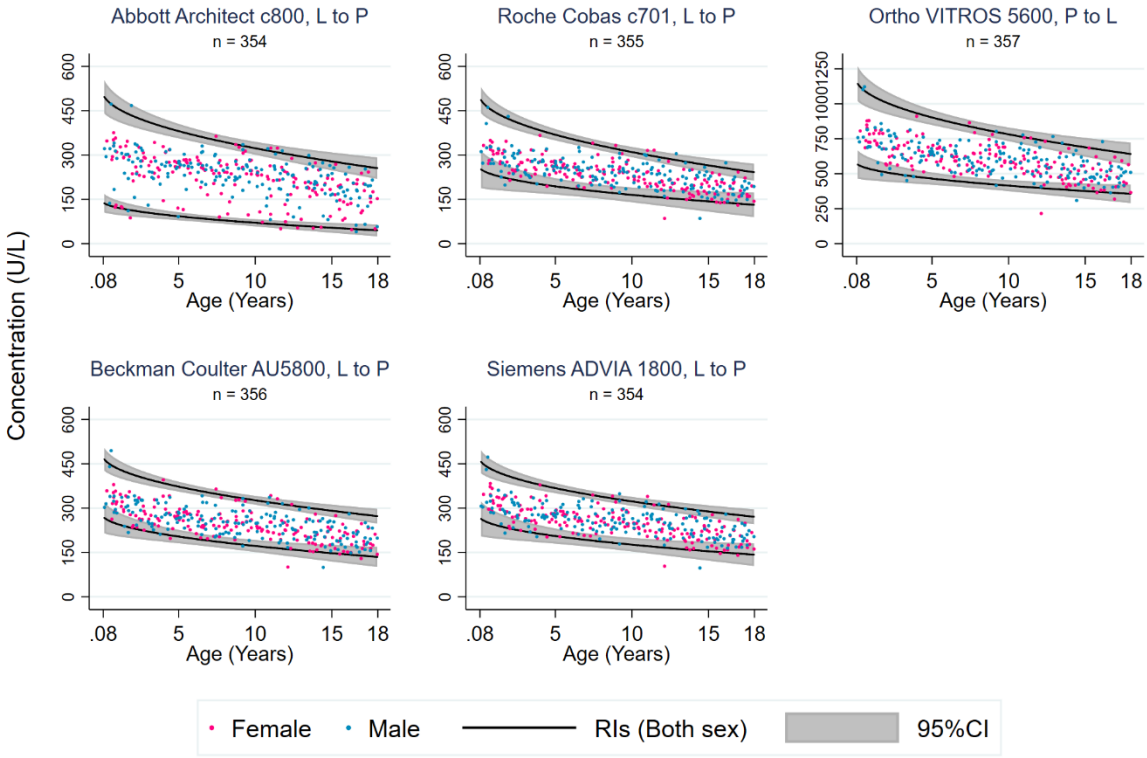


*Sample was not sufficient for testing (n = 119 on Abbott and 67 on Beckman)

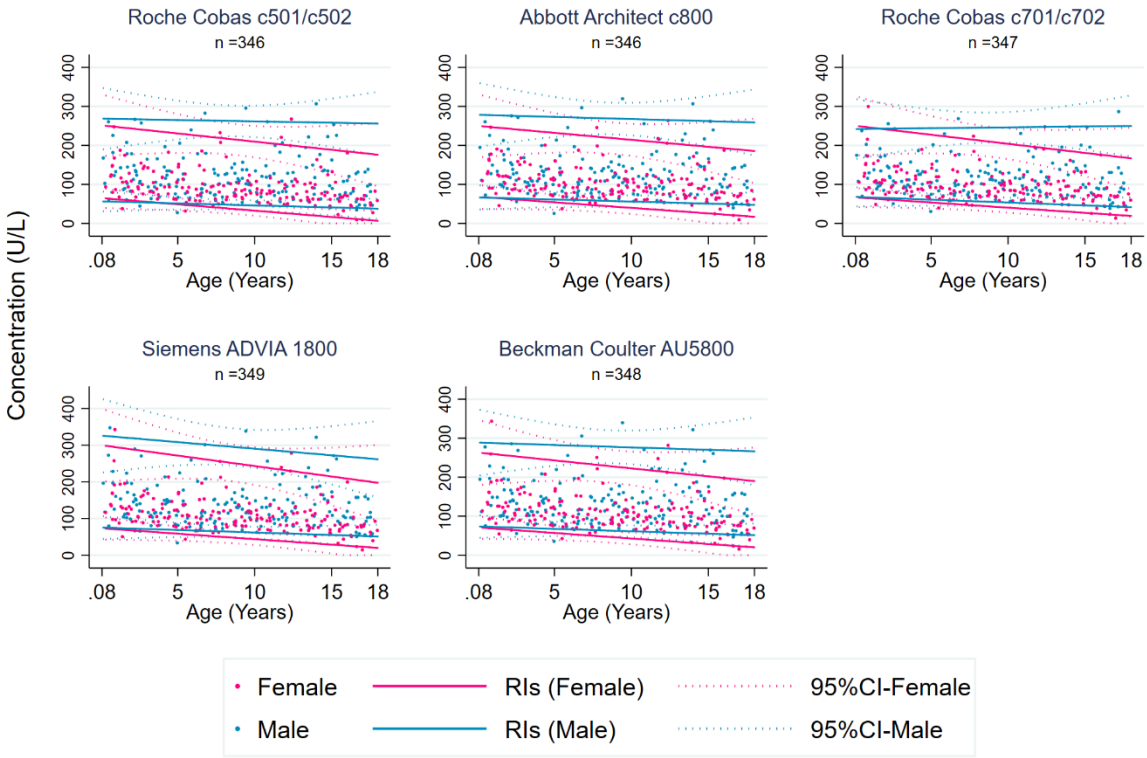
S-16: Reference intervals for albumin



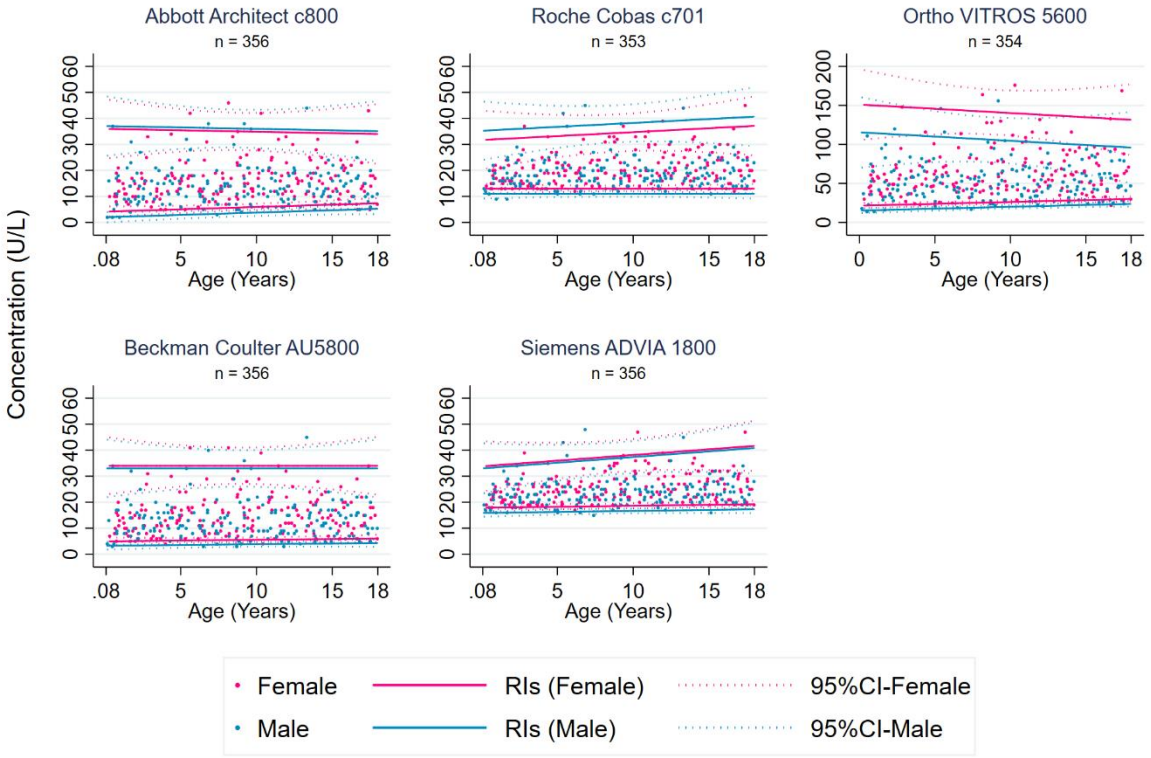
S-17: Reference intervals for lactate dehydrogenase



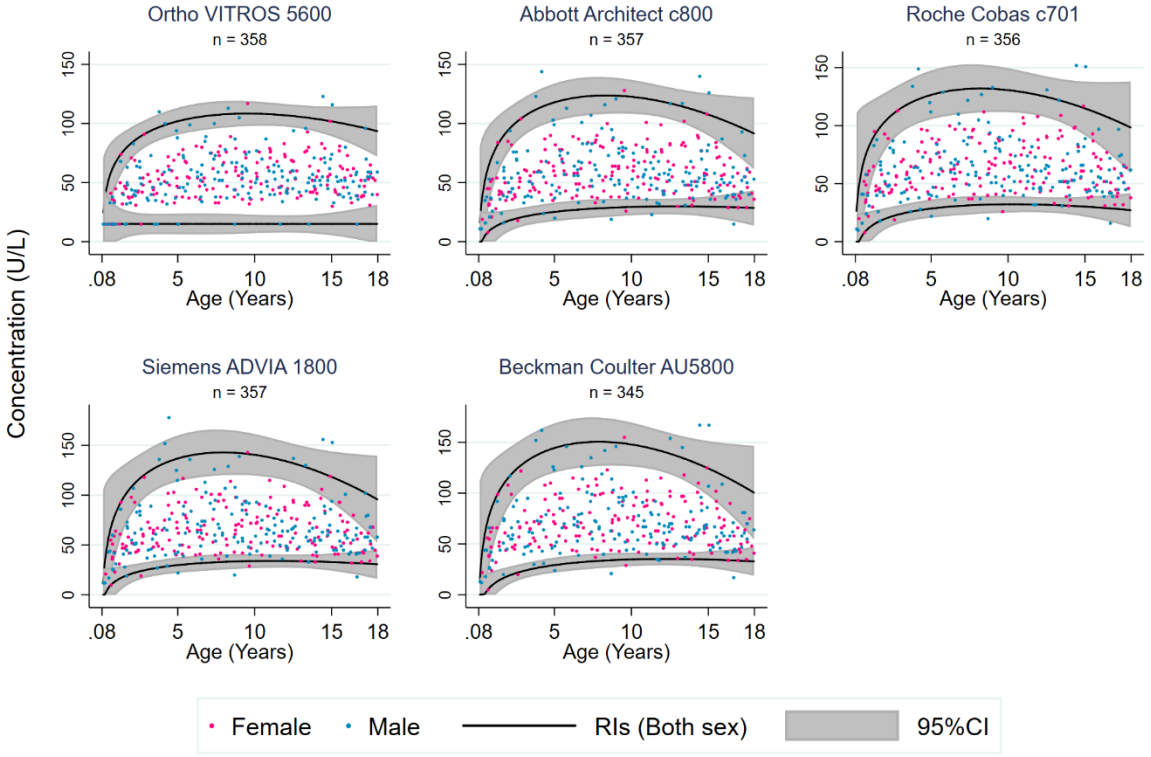
S-18: Reference intervals for creatine kinase



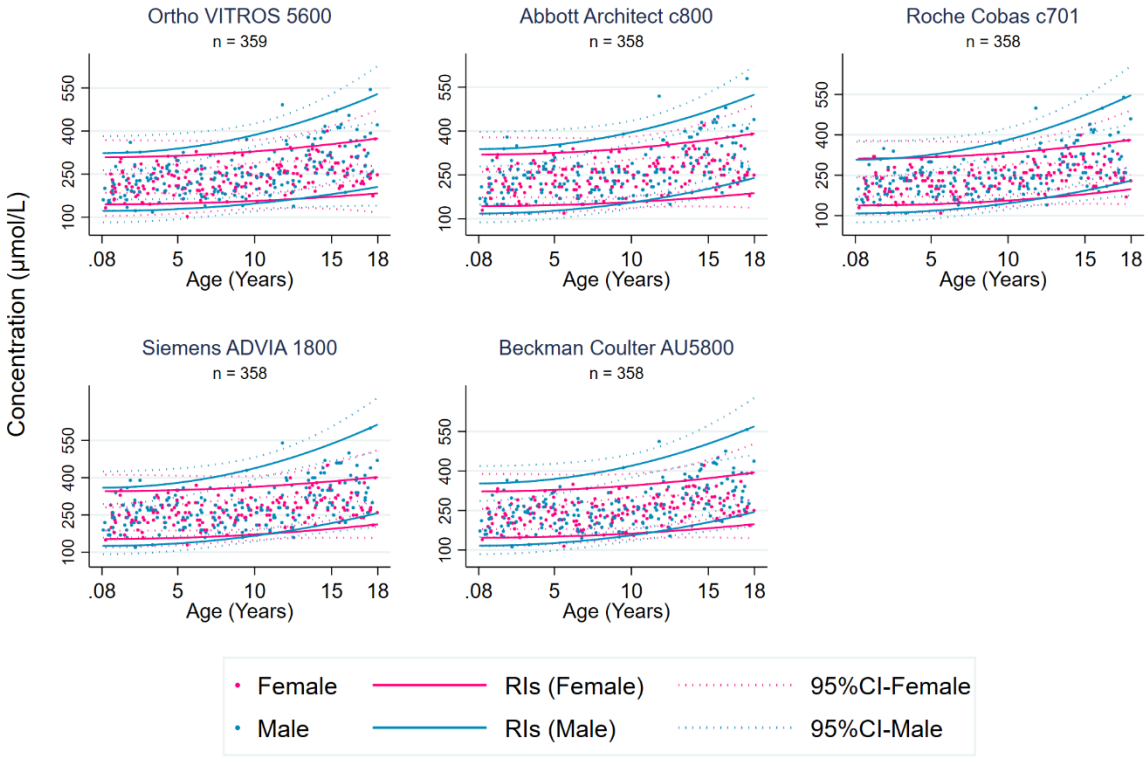
S-19: Reference intervals for lipase



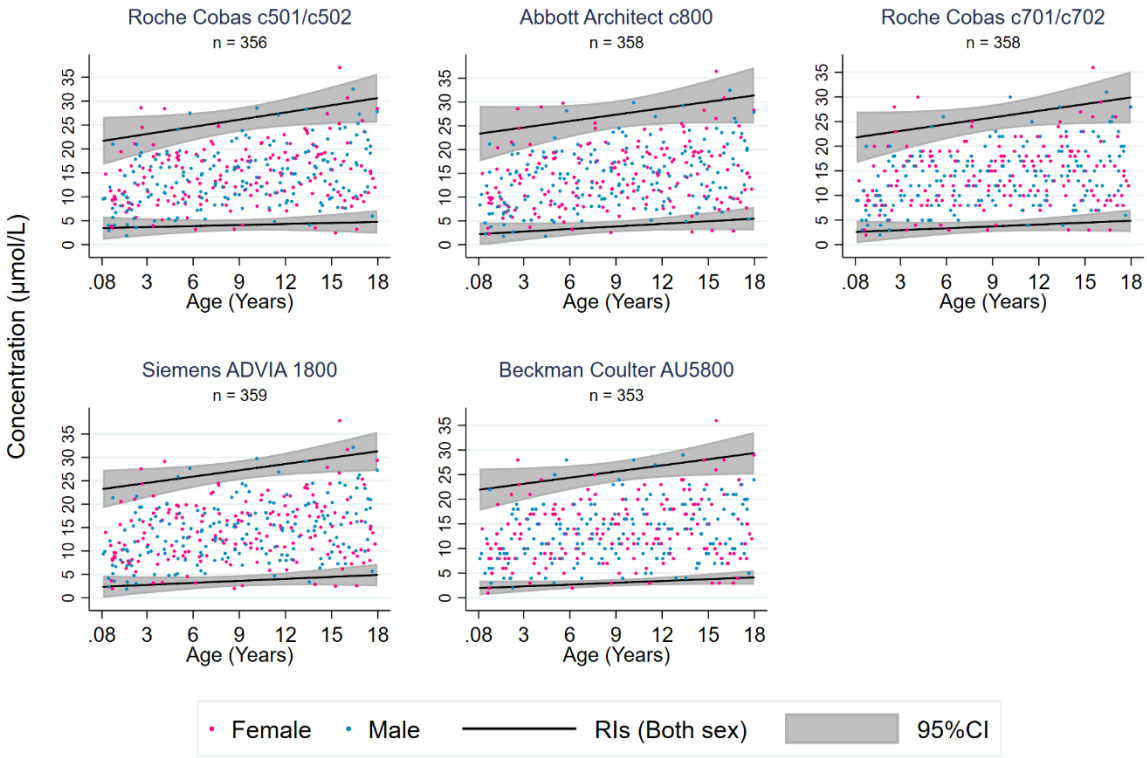
S-20: Reference intervals for amylase



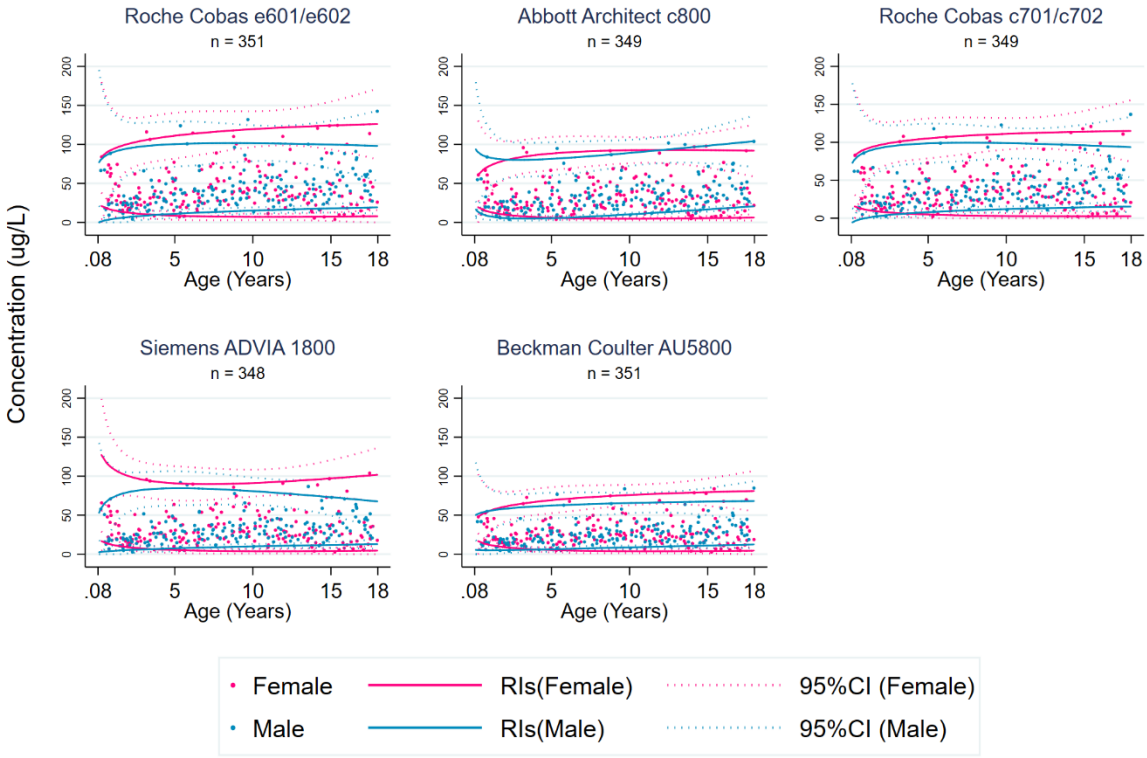
S-21: Reference intervals for uric acid



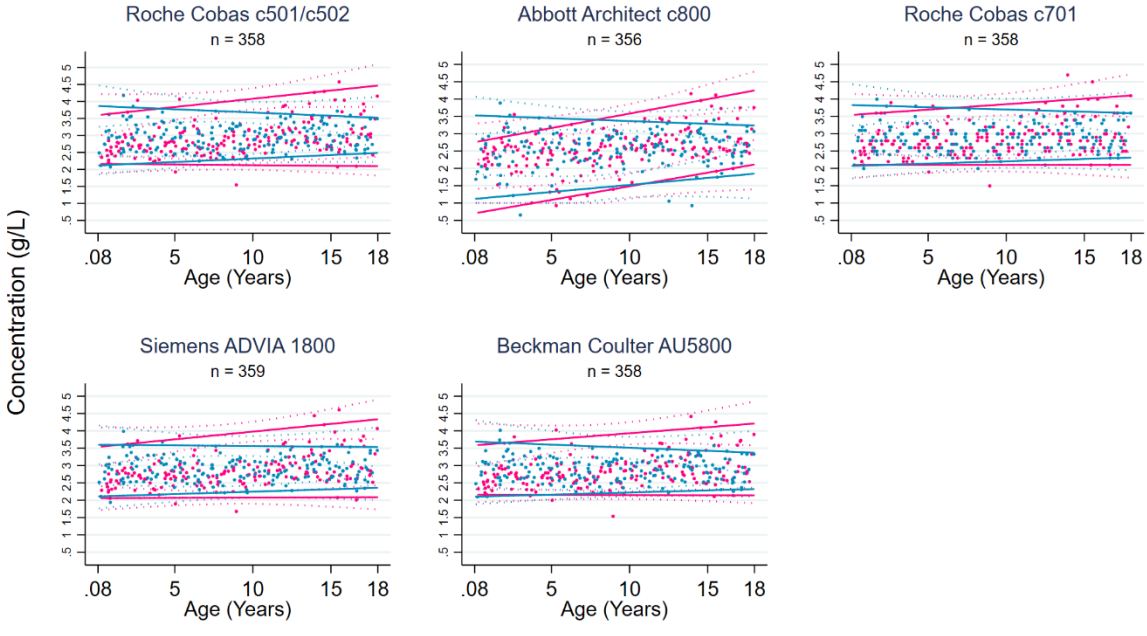
S-22: Reference intervals for iron



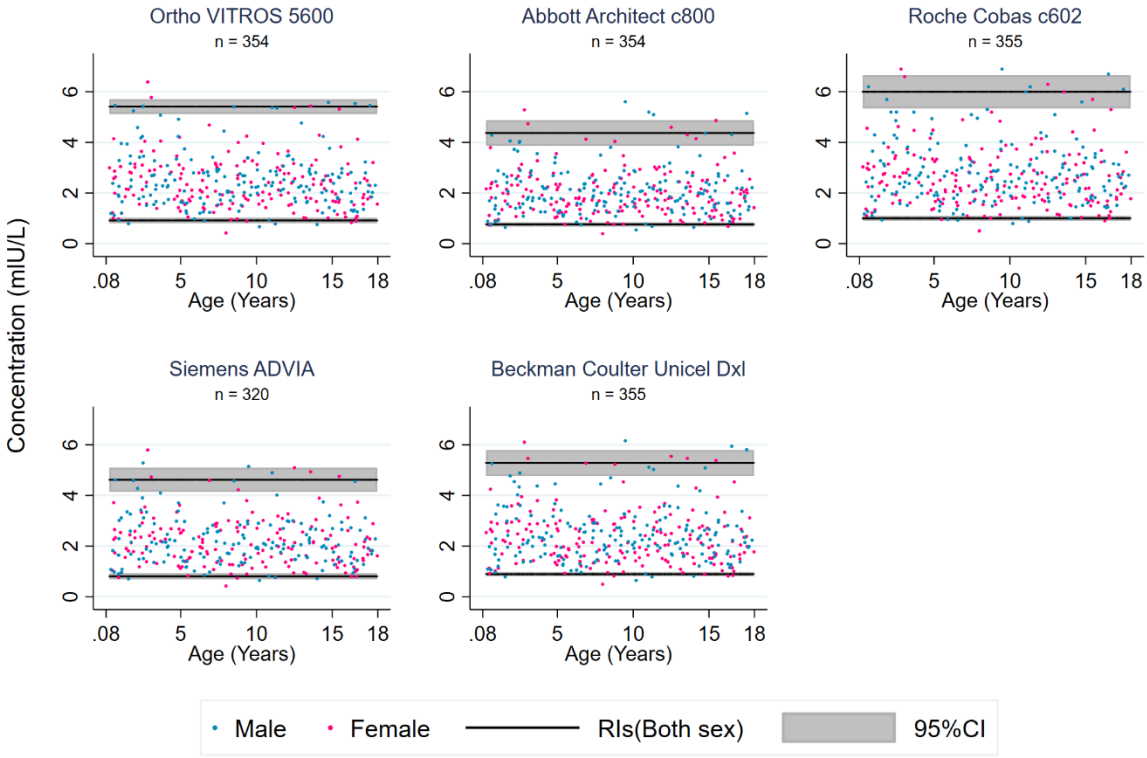
S-23: Reference intervals for ferritin



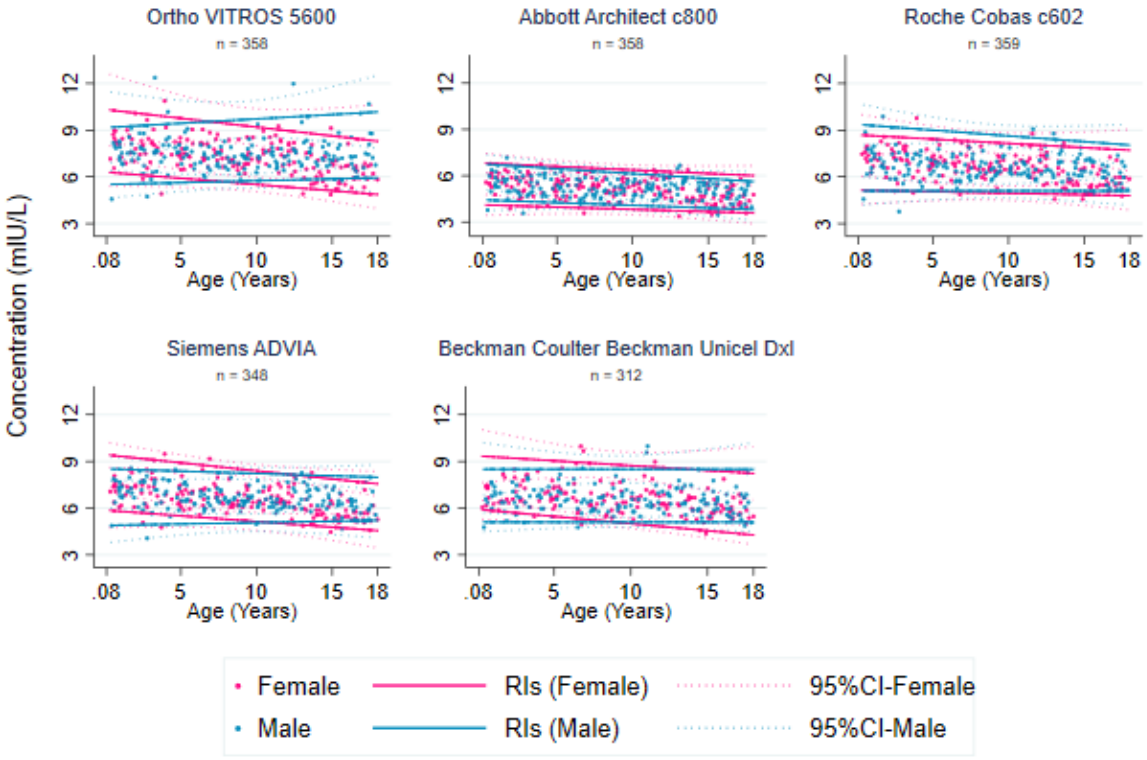
S-24: Reference intervals for transferrin



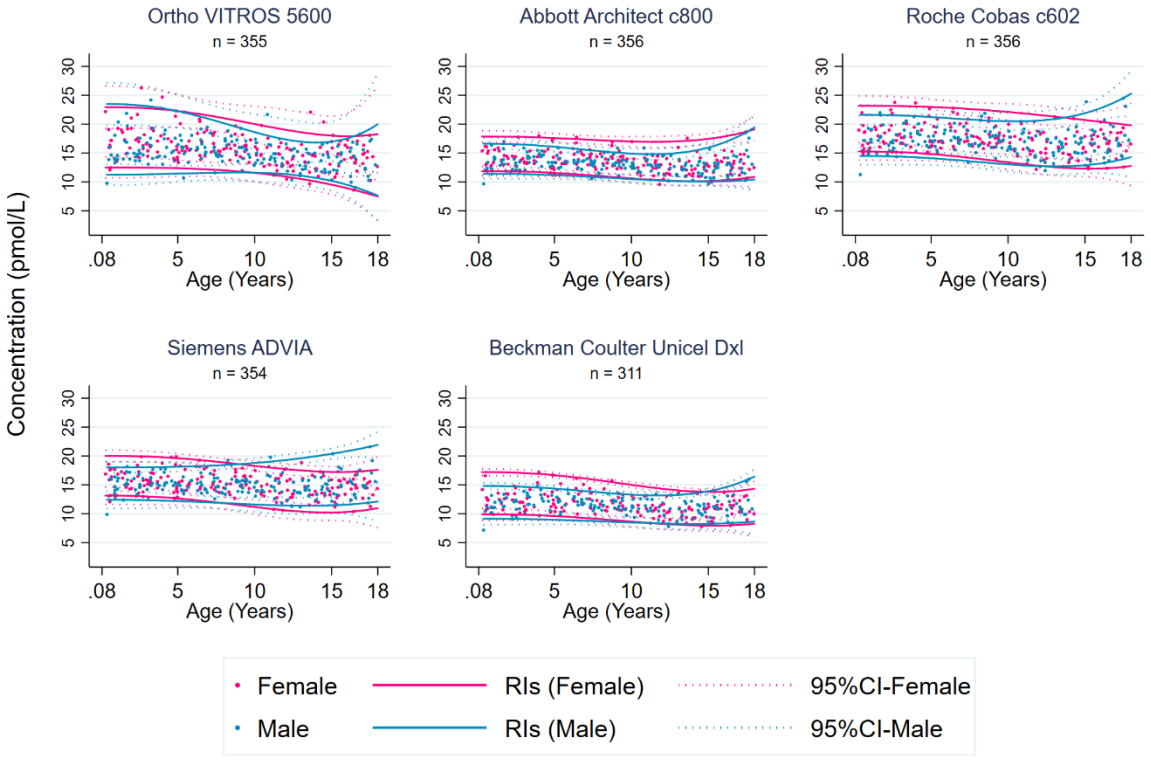
S-25: Reference intervals for thyroid stimulating hormone



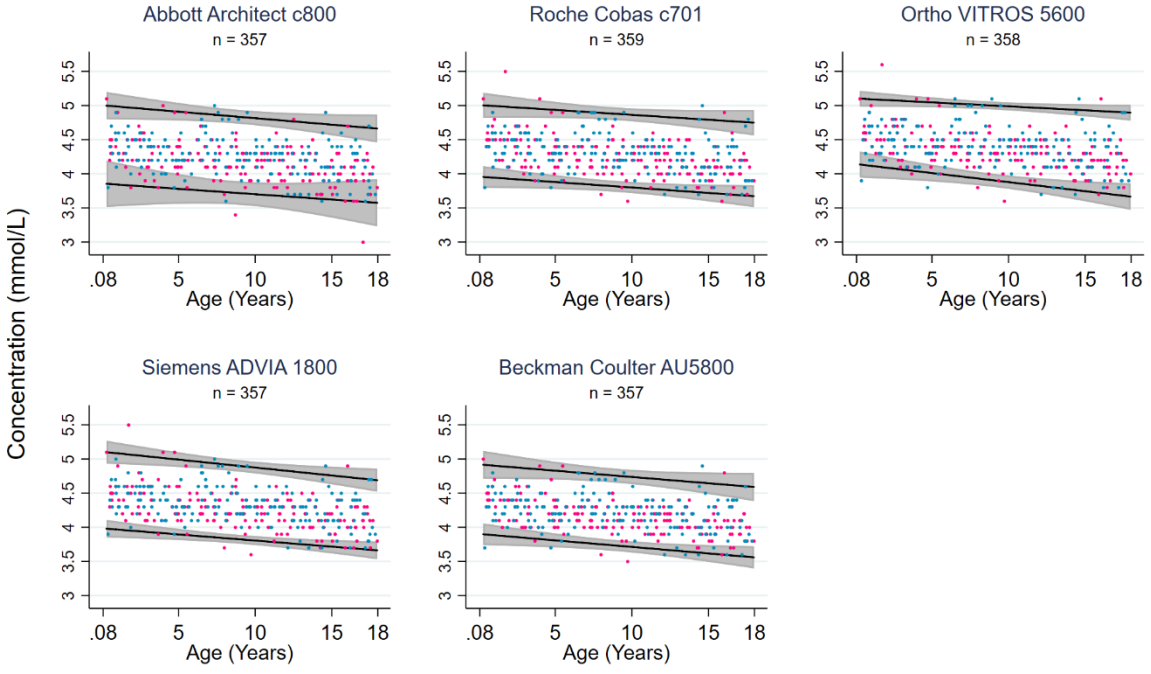
S-26: Reference intervals for free triiodothyronine



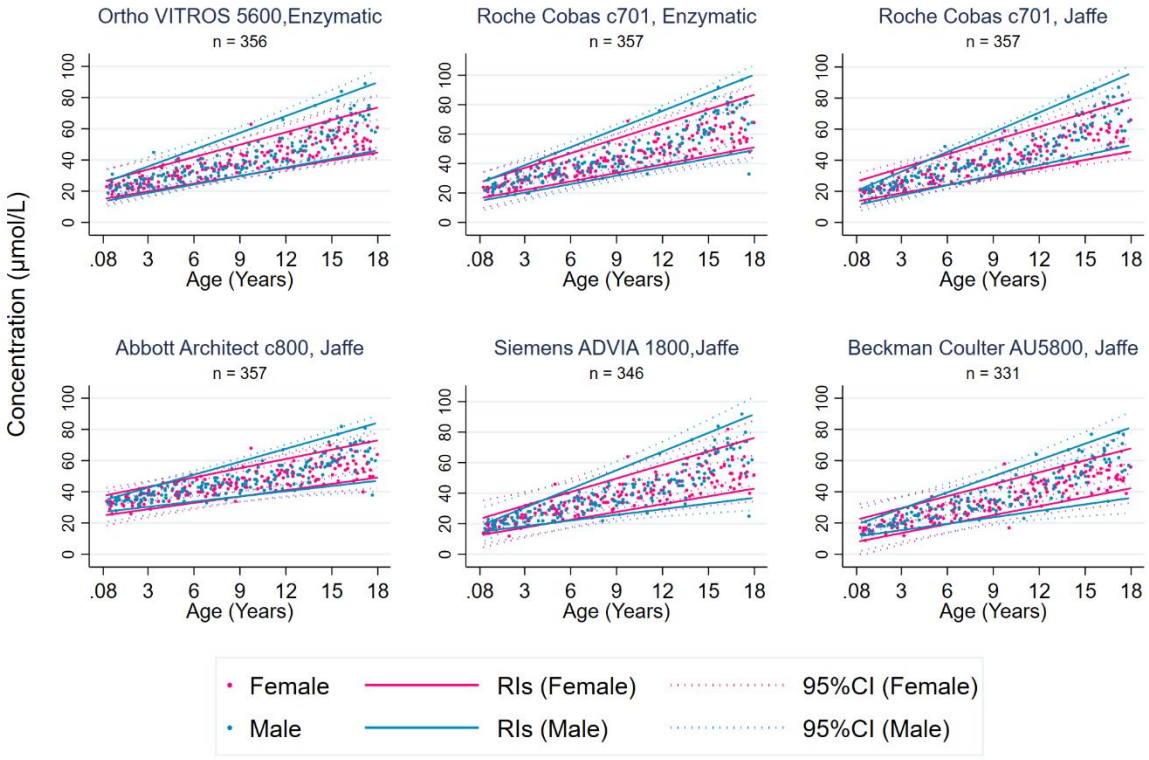
S-27: Reference intervals for free thyroxine



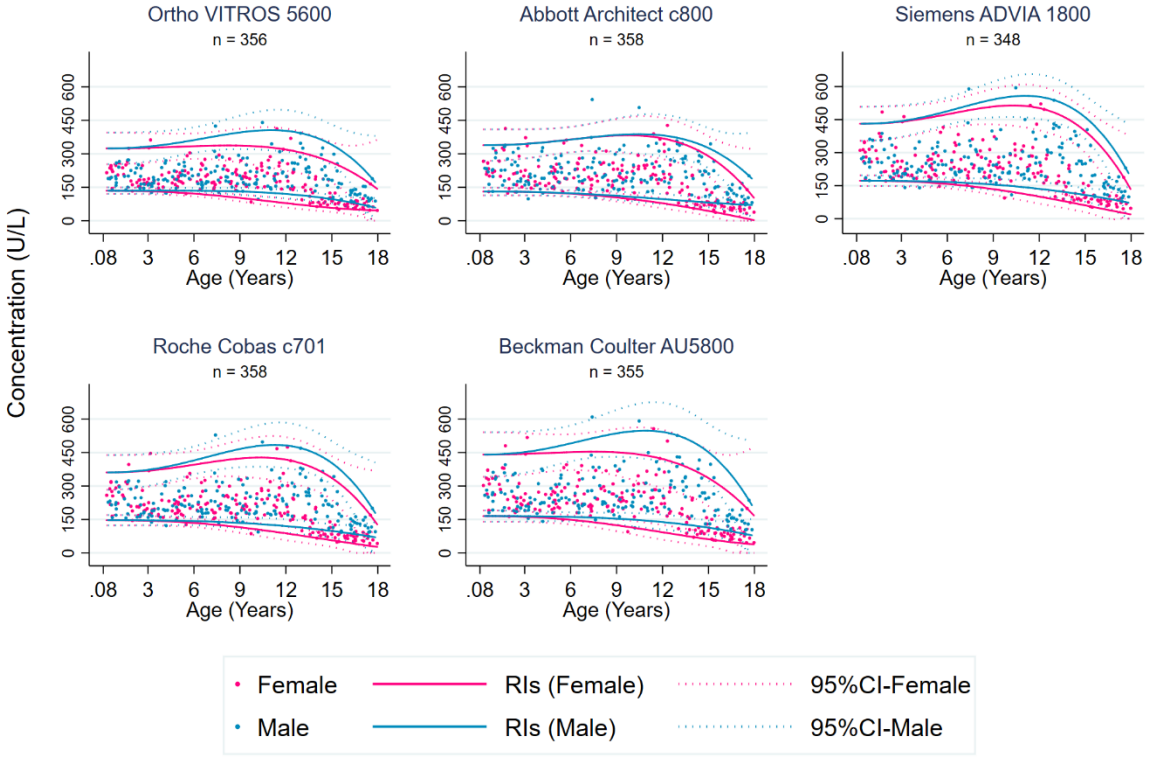
S-28: Reference intervals for potassium



S-29: Reference intervals for creatinine



S-30: Reference intervals for alkaline phosphatase



Appendix C: Suplementatry documents included in Chapter 4 (Hoq *et al*. Validation of the HAPPI Kids Continuous Age-Specific Pediatric Reference Intervals. The Journal of Applied Laboratory Medicine. 2020: jfaa045)

Supplementary Table 1: HKI Contributors

Role	**Person**	
Principle Investigator		
	Prof Paul Monagle	
	A/Prof Vera Ignjatovic	
Study Coordinators		
	Ms Vicky Karlaftis	
	Ms Chantal Attard	
	Ms Wincy Au-Yang	
	Ms Janet Burgess	
	Ms Jessica Cowley	
Participating Laboratories for sample testing		**Testing Scientist**
	The Royal Children's Hospital	Angela Chiriano
		Erik Haworth
		Chuong Tran
	The Royal Melbourne Hospital	Tina Nhan
	Melbourne Pathology	Mrs Daniella Angeleski
		Mr Warren Louey
	Monash Health	Dr Julie Newman
		Peter Edwards
	Dorevitch	Alan Webster
	Australian Clinical Laboratories	Daniel Cochrane
		Michelle Webber
Collaborating Clinicians for sample collection		**Site**
	Dr Ian McKenzie	Royal Children's Hospital
	Ms Jenny Ryan	Royal Women's Hospital
	Ms Anna Bourke	Royal Women's Hospital
	A/Prof Christine Rodda	Western Hospital
	Dr Martin Wright	Western Hospital
	A/Prof Wei Qi Fan	Northern Hospital
	Dr Simon Hauser	Northern Hospital
Pathology collectors		
	Ms Chauvy Burgess	

	Ms Jody Hand	
	Ms Kyra Hijniakoff	
	Ms Kathryn Bowers	
	Ms Lisa Howieson	
	Ms Aimee Chua	
Steering Committee		
	Prof Paul Monagle	Co-Chair
	A/Prof Vera Ignjatoivc	Co-Chair
	Dr Bill Renwick	Australian Clinical Laboratories
	Dr Sharon Choo	Royal Children's Hospital
	Dr Ellen Maxwell	Melbourne Pathology
	Cathy Cockshott	Dorevitch Pathology
	Dr Sue Matthews	Royal Children's Hospital
	Dr Kuang-Chih Hsiao	Royal Children's Hospital
	Dr Tina Yen	Royal Children's Hospital
	Mr Ray Czajko	Royal Melbourne Hospital
	Mr Peter Edwards	Biochemistry - Principal Scientist
	Mr Nick Drossos	Dorevitch
	Mr Daniel Cochrane	Australian Clinical Laboratories
	Ms Donna McNally	Australian Clinical Laboratories
	Ms Vicky Karlaftis	Study Coordinator
	Chantal Attard	Study Coordinator
	Jessica Cowley	Study Coordinator
	Mr Monsurul Hoq	Statistician
Biochemistry advisory panel		
	Dr Sue Matthews	Royal Children's Hospital
	Dr Tina Yen	Royal Children's Hospital
	A/Prof Ken Sikaris	Melbourne Pathology
	Dr Jim Doery	Monash Health
	Mr Ray Czajko	Royal Melbourne Hospital
Statistical analysis		
	Mr Monsurul Hoq	
	Prof John Carlin	
	A/Prof Susan Donath	

Supplementary Table 2: Analytical methods

Measurand	Analyser	Siemens ADVIA 1650/1800
Sodium	Analytical principle	Ion Selective Electrode Indirect (Diluted)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	CDC Flame Photometry Reference Method NIST SRM 909b
Potassium (Serum)	Analytical principle	Ion Selective Electrode Indirect (Diluted)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	CDC Flame Photometry Reference Method NIST SRM 909b
Chloride	Analytical principle	Ion Selective Electrode Indirect (Diluted)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	NIST Coulometric Reference Method NIST SRM 919
Bicarbonate	Analytical principle	Enzymatic
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Corning 965 Reference Method NIST SRM 192
Urea	Analytical principle	UREASE, Kinetic
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	CDC Reference Method NIST SRM 912a, NIST SRM 909b
Creatinine	Analytical principle 1	Alkaline Picrate (Jaffe), Kinetic
	Analytical principle 2	NA
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	IDMS Reference Method NIST SRM 967
Bilirubin, Total	Analytical principle	Chemical Oxidation (Vanadate)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	AACC Reference Method NIST SRM 916
Alkaline Phosphatase	Analytical principle	p-Nitrophenyl Phosphate Substrate - AMP Buffer - Rate Reaction IFCC Recommended
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	IFCC Reference Method
Aspartate Amino Transferase	Analytical principle 1	IFCC "Modified" without Pyridoxal-5-Phosphate
	Analytical principle 2	NA
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	IFCC Reference Method
Alanine Amino Transferase	Analytical principle 1	IFCC "modified" without Pyridoxal-5- phosphate
	Analytical principle 2	NA

	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Standardised against IFCC
Gamma-Glutamyl Transferase	Analytical principle	Other Gamma-Glutamyl-3-carboxy-4-nitroanilide Substrate
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Standardised against IFCC
Protein, total	Analytical principle	End Point with Blank
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Biuret Reference Method NIST SRM 927a
Albumin	Analytical principle 1	Bromocresol Purple
	Analytical principle 2	NA
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	ERM-DA470k / IFCC Reference Material
Calcium	Analytical principle	Arsenazo Dye
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Inductively Coupled Atomic Emission Method NIST SRM 915, NIST SRM 909b
Magnesium	Analytical principle	Xylidyl Blue
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Atomic Absorption Method NIST SRM 929
Phosphate	Analytical principle	Phosphomolybdate Formation
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Ammonium Molybdate Reference Method NIST SRM 937
Thyroid Stimulating Hormone	Analytical principle	Immunochemiluminometric Assay (ICMA - Sandwich Asay)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	WHO 3rd IS 81/565
Free Thyroxine	Analytical principle	Luminescent Immunoassay (LIA - Competitive Assay)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Traceable to an internal standard manufactured using U.S.P.(United States Pharmacopeia) material
Free Triiodothyronine	Analytical principle	Luminescent Immunoassay (LIA - Competitive Assay)
	Reagent	Siemens (Bayer)
	Traceability of Calibrator	Traceable to an internal standard manufactured using U.S.P.(United States Pharmacopeia) material

Supplementary Document 1: Secondary analysis assessing the validation of the RIs for Albumin and Creatinine using the test results from the Validation Cohort.

A: Albumin

The following figure revealed that the test results that were outside the RIs are distributed uniformly across age. However majority of the values are outside the upper reference limit.

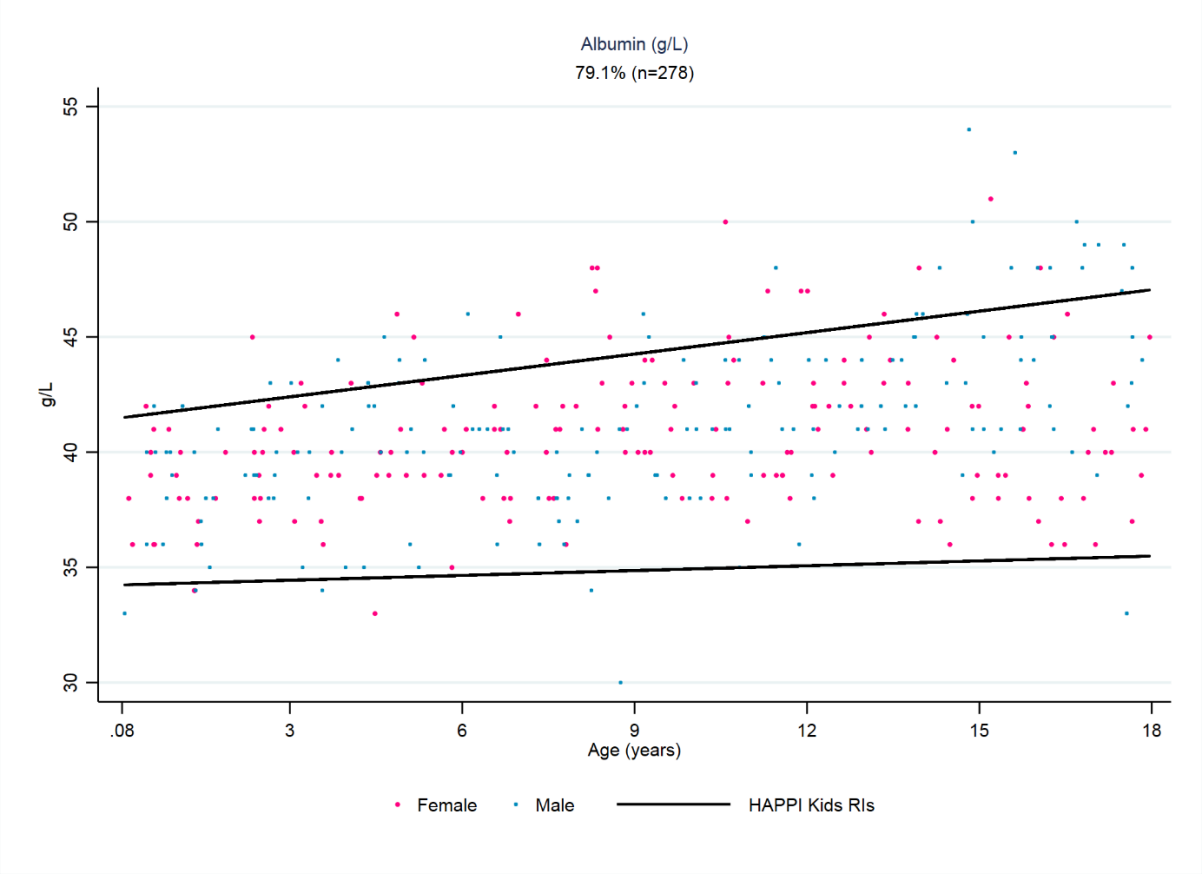


Figure 1: Validation of RIs for Albumin

The majority of the test results that were outside the RIs were tested on two dates i.e. 19th and 25th October 2017.

Table 5: Proportion of test results that were outside the RI by date of testing

Date of Testing	Proportion of test results that were outside the published RI, % (n)
18 May 2017	100 (1)
17 August 2017	0
18 August 2017	15.8 (3)
22 August 2017	3.3 (1)
11 October 2017	0

12 October 2017	10.2 (5)
13 October 2017	6.5 (5)
16 October 2017	11.5 (3)
17 October 2017	12.5 (3)
18 October 2017	12.5 (4)
19 October 2017	37.9 (11)
25 October 2017	59.6 (28)
Total	20.9 (58)

However, the QC reported the laboratories on those days were within limit.

Table 6 : Quality control data for albumin reported by the routine laboratory

		Pre-Run				Post-Run			
		Albumin L1		Albumin L2		Albumin L1		Albumin L2	
	Date	Result	Z score	Result	Z score	Result	Z score	Result	Z score
Panel 1 Batch 1 Run 1	17/08/2017	25.6	0.14	40.1	0.43	25.5	0.01	39.9	0.43
Panel 1 Batch 1 Run 2	22/08/2017	25.7	0.23	40.2	0.46	25.5	-0.04	38.8	0.05
Panel 1 Batch 1 Run 3	11/10/2017	24.8	-1.01	38.8	-1.20	24.7	-1.13	38.8	-1.21
Panel 1 Batch 1 Run 4	12/10/2017	25.4	-0.09	39.4	-0.52	24.6	-1.30	38.4	-1.75
Panel 1 Batch 2 Run 1	12/10/2017	24.6	-1.30	38.4	-1.75	24.7	-1.17	38.3	-1.89
Panel 1 Batch 2 Run 2	13/10/2017	24.7	-1.17	38.3	-1.89	24.7	-1.11	39.3	-0.59
Panel 1 Batch 2 Run 3	13/10/2017	24.7	-1.17	38.3	-1.89	24.7	-1.11	39.3	-0.59
Panel 1 Batch 2 Run 4	13/10/2017	24.7	-1.17	38.3	-1.89	24.7	-1.11	39.3	-0.59
Panel 1 Batch 3 Run 1	16/10/2017	24.7	-1.19	38.5	-1.68	24.5	-1.39	38.6	-1.54
Panel 1 Batch 3 Run 2	17/10/2017	26.2	0.94	40.8	1.23	126.1	0.81	40.9	1.36
Panel 1 Batch 3 Run 3	18/10/2017	25.8	0.41	41.2	1.79	26.1	0.83	40.6	0.99
Panel 1 Batch 4 Run 1	19/10/2017	26.1	0.84	40.9	1.39	25.9	0.60	40.6	1.05
Panel 1 Batch 4 Run 2	25/10/2017	26.0	0.70	40.9	1.40	26.5	1.41	41.0	1.56
Panel 1 Batch 4 Run 3	25/10/2017	26.5	1.41	41.0	1.56	26.4	1.33	41.0	1.46
S.D.		0.7 g/L		0.8 g/L		0.7 g/L		0.8 g/L	

The summary statistics of Albumin by date of testing also confirmed that the median Albumin on 19th and 25th October 2017 was 44 and 45 respectively and were higher than upper limit.

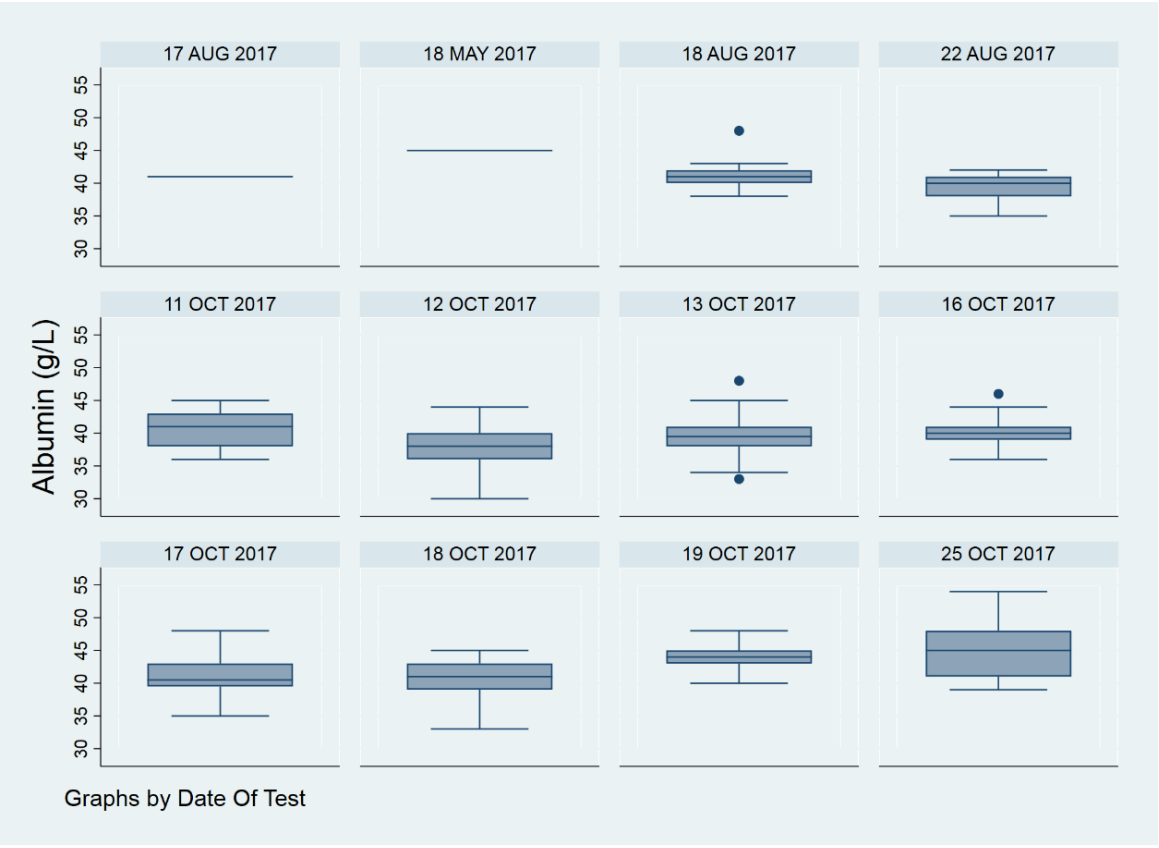


Figure 2: Box plot for albumin by date of test

Hence we decided to exclude the test results tested on October 19th and 25th for albumin from the Validation cohort.

Table 7 : Proportion of test results for calcium that were outside the published RI by date of testing

Date of Testing	Proportion of test results for calcium that were outside the published RI, % (n)
18 May 2017	0
17 August 2017	0
18 August 2017	7.7 (1)
22 August 2017	21.4 (3)
11 October 2017	0
12 October 2017	8.8 (3)
13 October 2017	8.8 (6)
16 October 2017	25.0 (4)
17 October 2017	12.5 (3)
18 October 2017	9.4 (3)
19 October 2017	0
25 October 2017	4.4 (2)
Total	9.0 (25)

Table 8: Proportion of test results for phosphate that were outside the published RIs by date of testing

Date of Testing	Proportion of test results for phosphate that were outside the published RI, % (n)
18 May 2017	0
17 August 2017	0
18 August 2017	7.7 (1)
22 August 2017	0
11 October 2017	0
12 October 2017	11.8 (4)
13 October 2017	4.4 (3)
16 October 2017	6.3 (1)
17 October 2017	8.3 (2)
18 October 2017	12.5 (4)
19 October 2017	0
25 October 2017	6.5 (3)
Total	6.5 (18)

The overall proportion of test results for Calcium and Phosphate that were outside the published RIs was satisfactory.

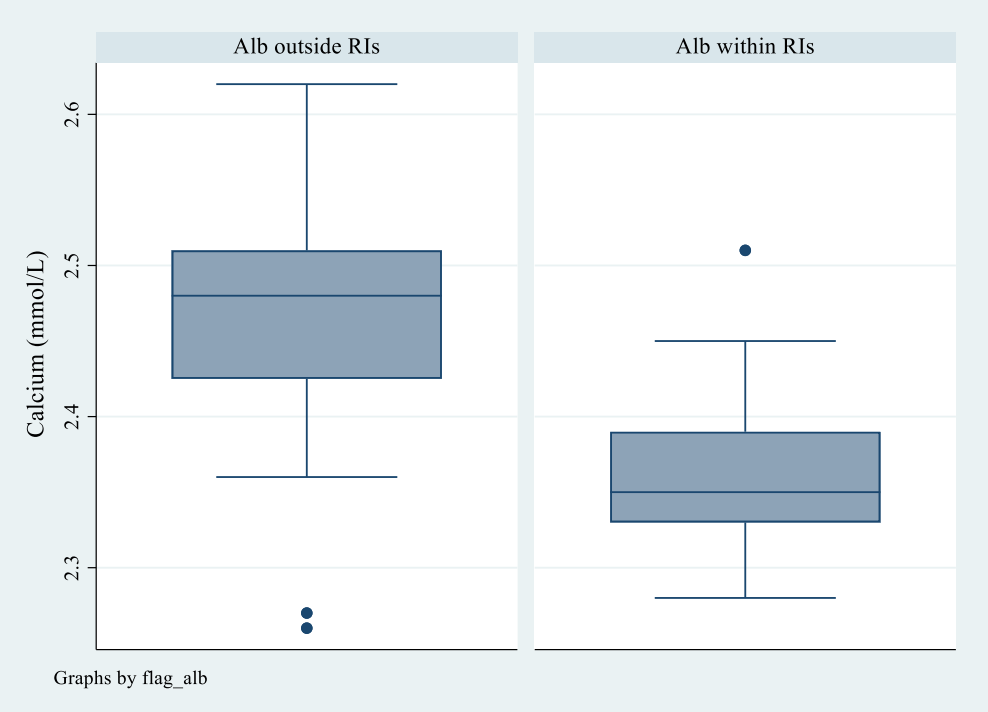


Figure 4: Comparison of mean Calcium tested on 2 days with high albumin flagging rate

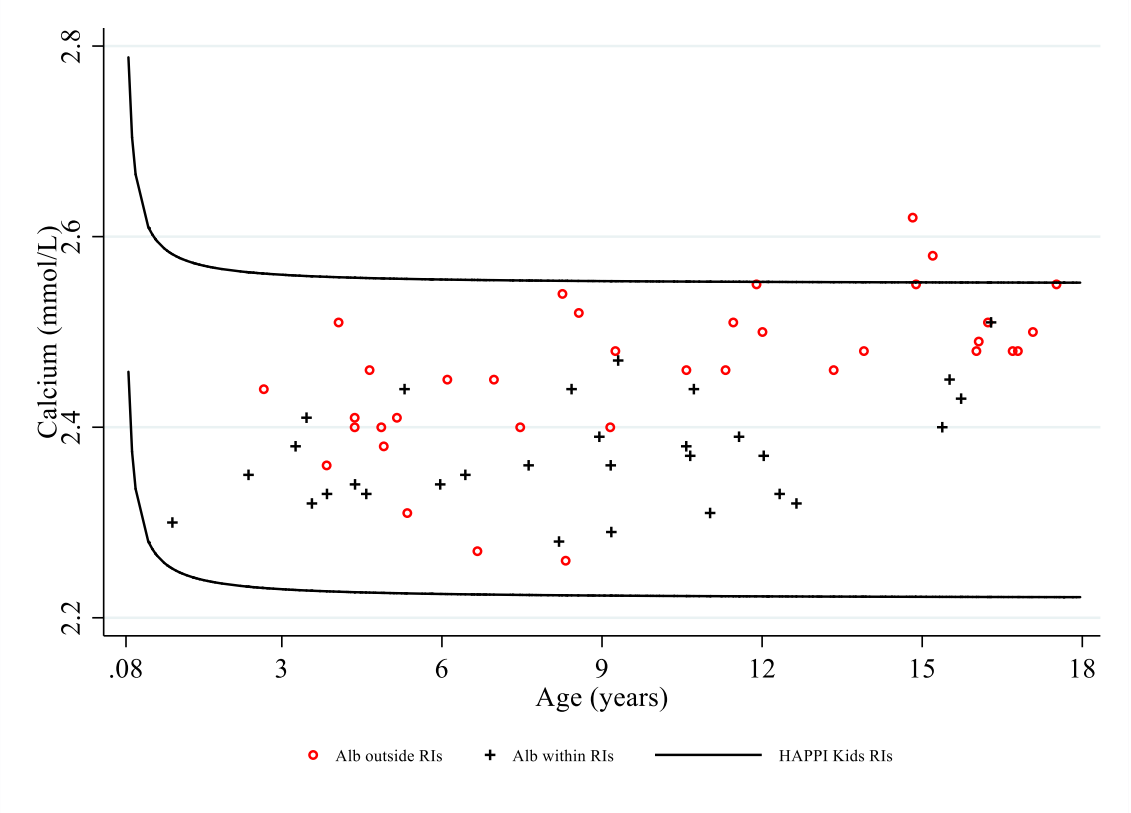


Figure 5: Distribution of the calcium on 2 days with high albumin flagging rate with respect to the published RIs

Based on figure 3 and 4, we assume that the pre-analytical issues have influenced the measurement of calcium. However, the results were within the RIs and validated the RIs.

B: Creatinine

The following figure revealed that the test results that were outside the RIs are distributed around early years of life.

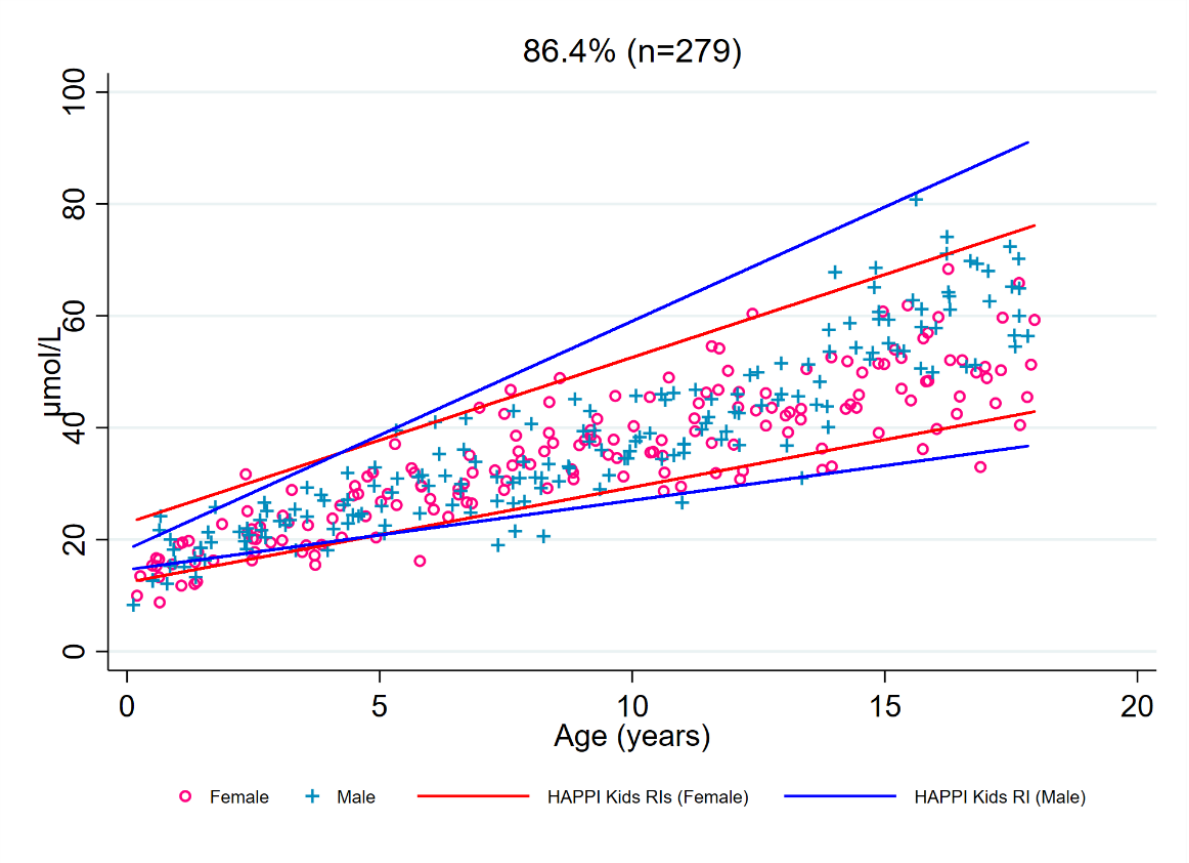


Figure 6: Validation of RIs for Creatinine

Table 9: Proportion of test results that were outside the RIs by age

Age	Proportion of test results that were outside the published RI, % (n)
1*	70.0 (7)
2	50.0 (6)
3	12.5 (2)
4	26.3 (5)
5	5.5 (1)
6	5.0 (1)
7	0
8	17.6 (3)
9	10.5 (2)
10	0
11	13.3 (2)
12	6.3 (1)
13	21.4 (3)

14	21.4 (3)
15	0
16	6.3 (1)
17	7.1 (1)
18	0
Total	13.6 (38)

*1 year refers to children 30 days to 365 days, 2 year refers to children 366 days to 730 days, etc.

The above table revealed that the proportion of test results that were outside the RIs are much higher for children aged less than 2 years compared to children older than 2 years.

Supplementary Document 2: Minimum sample size required for validating age-specific RIs

Probability of validating 95% reference intervals as a function of sample size when the criterion is that at least 90% of the samples fall within the interval

For validation of continuous reference intervals (ris),

- We considered each sample as an independent trial, where the sample will either fall within the ris or be outside the ris.
- For 95% ris, the probability that a sample will be within the ris is 0.95.
- For validation of 95% ris, the proportion of samples that should be within the ris is at least 90%.

Hence, based on the binomial distribution formula in (i) and using microsoft excel's "binom.dist" function (see figure s4.1), we calculated the cumulative probability of at least 90% of n samples that are within the ris for samples sizes between 20 and 240, and presented in the table s4.1 and in figure s4.2.

$$P\ (X \geq k) = 1 - P(X < k) = 1 - \sum_{j=0}^{k-1} \frac{n!}{!(k-1)!}\ p^{k-1}(1-p)^{n-(k-1)} \quad \ldots \text{(i)}$$

Where n = number of samples, p = proportion of samples within 95% RIs, and k = required number of samples within the RIs

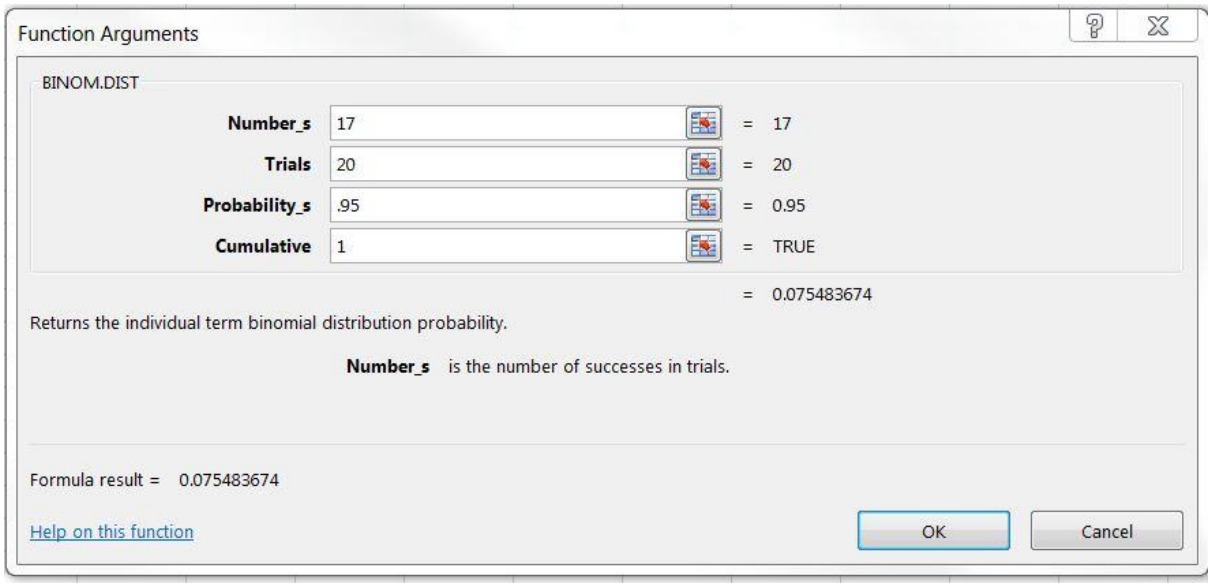


Figure 1: BINOM.DIST function in Microsoft Excel 2013

Table 1: Probability of validating 95% RIs with at least 90% samples within the RIs

Proportion of samples within 95% RIs	**Sample size n**	**Required no. of sampled that should be within RIs for validation k = n x 0.90**	**Pr (X >=k)***
0.95	20	18	0.9245
0.95	40	36	0.9520
0.95	60	54	0.9703
0.95	80	72	0.9816
0.95	100	90	0.9885
0.95	120	108	0.9928
0.95	140	126	0.9955
0.95	160	144	0.9971
0.95	180	162	0.9982
0.95	200	180	0.9988
0.95	220	198	0.9993
0.95	240	216	0.9995
0.95	260	234	0.9997

*Based on binomial distribution formula in (i)

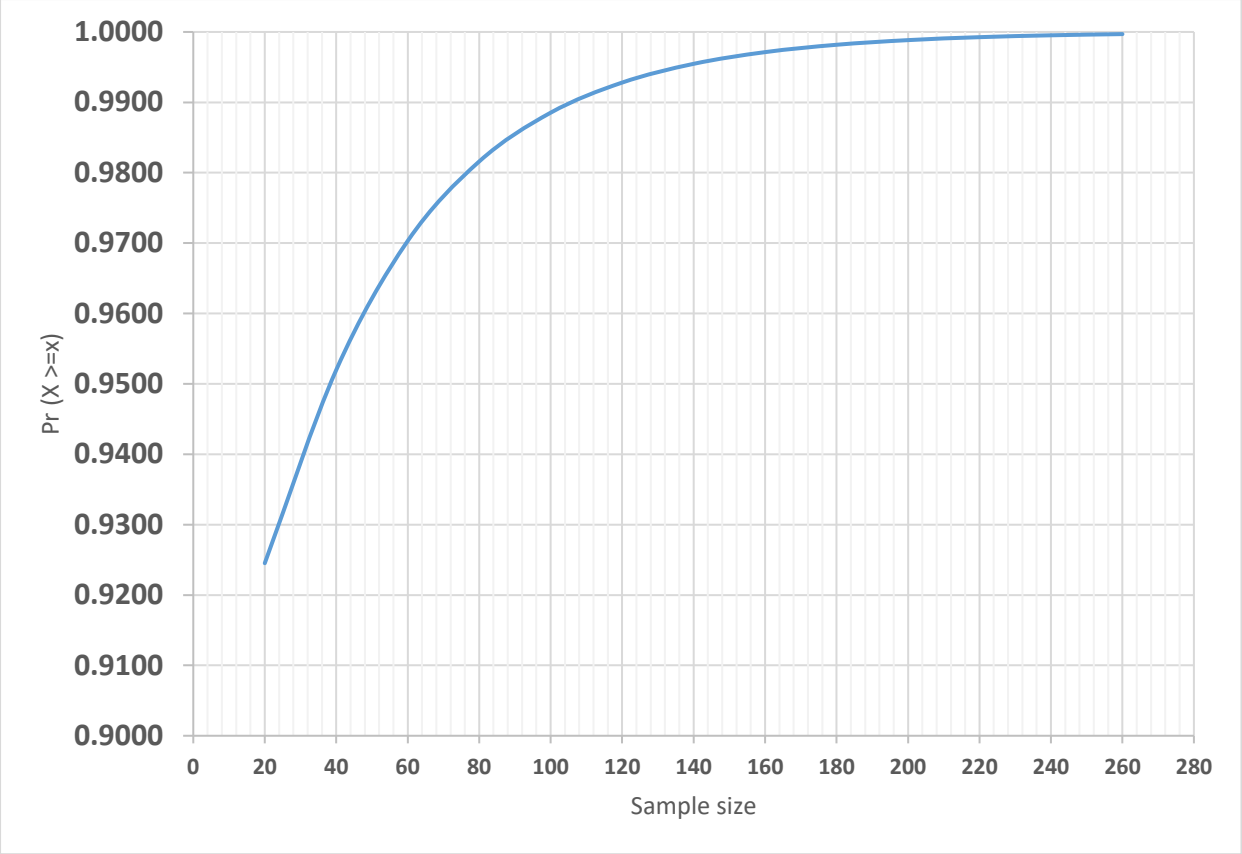
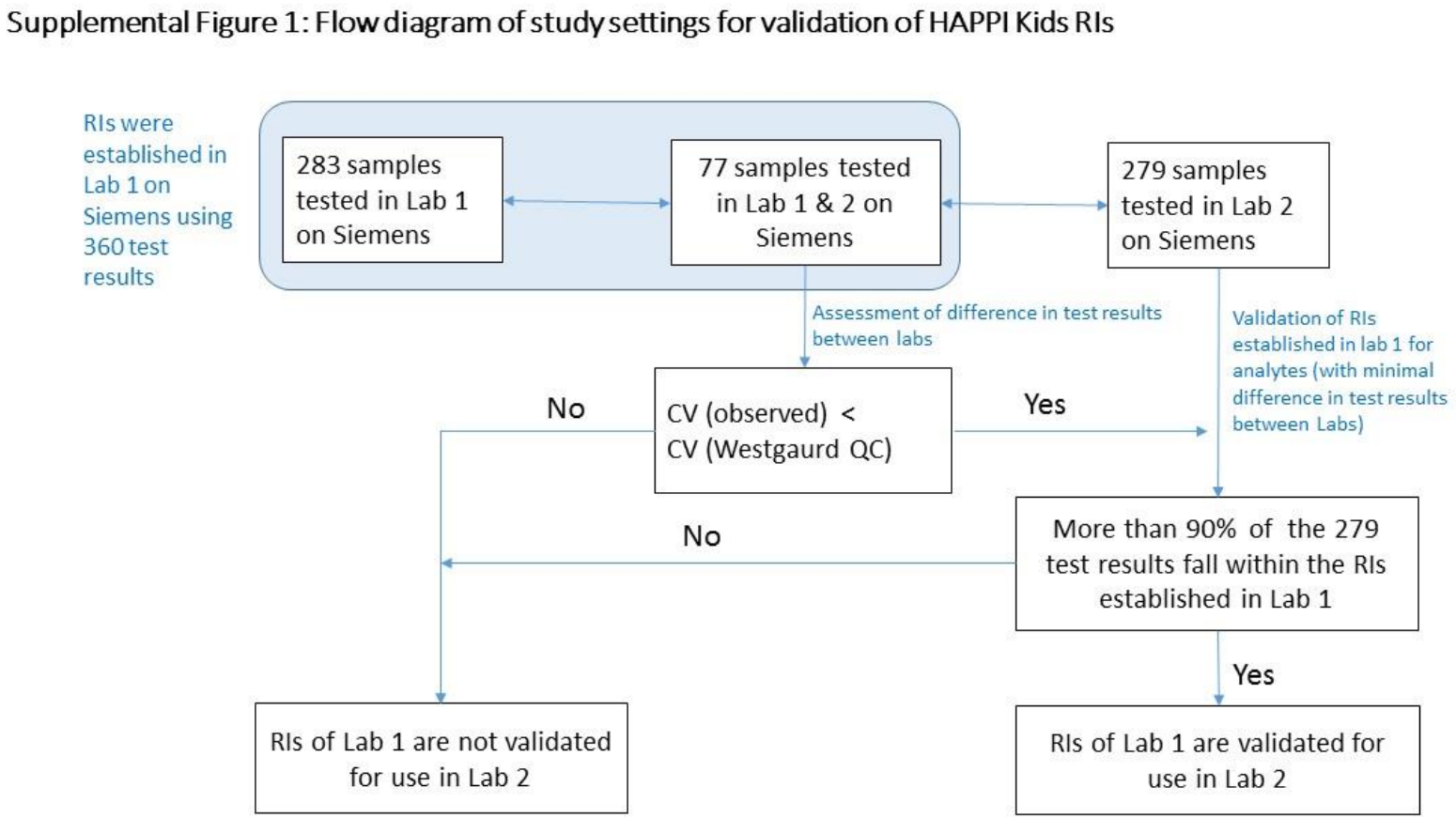


Figure.2: Probability of validating 95% RIs with at least 90% samples within the RIs

The probability of validating 95% RIs increases slightly as the sample size increases from 100 to 260. Hence, at least 100 paediatric samples should be tested for validation of 95% RIs with at least 90% samples falling with the interval.

Supplementary Figure 1: Flow diagram of study settings for validation of HAPPI Kids RIs



Supplementary Figure 2: Agreement and difference in test results between HAPPI Kids study laboratory and a routine laboratory

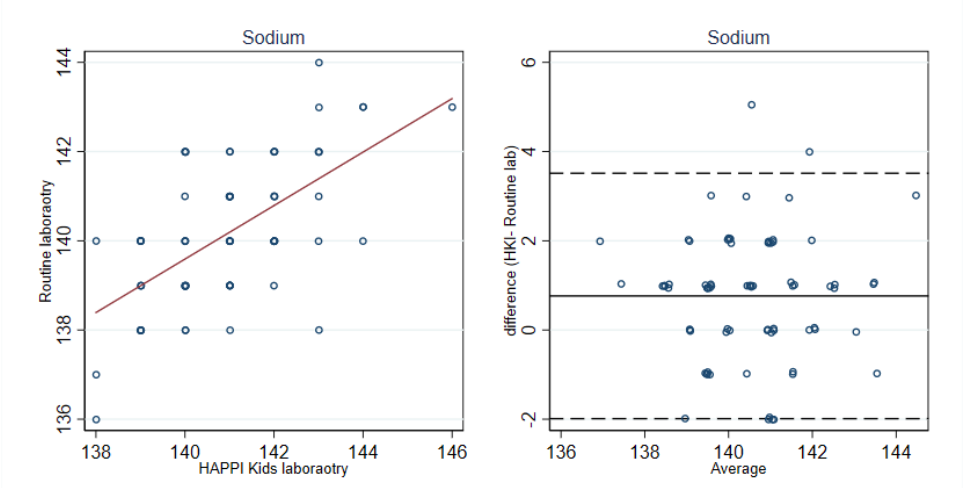


Figure S 2-A: Sodium

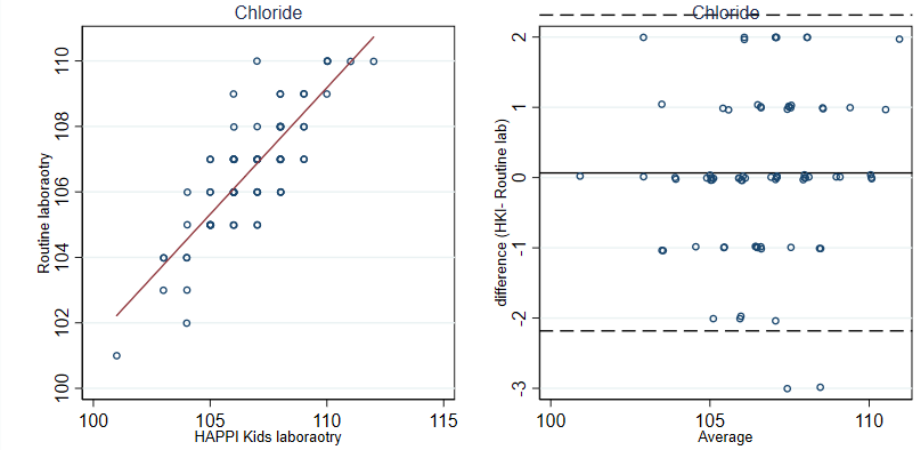


Figure S 2-B: Chloride

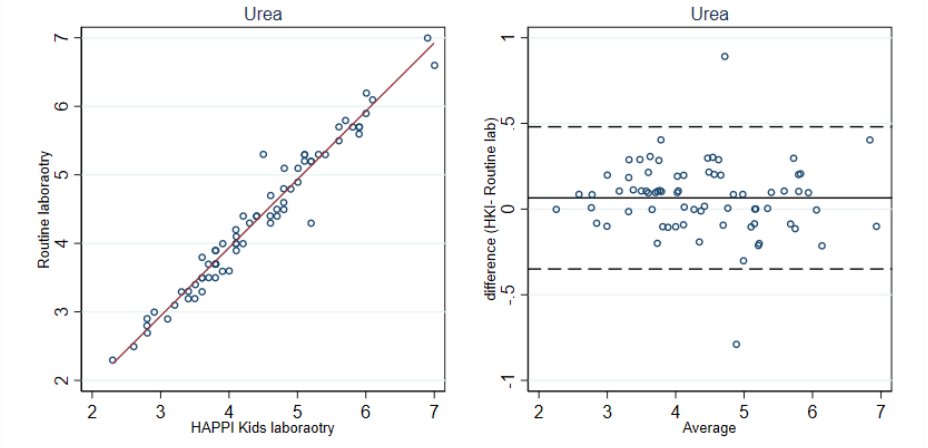


Figure S 2-C: Urea

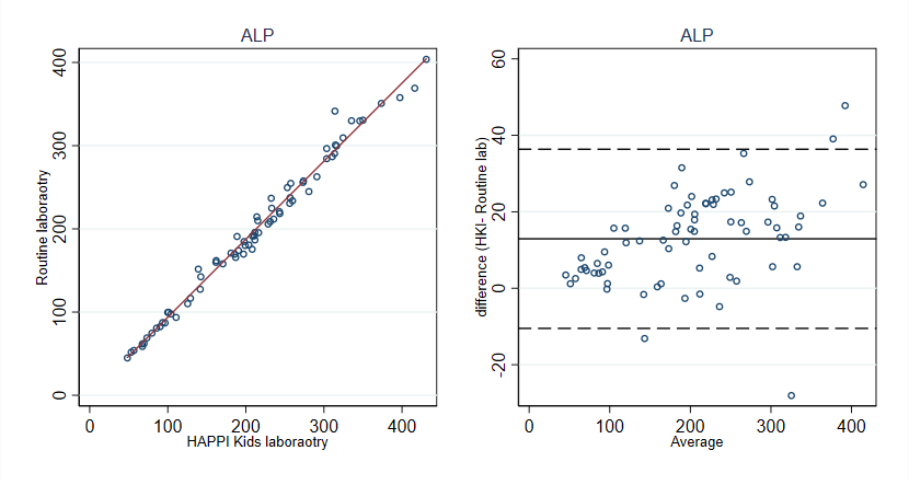


Figure S 2-D: ALP

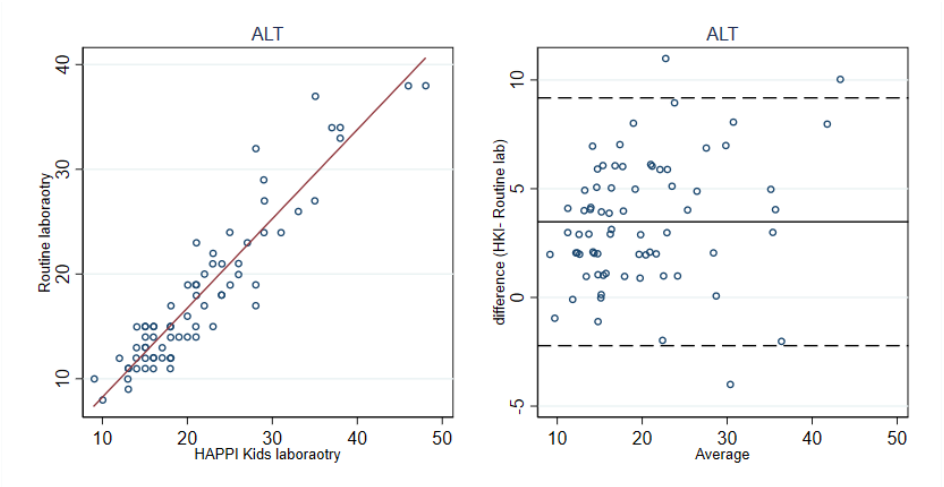


Figure S 2-E: ALT

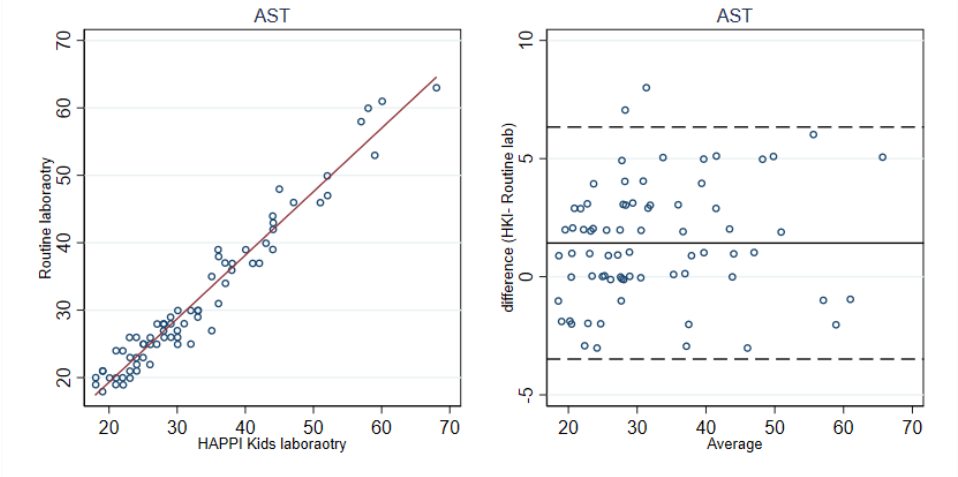


Figure S 2-F: AST

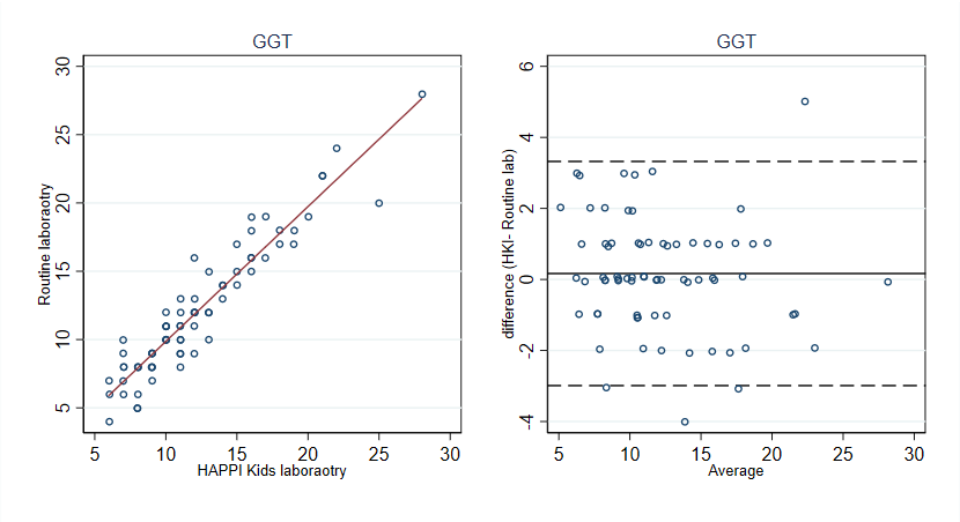


Figure S 2-G: GGT

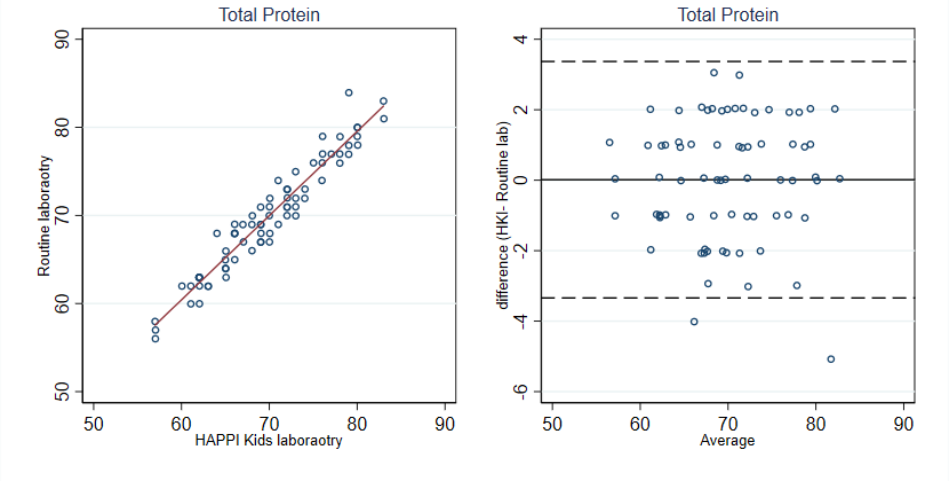


Figure S 2-H: Total Protein

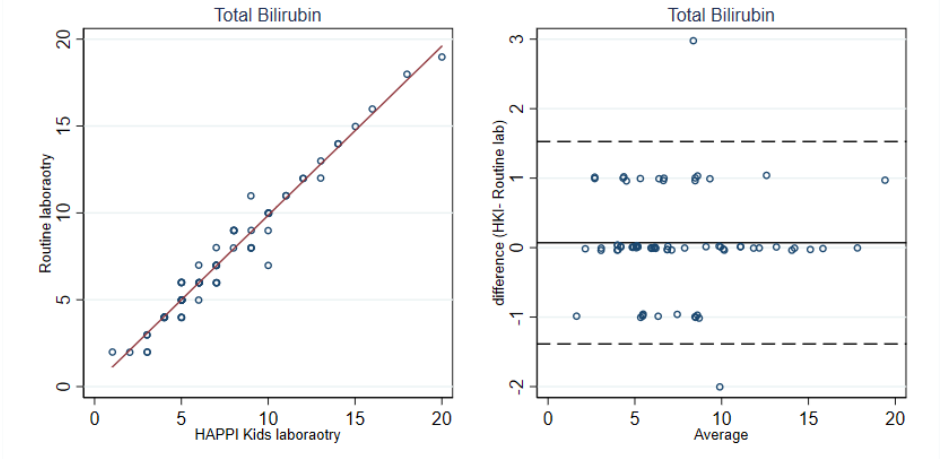


Figure S 2-I: Total Bilirubin

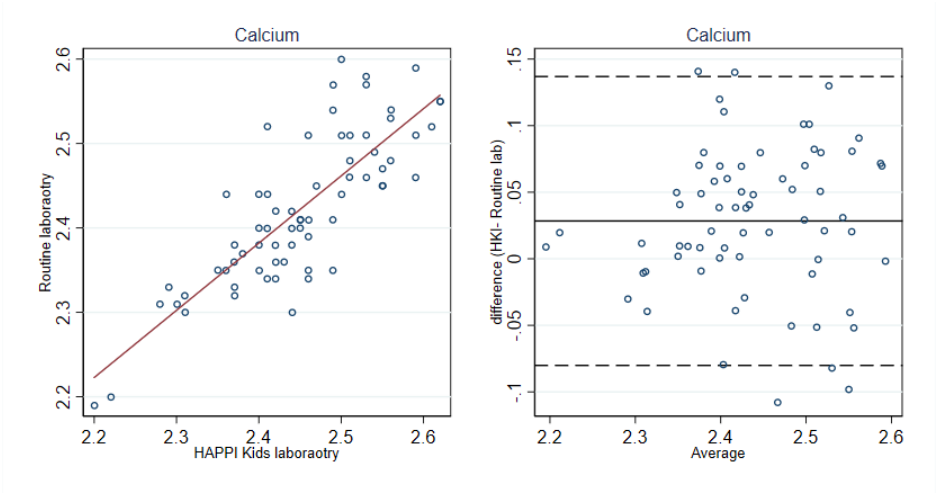


Figure S 2-J: Calcium

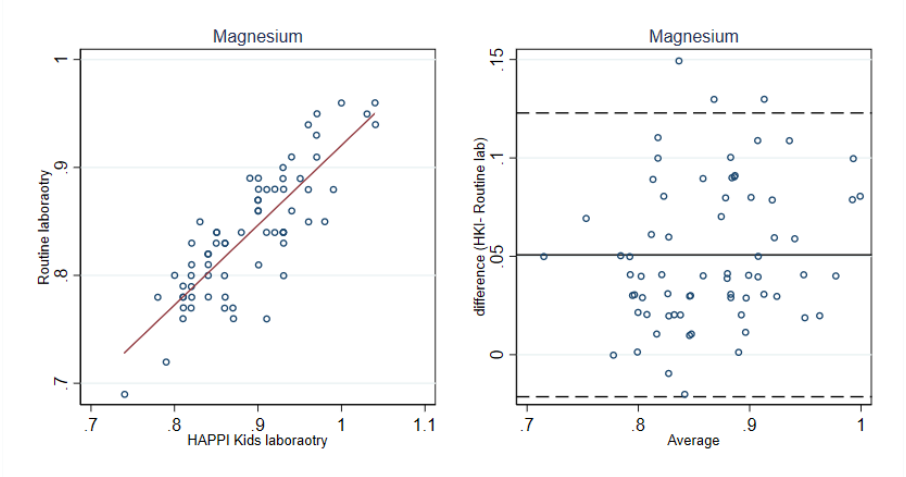


Figure S 2-K: Magnesium

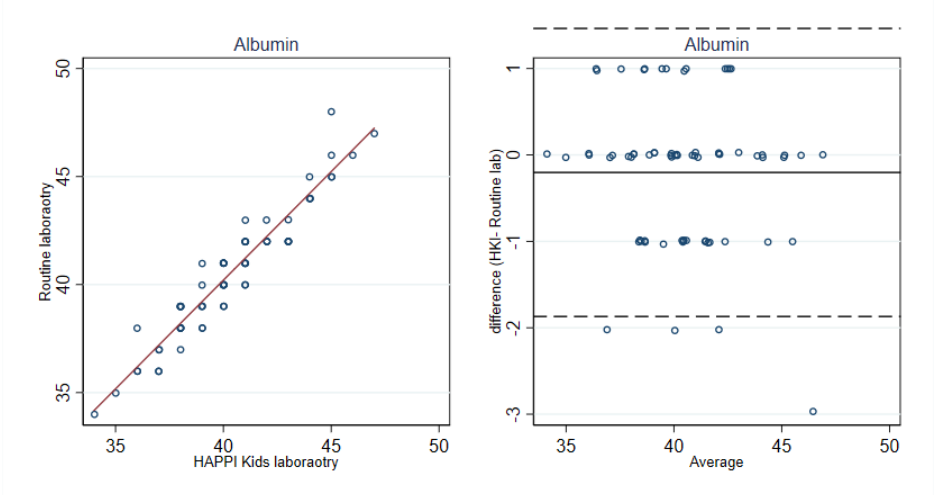


Figure S 2-L: Albumin

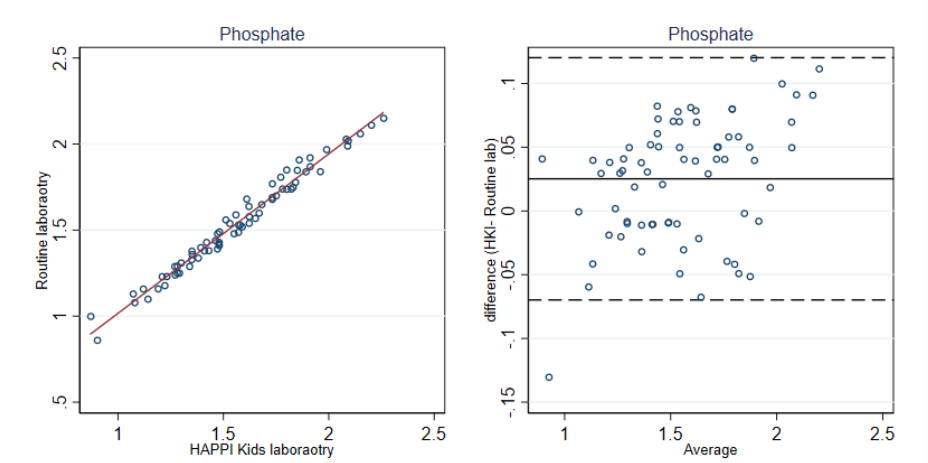


Figure S 2-M: Phosphate

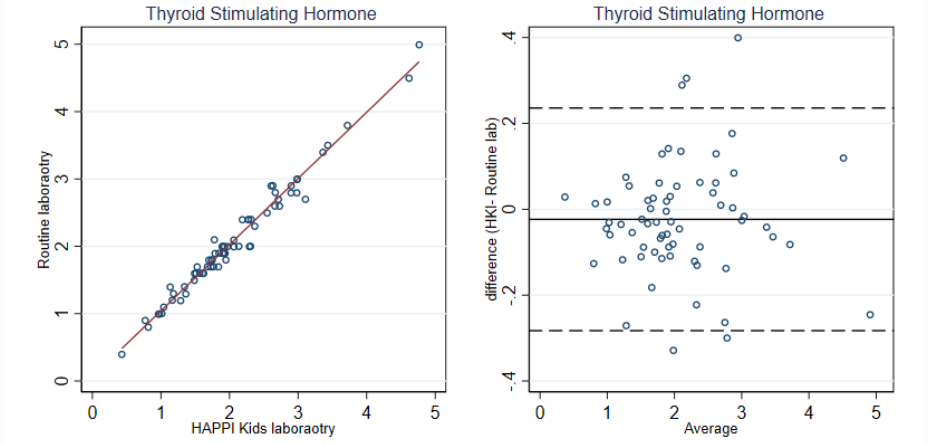


Figure S 2-N: TSH

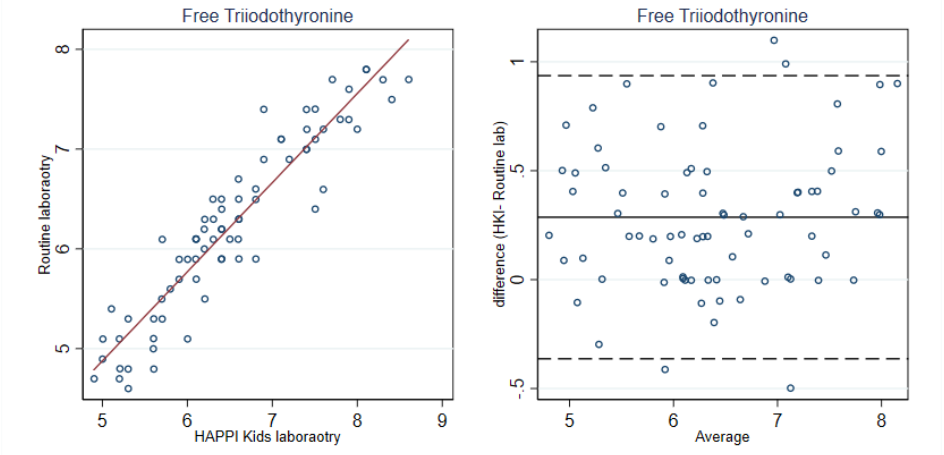


Figure 2 S - 0: FT3

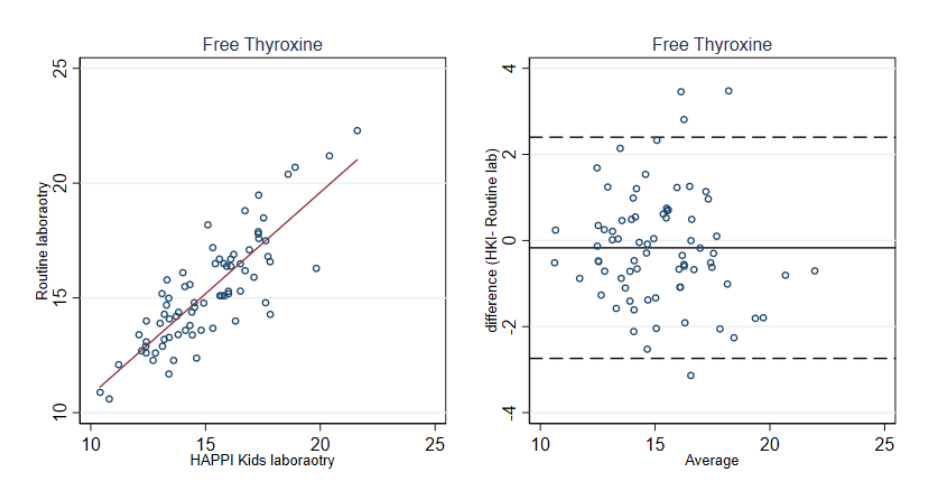


Figure 2 S-P: FT4

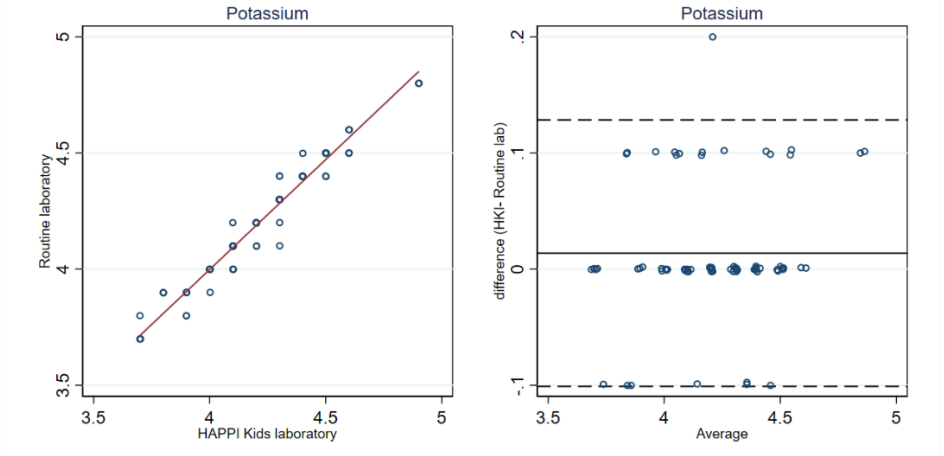


Figure 2 S – Q: Potassium

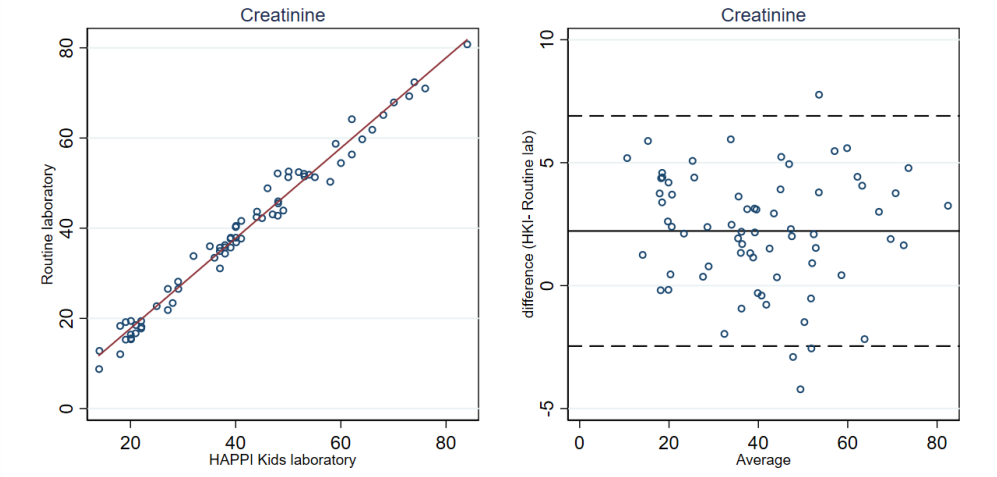
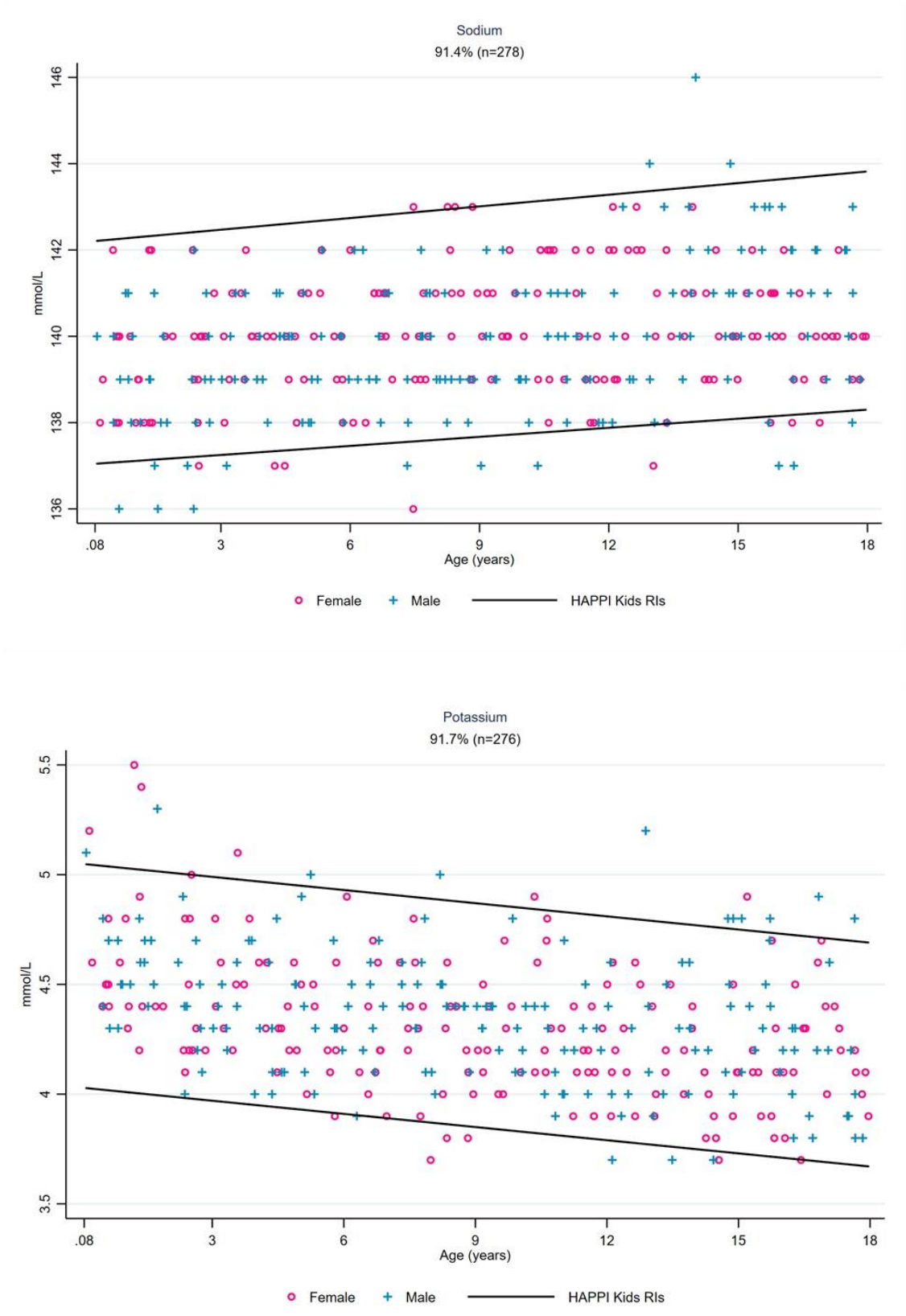
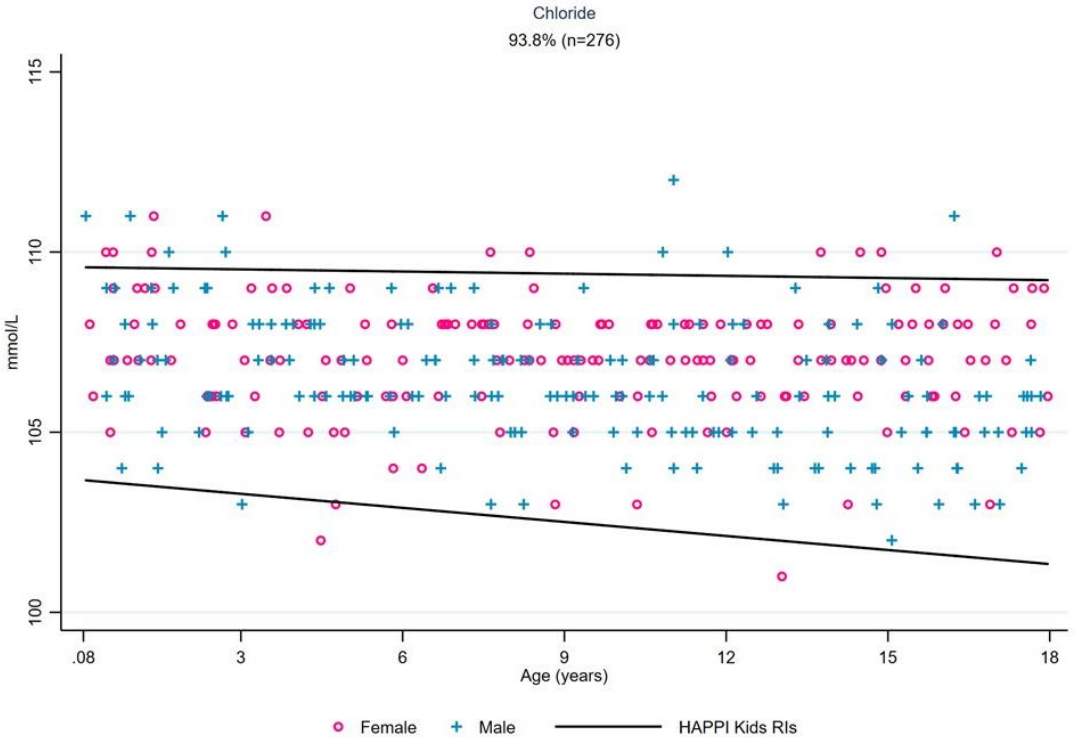
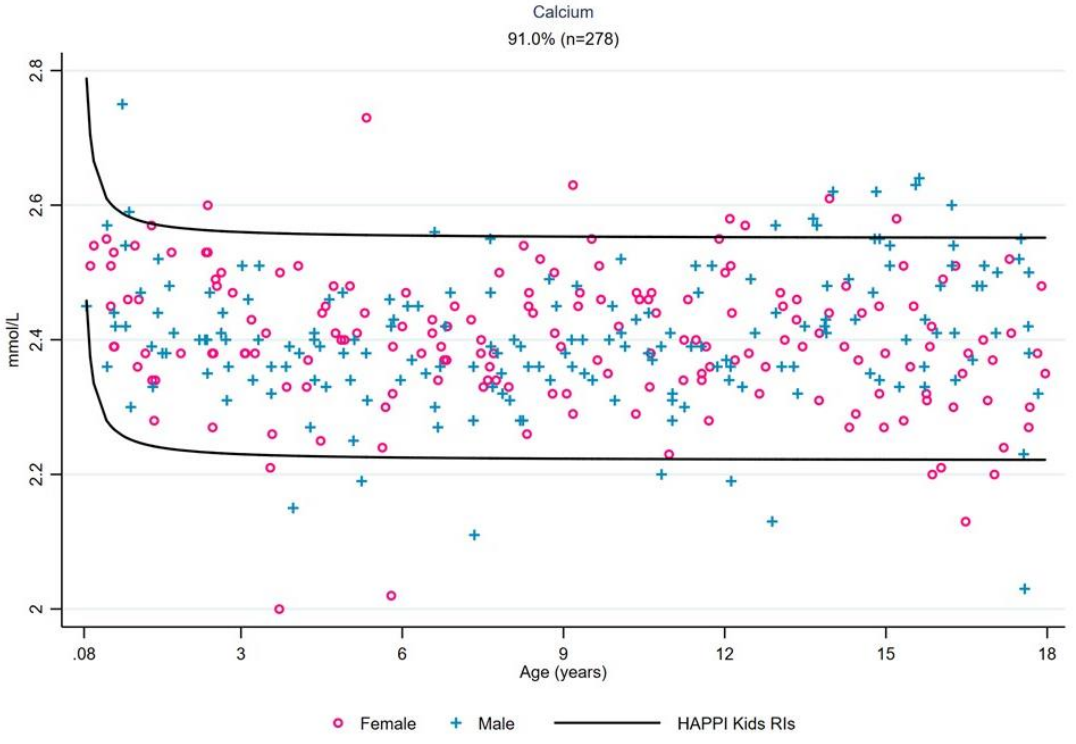


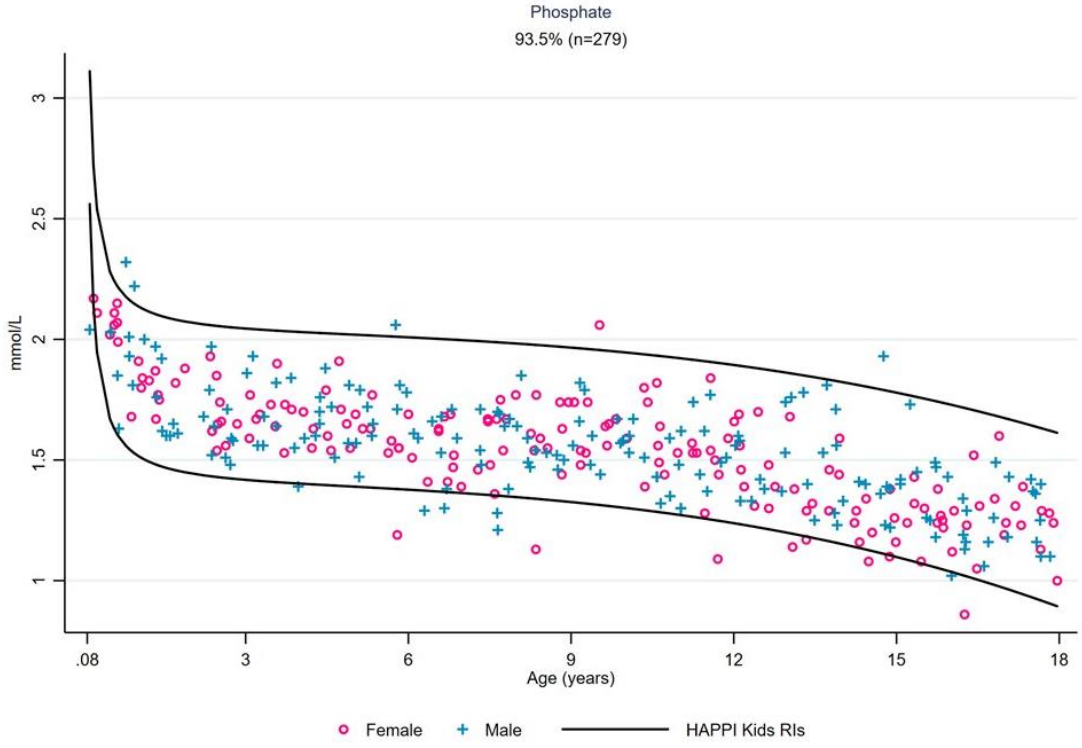
Figure 2 S – R: Creatinine

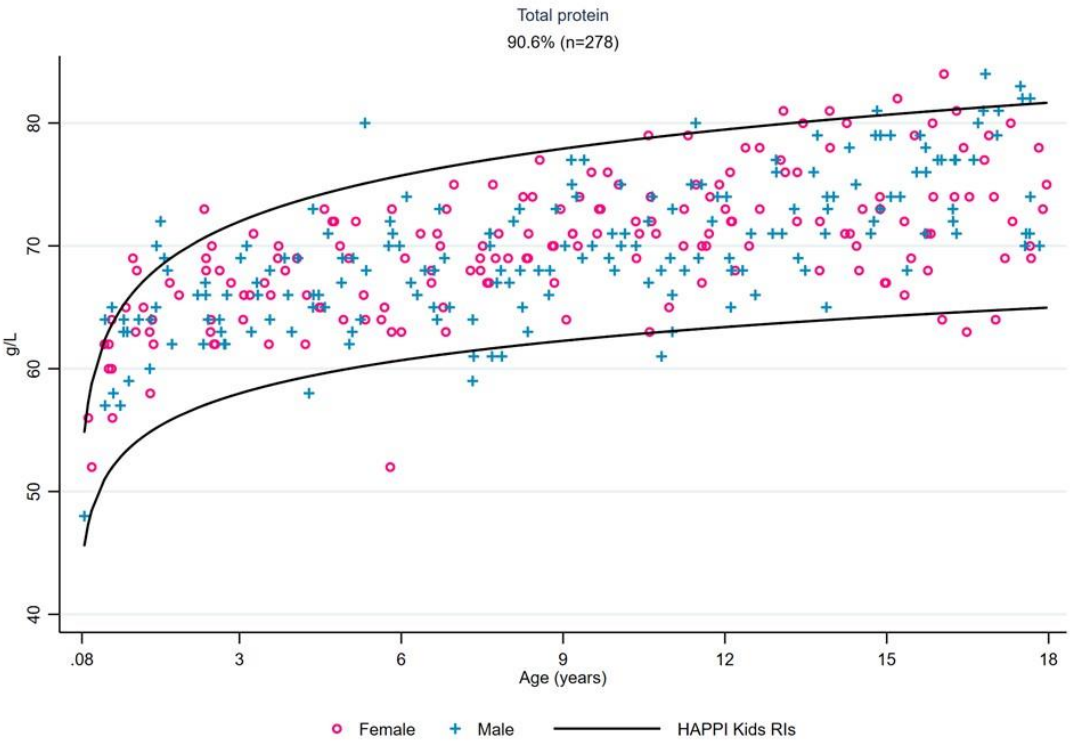
Supplementary Figure 3: Reference values from the validation cohort that were within the HAPPI Kids RIs.

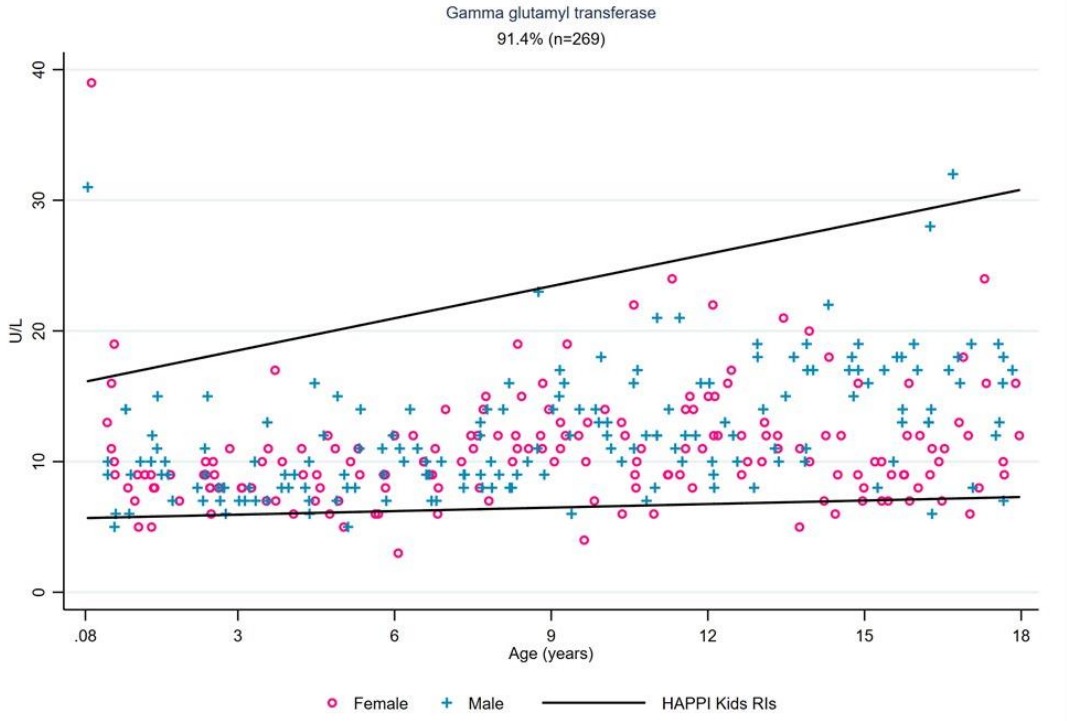


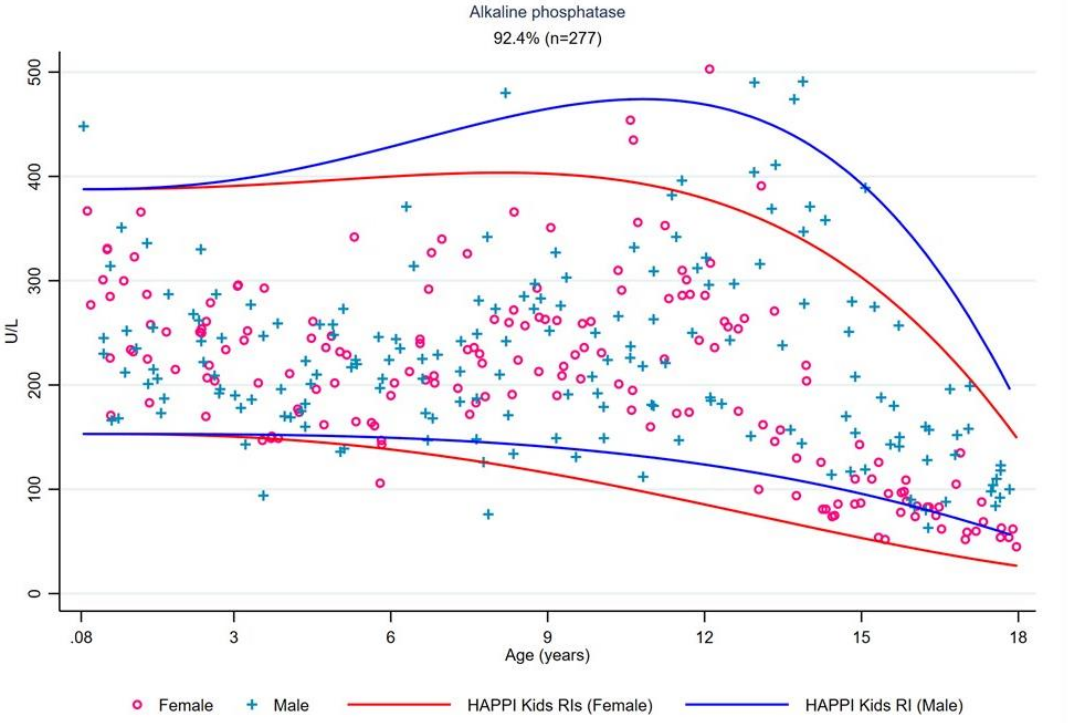

Chloride
93.8% (n=276)
mmol/L
115
110
105
100
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs

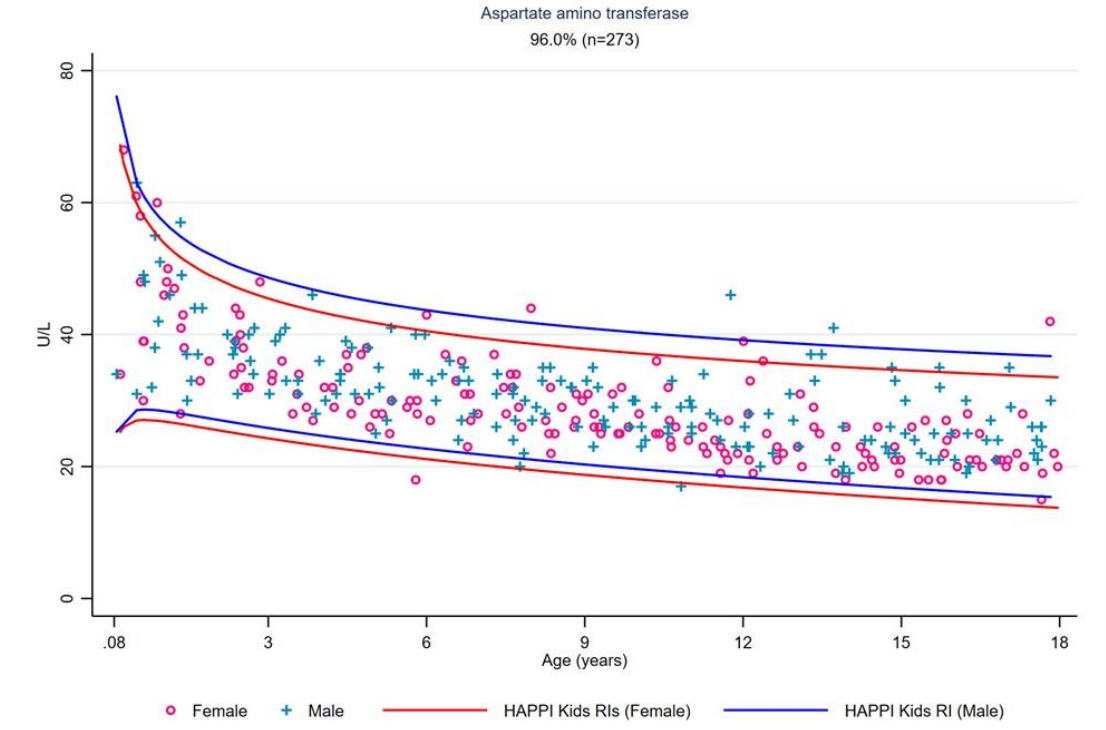

Calcium
91.0% (n=278)
mmol/L
2.8
2.6
2.4
2.2
2
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs

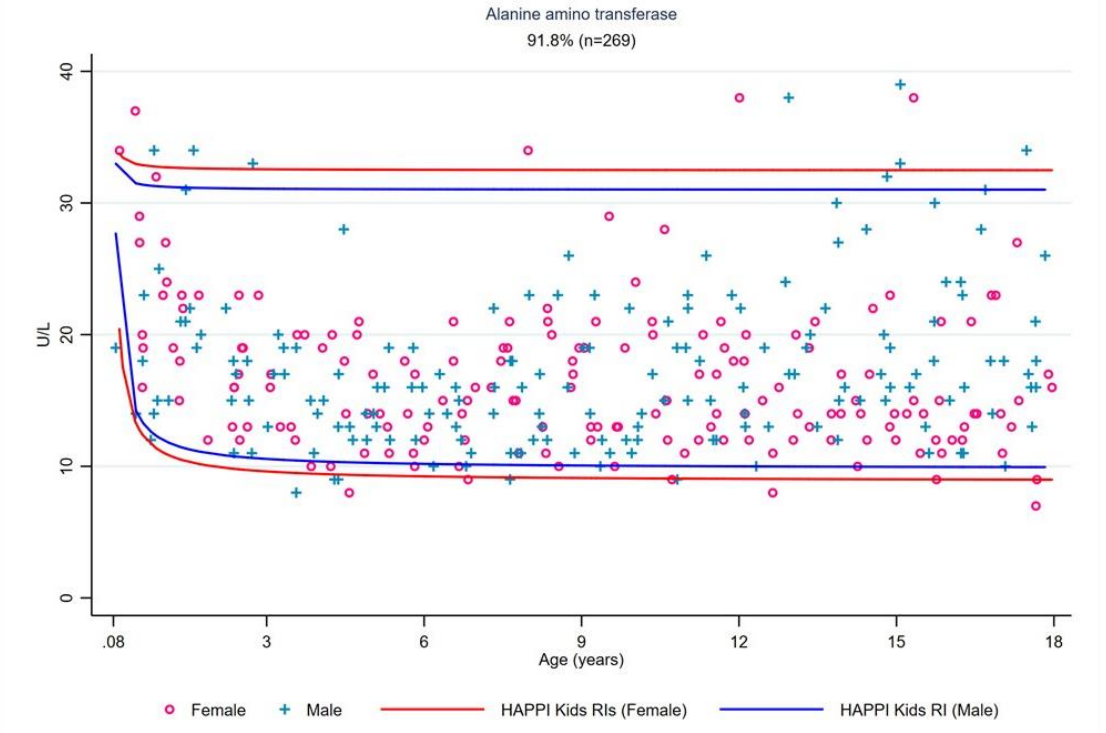

Phosphate
93.5% (n=279)
mmol/L
1
1.5
2
2.5
3
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs

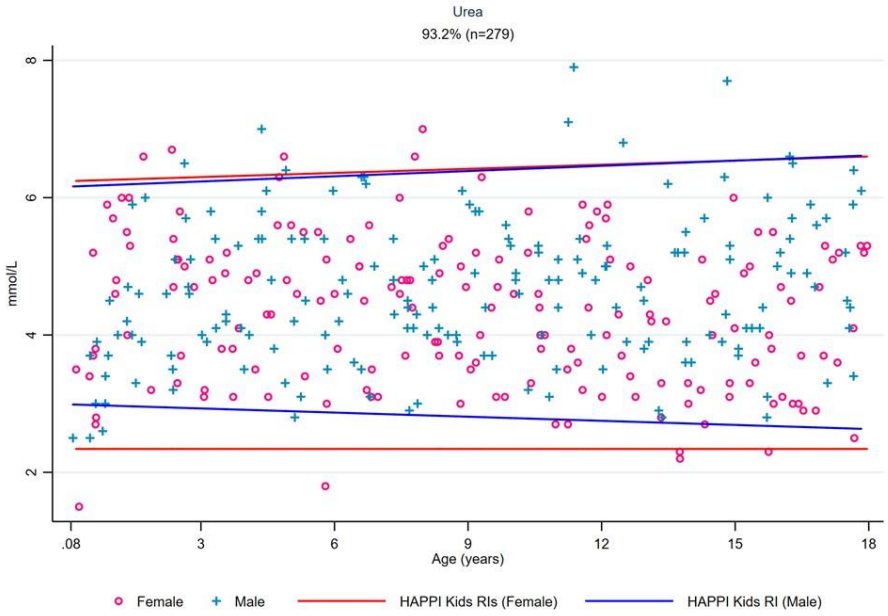

Total protein
90.6% (n=278)
g/L
40
50
60
70
80
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs

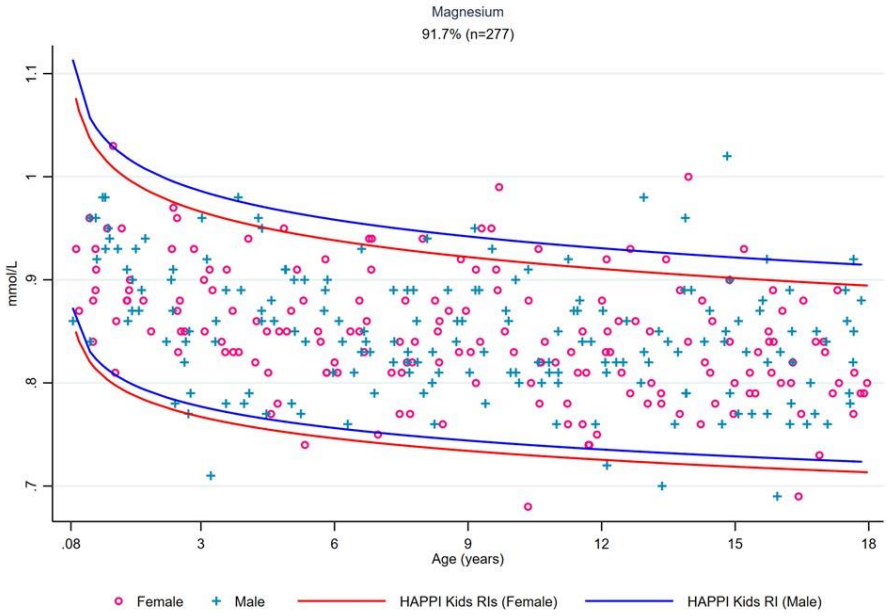

Gamma glutamyl transferase
91.4% (n=269)
U/L
0
10
20
30
40
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs

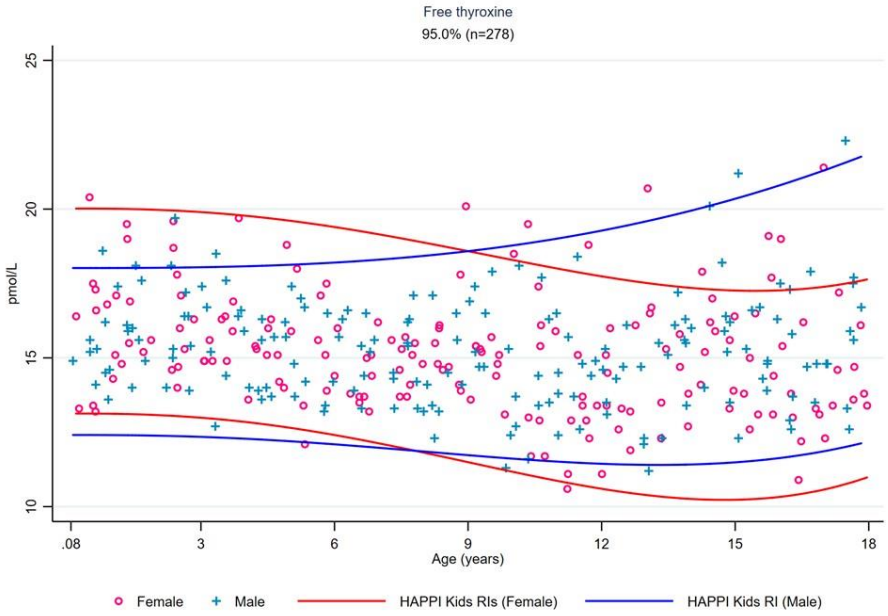

Alkaline phosphatase
92.4% (n=277)
U/L
0
100
200
300
400
500
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

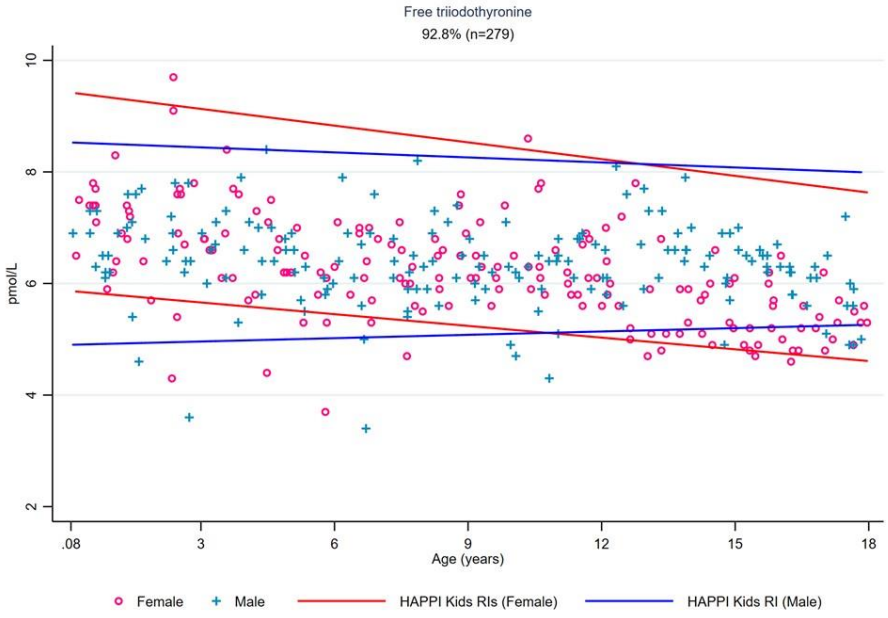

Aspartate amino transferase
96.0% (n=273)
U/L
0
20
40
60
80
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

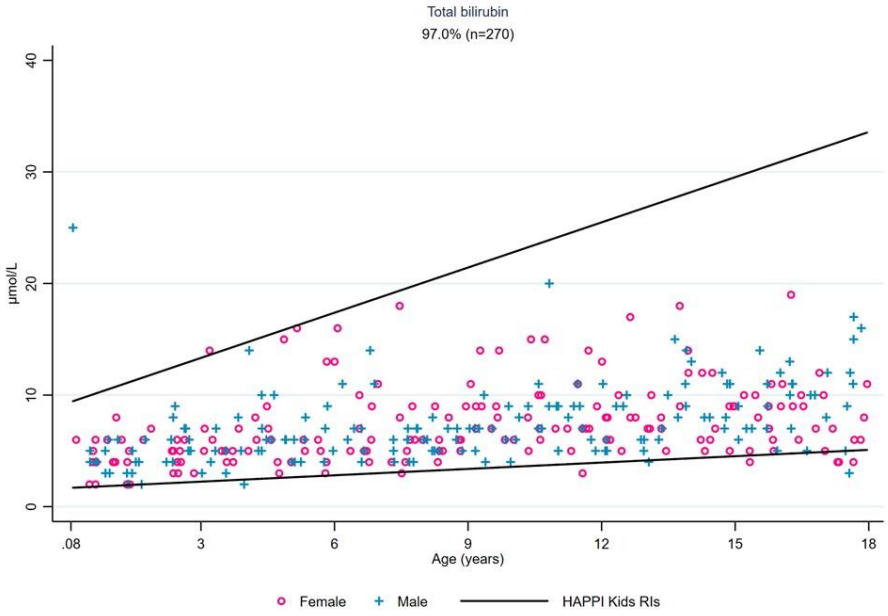

Alanine amino transferase
91.8% (n=269)
U/L
0
10
20
30
40
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

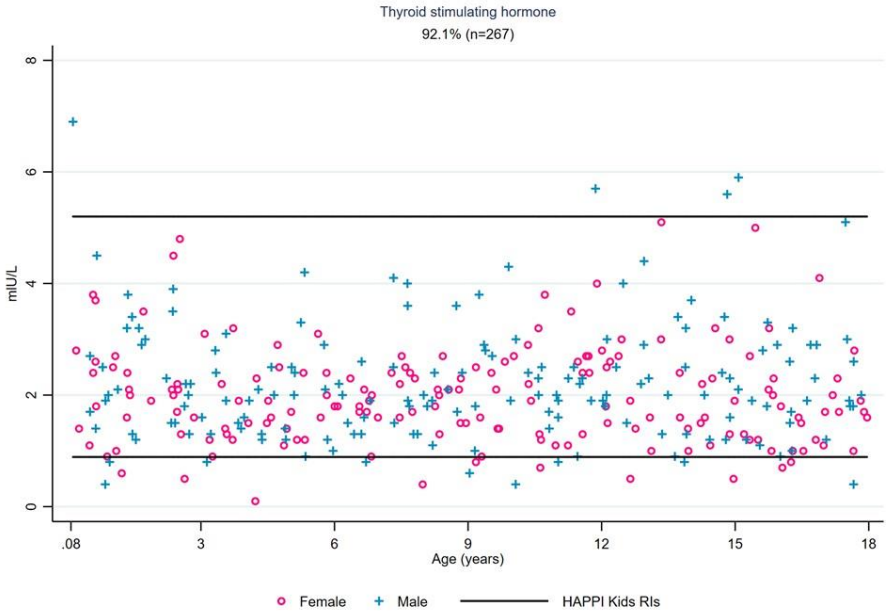

Urea
93.2% (n=279)
mmol/L
2
4
6
8
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

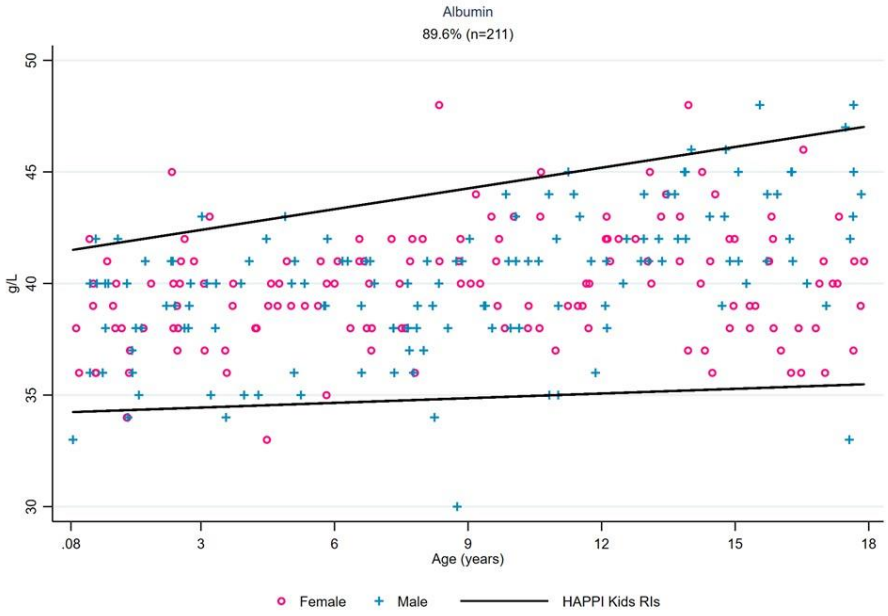

Magnesium
91.7% (n=277)
mmol/L
.7
.8
.9
1
1.1
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

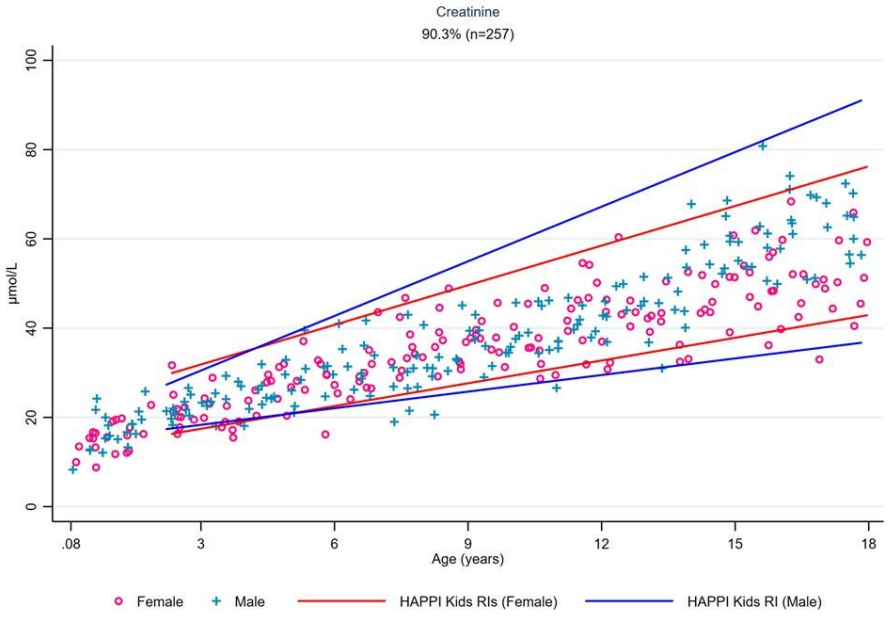

Free thyroxine
95.0% (n=278)
pmol/L
10
15
20
25
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)


Free triiodothyronine
92.8% (n=279)
pmol/L
2
4
6
8
10
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)


Total bilirubin
97.0% (n=270)
µmol/L
0
10
20
30
40
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs


Thyroid stimulating hormone
92.1% (n=267)
mIU/L
0
2
4
6
8
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs


Albumin
89.6% (n=211)
g/L
50
45
40
35
30
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs


Creatinine
90.3% (n=257)
µmol/L
100
80
60
40
20
0
.08
3
6
9
12
15
18
Age (years)
Female
Male
HAPPI Kids RIs (Female)
HAPPI Kids RI (Male)

Appendix D: Supplementary documents of the publication included in Chapter 5 (Hoq *et al*. Statistical methods used in the estimation of age-specific paediatric reference intervals for laboratory blood tests: A systematic review. *Clinical Biochemistry*. Aug 2020.)

Supplementary Document 1: Search strategy

Database(s): Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily 1946 to December 06, 2018

#	Searches	Results
1	(reference adj (value* or standard* or interval* or range* or limit* or curve* or chart*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	227518
2	(centile adj (chart* or curve*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	366
3	1 or 2	227660
4	(construct* or creat* or develop* or establish* or estimat* or calculat* or defin*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	7774889
5	(child or pediatric* or paedatric* or infant* or children or neonat* or Adolescen*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]	3854194
6	(((reference adj (value* or standard* or interval* or range* or limit* or curve* or chart*)) or (centile adj (chart* or curve*))) adj6 (construct* or creat* or develop* or establish* or estimat* or calculat* or defin*)).mp.	9171
7	6 and 5	3277
8	limit 7 to "english language"	2998
9	limit 8 to yr="1993 - Current"	2742

Pediatric total fractionated metanephrine. Age-related reference intervals in spot urine.

Roli L; Veronesi A; De Santis MC; Baraldi E.

Minerva Pediatrica. 2018 Dec 03.

[Journal Article]

UI: 30511558

Authors Full Name

Roli, Laura; Veronesi, Agnese; De Santis, Maria C; Baraldi, Enrica.

Database(s): **Embase Classic+Embase** 1947 to 2018 December 10

Search Strategy:

#	Searches	Results
1	(reference adj (value* or standard* or interval* or range* or limit* or curve* or chart*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	115108
2	(centile adj (chart* or curve*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	511
3	1 or 2	115428
4	(construct* or creat* or develop* or establish* or estimat* or calculat* or defin*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	10509932
5	(child or pediatric* or paedatric* or infant* or children or neonat* or Adolescen*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]	4036264
6	(((reference adj (value* or standard* or interval* or range* or limit* or curve* or chart*)) or (centile adj (chart* or curve*))) adj6 (construct* or creat* or develop* or establish* or estimat* or calculat* or defin*)).mp.	13493
7	6 and 5	3903
8	limit 7 to "english language"	3540
9	limit 8 to yr="1993 - Current"	3348
10	limit 9 to conference abstract status	816
11	9 not 10	2532

Accuracy of Administrative Health Data for Surveillance of Traumatic Brain Injury: A Bayesian Latent Class Analysis.

Lasry O., Dendukuri N., Marcoux J., Buckeridge D.L.

Epidemiology (Cambridge, Mass.). 29 (6) (pp 876-884), 2018. Date of Publication: 01 Nov 2018.

[Article]

Publisher

NLM (Medline)

Supplementary Table 1: Summary of all publications included in the systematic review

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a discrete or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Garcia-Part *et al* [1]	2018	Cross-sectional	Healthy children	Didn't specify	159	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Grunewald *et al* [2]	2018	Cross-sectional	Healthy children	Unused samples from hospital	456	Age-groups	Discrete RIs	Dixon/Reed method and Tukey's method	Formal tests	Robust	Yes	CLSI/IFCC
Jayasuriya *et al* [3]	2018	Cross-sectional	Healthy children	Available test results from Laboratory	569	Age-groups	Discrete RIs	Didn't mention	Formal tests	Both nonparametric and parametric	No	Didn’t mention
Jopling *et al* [4]	2018	Cross-sectional	Children with a particular condition	Unused samples from hospital	704	Age-groups and as a continuous variable	Both discrete and continuous Ris	Didn't mention	Didn't mention	Robust & quantile regression (using non-parametric function of age)	No	Other published paper
Karn *et al* [5]	2018	Cross-sectional	Healthy children	Communities or school, Hospitals	207	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Li *et al* [6]	2018	Cross-sectional	Healthy children	Communities or school, Hospitals	1394	Age-groups	Discrete RIs	Dixon/Reed method	Visually/graphically	Nonparametric	Yes	CLSI/IFCC
Lofving *et al* [7]	2018	Longitudinal	Healthy children	From another study	456	Age-groups	Discrete RIs	Clinically implausible	Visually/graphically	Parametric	No	Other published paper
Oh *et al* [8]	2018	Cross-sectional	Healthy children	From another study	2487	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Peng *et al* [9]	2018	Cross-sectional	Healthy children	Available test results from Laboratory	136546	Age-groups	Discrete RIs	Tukey's method	Visually/graphically	Nonparametric	No	CLSI/IFCC
Ridefelt *et al* [10]	2018	Cross-sectional	Healthy children	From another study	1988	Age-groups	Discrete RIs	Dixon/Reed method and extremely large values	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Strand *et al* [11]	2018	Cross-sectional	Healthy children	Didn't specify	1340	Age-groups and as a continuous variable	Both discrete and continuous RIs	Dixon/Reed method	Didn't mention	Nonparametric & GAMLSS	Yes	Statistical methods paper
Uaariyapanichkul *et al* [12]	2018	Cross-sectional	Healthy children	Communities or school, Hospitals	277	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Wang *et al* [13]	2018	Cross-sectional	Healthy children	Hospitals	1143	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric and other methods	Yes	CLSI/IFCC
Weidhofer *et al* [14]	2018	Cross-sectional	Healthy children	Available test results from Laboratory		Integer ages (overlapping age-groups)	Continuous RIs	Didn't mention	Didn't mention	Indirect algorithm	No	Other published paper
Welter *et al* [15]	2018	Cross-sectional	Healthy children	Communities or school	1067	Age-groups	Discrete RIs	Tukey's method	Formal tests	Nonparametric	Yes	CLSI/IFCC
Ziegelasch *et al* [16]	2019	Longitudinal	Healthy children	From another study	2803	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	GAMLSS	No	Statistical methods paper
Ata *et al* [17]	2018	Cross-sectional	Healthy children	From another study	6116	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS (but *gamlss* package in R)	No	Statistical methods paper
de Melo *et al* [18]	2018	Cross-sectional	Healthy children	Hospitals	246	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Fulgoni *et at* [19]	2019	Cross-sectional	Healthy children	From another study	23831	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric and Piecewise Regression-Derived Reference Intervals	No	Other published paper
Aral *et al* [20]	2018	Cross-sectional	Healthy children	Available test results from Laboratory	11230	Age-groups	Discrete RIs	Extremely large values	Both visually and using format tests	Parametric	No	Didn’t mention
Adeli *et al* [21]	2017	Cross-sectional	Healthy children	Communities or school, Hospitals	2188	Age-groups	Discrete RIs	Tukey's method and extremely large values	Both visually and using format tests	Both nonparametric and robust	Yes	CLSI/IFCC
Gannage-Yared *et al* [22]	2017	Cross-sectional	Healthy children	Communities or school	974	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Gitaka *et al* [23]	2017	Cross-sectional	Healthy children	Communities or school	1070	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Kulle *et al* [24]	2017	Cross-sectional	Healthy children	Unused samples from hospital	4334	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Statistical methods paper
Lorennz *et al* [25]	2017	Cross-sectional	Healthy children	Available test results from Laboratory	512	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn’t mention
MacQueen *et al* [26]	2017	Cross-sectional	Didn’t mention	Available test results from Laboratory	20614	Integer ages	Continuous RIs	Didn't mention	Didn't mention	Did not specify (RIs presented as a chart).	No	Didn’t mention
Narula *et al* [27]	2017	Cross-sectional	Healthy children	Hospitals	403	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric		Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Parkin *et al* [28]	2017	Didn't mention	Healthy children	Hospitals	4334	Age-groups	Discrete RIs	Horn *et al*	Visually/graphically	Others	Yes	CLSI/IFCC
Saravanabhavan *et al* [29]	2017	Cross-sectional	Healthy children	From another study	250	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Sokou *et al* [30]	2017	Cross-sectional	Healthy children	Hospitals	282	Age-groups	Discrete RIs	Didn't mention	Both visually and using format tests	Nonparametric	No	Other published paper
Supriya *et al* [31]	2017	Cross-sectional	Healthy children	Didn't specify	3441	Age-groups	Discrete RIs	Tukey's method and extremely large values	Didn't specify	Nonparametric	No	Other published paper
Wanjian *et al* [32]	2017	Cross-sectional	Healthy children	Hospitals	502	Age-groups	Discrete RIs	Dixon/Reed method	Both visually and using format tests	Both nonparametric and parametric	No	CLSI/IFCC
Zhai *et al* [33]	2017	Cross-sectional	Healthy children	Hospitals	3210	Age-groups	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	No	Didn’t mention
Zierk *et al* [34]	2017	Cross-sectional	Healthy children	Available test results from Laboratory	361405	Integer ages	Both discrete and continuous Ris	Didn't mention	Didn't mention	Nonparametric (centre-wise) & Indirect algorithm.	No	Other published paper
Arslan *et al* [35]	2016	Cross-sectional	Didn’t mention	Hospitals	320	Age-groups	Discrete RIs	Dixon/Reed method	Formal tests	Both nonparametric and robust	Yes	CLSI/IFCC
Berron-Ruiz *et al* [36]	2016	Cross-sectional	Healthy children	Didn't specify	112	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Choi *et al* [37]	2016	Cross-sectional	Healthy children	From another study	1585	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric method but geometric mean instead of arithmetic mean	Yes	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Cuerq *et al* [38]	2016	Didn't mention	Healthy children	Hospitals	156	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Robust	No	CLSI/IFCC
Demirdjian *et al* [39]	2019	Cross-sectional	Healthy children	Available test results from Laboratory	864	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Robust	Yes	Other published paper
Fuzukzumi *et al* [40]	2016	Cross-sectional	Didn’t mention	Hospitals	283	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Gennai *et al* [41]	2016	Cross-sectional	Children with a particular condition	Hospitals	1502	Age-groups and as a continuous variable	Both discrete and continuous Ris	Didn't mention	Formal tests	Nonparametric and regression based method using parametric function of age	No	Didn’t mention
Kim *et al* [42]	2016	Cross-sectional	Children with a particular condition	Hospitals	413	Age-groups	Discrete RIs	Didn't mention	Formal tests	Robust	Yes	CLSI/IFCC
Kira *et al* [43]	2016	Cross-sectional	Healthy children	Communities or school	538	Age-groups	Discrete RIs	Didn't mention	Formal tests	Both nonparametric and parametric	Yes	Didn’t mention
La’ulu *et al* [44]	2016	Cross-sectional	Healthy children	From another study	2540	Age-groups	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	Yes	Didn’t mention
Lausten-Thomsen *et al* [45]	2016	Cross-sectional	Healthy children	Didn't specify	1196	As a continuous variable	Continuous RIs	Didn't mention	Formal tests	GAMLSS	No	Statistical methods paper
Lopez-Ruzafa *et al* [46]	2016	Cross-sectional	Healthy children	From another study	1037	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	Yes	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
MacQueen *et al* [47]	2016	Cross-sectional	Healthy children	Available test results from Laboratory	5281	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Okebe *et al* [48]	2016	Cross-sectional	Healthy children	Communities or school	1261	As a continuous	Continuous RIs	Tukey's method and based on Q-Q plot	Didn't specify	GAMLSS	No	Statistical methods paper
Shashaj *et al* [49]	2016	Cross-sectional	Healthy children	Hospitals	2573	Integer ages	Continuous RIs	Clinically implausible	Didn't mention	LMS	No	Didn’t mention
Toulon *et al* [50]	2016	Cross-sectional	Healthy children	Hospitals	1437	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	CLSI/IFCC
Bateman *et al* [51]	2015	Longitudinal	Children with a particular condition	Hospitals	218	Integer ages	Continuous RIs	Didn't mention	Didn't mention	Locally weighted scatterplot smoothing (LOWESS)	No	Statistical methods paper
Castagna *et al* [52]	2015	Cross-sectional	Healthy children	Hospitals	2740	Integer ages	Continuous RIs	Didn't mention	Didn't specify	LMS	No	Statistical methods paper
Franscini *et al* [53]	2015	Didn't mention	Healthy children	Hospitals		As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Regression based method using parametric function of age	No	Statistical methods paper
Ghasemi *et al* [54]	2015	Cross-sectional	Healthy children	From another study	1502	Age-groups	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	No	CLSI/IFCC
Jia *et al* [55]	2015	Cross-sectional	Healthy children	Communities or school	1027	Integer ages	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	No	Other published paper

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Lausten-Thomsen *et al* [56]	2015	Cross-sectional	Healthy children	Communities or school	1193	Integer ages	Both discrete and continuous Ris	Didn't mention	Formal tests	Nonparametric and regression based method using parametric function of age	Yes	Didn't mention
Perez-de-Heredia *et al* [57]	2015	Cross-sectional	Healthy children	Communities or school	1089	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn't mention
Schlapbach *et al* [58]	2015	Cross-sectional	Healthy children	Hospitals	311	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn't mention
Slhessarenko *et al* [59]	2015	Cross-sectional	Healthy children	Communities or school	1994	Age-groups	Discrete RIs	+/- 3SD	Visually/graphically	Parametric	No	CLSI/IFCC
Soeby *et al* [60]	2015	Cross-sectional	Healthy children	Available test results from Laboratory	11702	Age-groups	Discrete RIs	Visually and by fitting a linear model	Visually/graphically	Nonparametric	Yes	CLSI/IFCC
Teixeira *et al* [61]	2015	Cross-sectional	Healthy children	Unused samples from hospital	311	Age-groups	Discrete RIs	Dixon/Reed method	Both visually and using format tests	Robust	Yes	CLSI/IFCC
Temel *et al* [62]	2015	Cross-sectional	Healthy children	Hospitals	21897	Integer ages	Both discrete and continuous Ris	Didn't mention	Didn't mention	Nonparametric & Multivariate Additive Regression Spline (MARS)	No	Statistical methods paper
Tosato *et al* [63]	2015	Cross-sectional	Healthy children	Hospitals	195	Age-groups	Discrete RIs	Dixon/Reed method	Formal tests	Nonparametric	No	Didn't mention
Uemura *et al* [64]	2015	Cross-sectional	Healthy children	Hospitals	1137	Age-groups and Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn't mention
Yamamoto *et al* [65]	2015	Cross-sectional	Healthy children	Didn't specify	118	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Zierk *et al* [66]	2015	Cross-sectional	Healthy children	Available test results from Laboratory		Integer ages (overlapping age-groups)	Both discrete and continuous Ris	Didn't mention	Didn't mention	Kernel method (nonparametric) & using smooth spline method in R	No	Statistical methods paper
Zwiers *et al* [67]	2015	Cross-sectional	Healthy children	Hospitals	106	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	GAMLSS	No	Statistical methods paper
Cho *et al* [68]	2014	Cross-sectional	Healthy children	Hospitals	2474	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	CLSI/IFCC
Bidlingmaier *et al* [69]	2014	Didn't mention	Healthy children	From another study	4252	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
Christensen *et al* [70]	2014	Cross-sectional	Children with a particular condition	Hospitals	175	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Both nonparametric and parametric	No	Didn’t mention
Duchamp *et al* [71]	2014	Cross-sectional	Healthy children	Hospitals	242	Age-groups	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	No	Didn’t mention
Erhardt *et al* [72]	2014	Cross-sectional	Healthy children	From another study	539	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	GAMLSS	No	Statistical methods paper
Friedrich *et al* [73]	2014	Didn't mention	Didn’t mention	From another study	4218	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
Greaves *et al* [74]	2014	Cross-sectional	Children with a particular condition	Communities or school	107	Age-groups	Discrete RIs	Didn't mention	Visually/graphically	Parametric	No	Didn’t mention
Hahn *et al* [75]	2014	Cross-sectional	Didn’t mention	Didn't specify	883	Age-groups	Discrete RIs	Tukey's method	Formal tests	Parametric	Yes	CLSI/IFCC

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Hirfanoglu *et al* [76]	2014	Cross-sectional	Healthy children	Hospitals	583	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Hu *et al* [77]	2014	Cross-sectional	Healthy children	Hospitals	308	Integer ages	Continuous RIs	Didn't mention	Formal tests	Regression based method using parametric function of age	No	Didn’t mention
Koester-Weber *et al* [78]	2014	Cross-sectional	Healthy children	Communities or school	972	Integer ages	Continuous RIs	Didn't mention	Didn't specify	LMS	No	Didn’t mention
Loh *et al* [79]	2014	Cross-sectional	Healthy children	Available test results from Laboratory	47311	Integer ages	Continuous RIs	Clinically implausible	Didn't mention	LMS	No	Statistical methods paper
Macit *et al* [80]	2014	Cross-sectional	Healthy children	Didn't specify	945	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Moraes-Pinto *et al* [81]	2014	Cross-sectional	Healthy children	Didn't specify	408	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Odutola *et al* [82]	2014	Cross-sectional	Healthy children	Communities or school	717	As a continuous	Continuous RIs	Tukey's method and extremely large values	Didn't specify	Regression based method using parametric function of age	No	Statistical methods paper
Oh *et al* [83]	2014	Cross-sectional	Children with a particular condition	Available test results from Laboratory	331	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Didn’t mention
Savino *et al* [84]	2014	Cross-sectional	Healthy children	Hospitals	317	Age-groups	Discrete RIs	Didn't mention	Formal tests	Quantile regression	Yes	Didn’t mention
Schlenz *et al* [85]	2014	Cross-sectional	Healthy children	Communities or school	16228	As a continuous variable	Continuous RIs	Extremely large values	Didn't mention	GAMLSS	No	Statistical methods paper

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Shaw *et al* [86]	2014	Cross-sectional	Didn't mention	Available test results from Laboratory	200627	Age-groups	Discrete RIs	Tukey's method	Didn't specify	Modified Hoffmann approach	No	Statistical methods paper
Sheikhbahaei *et al* [87]	2014	Cross-sectional	Healthy children	Hospitals	246	Age-groups	Discrete RIs	Extremely large values	Formal tests	Nonparametric	No	Didn't mention
Sun *et al* [88]	2014	Cross-sectional	Children with a particular condition	Hospitals	308	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Voss *et al* [89]	2014	Cross-sectional	Healthy children	Communities or school	140	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric	No	Didn't mention
Wolters *et al* [90]	2014	Cross-sectional	Healthy children	From another study	1653	As a continuous	Continuous RIs	Didn't mention	Didn't mention	GAMLSS	No	Statistical methods paper
Yi *et al* [91]	2014	Cross-sectional	Healthy children	From another study	2116	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn't mention
Mortensen *et al* [92]	2014	Didn't mention	Healthy children	From another study	7841	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention
Aldrimer *et al* [93]	2013	Cross-sectional	Healthy children	Communities or school, Hospitals	689	Age-groups	Discrete RIs	Dixon/Reed method	Didn't specify	Nonparametric	No	CLSI/IFCC
Aitkenhead *et al* [94]	2013	Cross-sectional	Healthy children	Available test results from Laboratory	2369	Age-groups	Discrete RIs	Dixon/Reed method	Didn't specify	Both nonparametric and parametric	Yes	CLSI/IFCC
Biino *et al* [95]	2013	Cross-sectional	Healthy children	Communities or school	43807	Age-groups	Discrete RIs	Dixon/Reed method and extremely large values	Visually/graphically	Both nonparametric and parametric	Yes	CLSI/IFCC

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Can *et al* [96]	2013	Cross-sectional	Healthy children	Didn't specify	4102	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Guven *et al* [97]	2013	Cross-sectional	Healthy children	Hospitals	571	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Heinze *et al* [98]	2013	Cross-sectional	Healthy children	Hospitals	100	Integer ages	Continuous RIs	Resampling	Didn't mention	LMS	No	Statistical methods paper
Hilsted *et al* [99]	2013	Cross-sectional	Healthy children	Hospitals	2141	Age-groups	Discrete RIs	Dixon/Reed method	Formal tests	Nonparametric	No	Other published paper
Joncquel-Chevalier Curt *et al* [100]	2013	Cross-sectional	Children with a particular condition	Didn't specify	6334	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn’t mention
Kulle *et al* [101]	2013	Cross-sectional	Healthy children	Unused samples from hospital	754	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Mendez *et al* [102]	2013	Cross-sectional	Didn’t mention	Hospitals	73	Age-groups	Discrete RIs	Dixon/Reed method	Both visually and using format tests	Nonparametric	Yes	
Morovat *et al* [103]	2013	Cross-sectional	Healthy children	Didn't specify	407	Age-groups	Discrete RIs	Tukey's method	Formal tests	Both nonparametric and parametric	Yes	Didn’t mention
Osta *et al* [104]	2013	Cross-sectional	Healthy children	Didn't specify	200	Age-groups	Discrete RIs	Didn't mention	Formal tests	Not specified	No	Didn’t mention
Radicioni *et al* [105]	2013	Cross-sectional	Healthy children	Hospitals	692	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	CLSI/IFCC
Tauseef Bukhari *et al* [106]	2013	Cross-sectional	Healthy children	Hospitals	2000	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Zhu *et al* [107]	2013	Cross-sectional	Healthy children	Hospitals	247	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Both nonparametric and parametric	Yes	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Eisenhofer *et al* [108]	2013	Cross-sectional	Children with a particular condition	Didn't specify	1226	As a continuous	Continuous RIs	Didn't mention	Didn't specify	Quantile regression (using parametric function of age)	No	Didn’t mention
Zec *et al* [109]	2012	Cross-sectional	Healthy children	Hospitals	966	Age-groups	Discrete RIs	Dixon/Reed method and Tukey's method	Didn't specify	Parametric and nonparametric method	Yes	CLSI/IFCC
Troy *et al* [110]	2012	Longitudinal	Healthy children	Didn't specify	269	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Shatat *et al* [111]	2012	Cross-sectional	Healthy children	From another study	1725	Age-groups	Discrete RIs	Didn't mention	Didn't mention	quantile regression	No	Didn’t mention
Shahabi *et al* [112]	2012	Cross-sectional	Healthy children	From another study	792	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Ridefelt *et al* [113]	2012	Cross-sectional	Healthy children	Communities or school, Hospitals	692	Age-groups	Discrete RIs	Dixon/Reed method	Didn't specify	Nonparametric	No	CLSI/IFCC
Olney *et al* [114]	2012	Cross-sectional	Healthy children	Communities or school	258	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
Liu *et al* [115]	2012	Cross-sectional	Healthy children	Hospitals	4044	Integer ages	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	No	Didn’t mention
Lin *et al* [116]	2012	Cross-sectional	Healthy children	Communities or school	2115	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Lem *et al* [117]	2012	Cross-sectional	Healthy children	Hospitals	512	As a continuous variable	Continuous RIs	Didn't mention	Formal tests	LMS	No	Statistical methods paper
Kardar *et al* [118]	2012	Cross-sectional	Healthy children	Hospitals	800	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Isojima *et al* [119]	2012	Cross-sectional	Healthy children	From another study	586	Integer ages	Continuous RIs	Extremely large values	Didn't mention	LMS	No	Statistical methods paper
Hyun *et al* [120]	2012	Cross-sectional	Healthy children	Hospitals	1378	Integer ages and age groups (tanner stage)	Both discrete and continuous Ris	Didn't mention	Didn't mention	Nonparametric & quantile regression (using nonparametric function of age)	No	Didn't mention
Gonzalez-Gross *et al* [121]	2012	Cross-sectional	Healthy children	Communities or school	1051	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
D'Apolito *et al* [122]	2012	Cross-sectional	Healthy children	Unused samples from hospital	131	Age-groups	Discrete RIs	Tukey's method	Didn't specify	Both nonparametric and parametric	No	Didn't mention
Chaler *et al* [123]	2012	Cross-sectional	Didn't mention	Available test results from Laboratory	7581	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Using Statistic 9 program	No	Didn't mention
Cangemi *et al* [124]	2012	Cross-sectional	Didn't mention	Available test results from Laboratory	3127	Integer ages and continuous age	Both discrete and continuous Ris	Using the CBstata program	Formal tests	Nonparametric and regression based method using parametric function of age	No	Other published paper
Bedogni *et al* [125]	2012	Cross-sectional	Didn't mention	Hospitals	24403	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Quantile regression (using parametric function of age)	No	Statistical methods paper
Appel *et al* [126]	2012	Cross-sectional	Healthy children	Didn't specify	218	Age-groups	Discrete RIs	Didn't mention	Didn't mention	did not specify	No	Didn't mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Yuksel *et al*[127]	2011	Cross-sectional	Healthy children	Communities or school	566	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention
Yi *et al* [128]	2011	Cross-sectional	Healthy children	Didn't specify	108	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Nonparametric	No	Didn't mention
Uemura *et al*[129]	2011	Cross-sectional	Didn't mention	Hospitals	1151	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn't mention
Turan *et al* [130]	2011	Cross-sectional	Healthy children	Hospitals	1741	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric	No	Didn't mention
Tamimi *et al* [131]	2011	Cross-sectional	Healthy children	Communities or school	1168	Age-groups	Discrete RIs	based on transformation	Didn't specify	Parametric	No	Didn't mention
Roehrl *et al* [132]	2011	Cross-sectional	Didn't mention	Available test results from Laboratory	2571	Age-groups	Discrete RIs	Tukey's method	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Marwaha *et al* [133]	2011	Cross-sectional	Healthy children	Communities or school	3076	Integer ages	Continuous RIs	Extremely large values	Didn't mention	LMS	No	Statistical methods paper
Huang *et al* [134]	2011	Cross-sectional	Didn't mention	Available test results from Laboratory	1769	Age-groups	Discrete RIs	Tukey's method	Didn't specify	Both nonparametric and robust	Yes	CLSI/IFCC
Verburg *et al* [135]	2011	Cross-sectional	Healthy children	Unused samples from hospital	654	Integer ages	Continuous RIs	removed 10 outliers - But did not mention	Formal tests	LMS	No	Statistical methods paper
Clifford *et al* [136]	2011	Cross-sectional	Healthy children	Didn't specify	1765	Age-groups	Discrete RIs	Outliers were eliminated manually	Both visually and using format tests	Nonparametric	Yes	CLSI/IFCC
Bouvier *et al* [137]	2011	Cross-sectional	Healthy children	Hospitals	236	Integer ages and a continuous variable	Both discrete and continuous Ris	Didn't mention	Didn't mention	Parametric & Quantile regression (nonparametric function of age)	No	Didn't mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Alberti *et al* [138]	2011	Cross-sectional	Healthy children	From another study	1119	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Regression based method using parametric function of age	No	Statistical methods paper
Langer *et al* [139]	2011	Cross-sectional	Healthy children	Hospitals	150	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Not specified	No	Didn’t mention
Xu *et al* [140]	2010	Cross-sectional		Communities or school	837	Integer ages	Continuous RIs	Clinically implausible	Didn't specify	Regression based method using parametric function of age	No	Didn’t mention
Takala *et al* [141]	2010	Longitudinal	Healthy children	Hospitals	150	As a continuous variable	Continuous RIs	Values with the absolute value of residuals that were > 3 from the calculated RIs.	Visually/graphically	Regression based method using parametric function of age	No	Other published paper
Southcott *et al* [142]	2010	Longitudinal	Healthy children	Communities or school	852	Age-groups	Discrete RIs	Clinically implausible	Didn't mention	Nonparametric	No	Didn’t mention
Salameh *et al* [143]	2010	Cross-sectional	Healthy children	Communities or school	499	Age-groups	Discrete RIs	Residual	Didn't mention	Segmented Regression	No	Didn’t mention
Piatosa *et al* [144]	2010	Cross-sectional	Didn’t mention	Didn't specify	264	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Oswald *et al* [145]	2010	Didn't mention	Healthy children	Hospitals	407	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	CLSI/IFCC
Morbach *et al* [146]	2010	Cross-sectional	Healthy children	Hospitals	158	Age-groups	Both discrete and continuous Ris	Didn't mention	Didn't mention	Nonparametric & GAM where a smooth spline was used	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Marwaha *et al* [147]	2010	Cross-sectional	Healthy children	From another study	3700	Integer ages	Discrete RIs	Visually by histogram and confirmed by Leverage statistics, Cook's distance and Jack-knife residuals.	Didn't specify	Both nonparametric and parametric	No	Didn’t mention
Kushnir *et al*[148]	2010	Cross-sectional	Healthy children	Didn't specify	2157	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Robust	No	Statistical methods paper
Klarmann *et al* [149]	2010	Cross-sectional	Healthy children	Didn't specify	592	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	CLSI/IFCC
Fideleff *et al* [150]	2010	Cross-sectional	Didn’t mention	Didn't specify	338	Age-groups	Discrete RIs	Did not mention the method	Formal tests	Nonparametric	No	Didn’t mention
Dawson *et al* [151]	2010	Cross-sectional	Healthy children	Hospitals	813	Integer ages	Continuous RIs	Didn't mention	Didn't specify	LMS	No	Statistical methods paper
Christensen *et al* [152]	2010	Cross-sectional	Didn’t mention	Available test results from Laboratory	63371	Integer ages	Continuous RIs	Didn't mention	Didn't mention	Not specified (Regression based parametric function of age from graph).	No	Didn’t mention
Buchanan *et al* [153]	2010	Cross-sectional	Healthy children	Hospitals	655	Age-groups	Discrete RIs	Dixon/Reed method	Didn't mention	Nonparametric	Yes	CLSI/IFCC
Boer *et al* [154]	2010	Cross-sectional		Available test results from Laboratory	12205	As a continuous	Continuous RIs	Didn't mention	Didn't mention	The mean values are estimated in three groups based on prior knowledge.	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Wu *et al* [155]	2009	Cross-sectional	Didn't mention	Hospitals, Available test results from Laboratories	337	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention
Wiedmeier *et al* [156]	2009	Cross-sectional		Hospitals, Available test results from Laboratories	34146	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn't mention
Sfar *et al* [157]	2009	Cross-sectional	Didn't mention	Communities or school	200	A continuous variable	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Didn't mention
Ooi *et al* [158]	2009	Cross-sectional	Healthy children	Hospitals	436	Age-groups	Discrete RIs	Didn't mention	Formal tests	Both nonparametric and parametric	No	Didn't mention
Lai *et al*[159]	2009	Cross-sectional	Didn't mention	Communities or school	4326	Age-groups	Discrete RIs	Extremely large values	Didn't specify	Nonparametric	No	Didn't mention
Jopling *et al* [160]	2009	Cross-sectional	Didn't mention	Available test results from Laboratory	22621	Integer ages	Discrete RIs	Extremely large values	Didn't mention	Nonparametric	No	Didn't mention
Johnson-Davis *et al* [161]	2009	Cross-sectional	Healthy children	Didn't specify	1136	Age-groups	Discrete RIs	Didn't mention	Visually/graphically	Nonparametric and other	Yes	CLSI/IFCC
Heiduk *et al* [162]	2009	Cross-sectional	Didn't mention	Unused samples from hospital	833	Age-groups and as a continuous variable	Both discrete and continuous Ris	Extremely large values	Didn't mention	Nonparametric & quantile regression (using nonparametric function of age)	No	CLSI/IFCC

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Adetifa *et al*[163]	2009	Cross-sectional	Healthy children	Hospitals	1279	As a continuous variable	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Statistical methods paper
Abrahamsson *et al* [164]	2009	Longitudinal	Children with a particular condition	Hospitals	471	As a continuous variable	Continuous RIs	Didn't mention	Didn't specify	Piecewise linear regression method	No	Didn’t mention
Thayyil *et al* [165]	2008	Cross-sectional	Didn’t mention	Available test results from Laboratory	172	Integer ages	Discrete RIs	Didn't mention	Didn't specify	Only mean	No	Didn’t mention
Soldin *et al*[166]	2008	Cross-sectional	Didn’t mention	Available test results from Laboratory	1352	Age-groups	Discrete RIs	Hoffman	Didn't specify	Hoffman	No	Statistical methods paper
Shine, B *et al* [167]	2008	Cross-sectional	Didn’t mention	Available test results from Laboratory	75328	Integer ages	Both discrete and continuous Ris	Hoaglin (Performance of some resistant rules for outlier labelling. J Am Stat Assoc 1986;81:991-999).	Didn't mention	Nonparametric & LMS	No	Statistical methods paper
Kushnir *et al* [168]	2008	Cross-sectional	Healthy children	Didn't specify	319	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric and other	No	Didn’t mention
Kratzsch *et al* [169]	2008	Cross-sectional	Healthy children	From another study	1004	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Nonparametric	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Sobrero *et al* [170]	2007	Cross-sectional	Didn't mention	Didn't specify	228	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Not specified	Yes	Didn't mention
Rauchenzauner *et al* [171]	2007	Cross-sectional	Healthy children	Hospitals	572	As a continuous	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Statistical methods paper
Lippi *et al* [172]	2007	Cross-sectional	Didn't mention	Available test results from Laboratory	1758	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric method but geometric mean instead of arithmetic mean	No	Didn't mention
Koskas *et al* [173]	2007	Cross-sectional	Didn't mention	Unused samples from hospital	696	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Not specified	No	Didn't mention
Kong *et al* [174]	2007	Cross-sectional	Healthy children	Communities or school	1692	Integer ages	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
Arndt *et al* [175]	2007	Cross-sectional	Didn't mention	Available test results from Laboratory	128	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Mean and percentiles	No	Didn't mention
Akuyam *et al* [176]	2007	Cross-sectional	Healthy children	Hospitals	115	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric	No	Didn't mention
Kapelari *et al* [177]	2008	Cross-sectional	Didn't mention	Available test results from Laboratory	1209	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn't mention
Castellani *et al* [178]	2008	Cross-sectional	Didn't mention	Hospitals	394	As a continuous variable	Continuous RIs	Tukey's method	Didn't specify	Quantile regression (using nonparametric function of age)	No	Other published paper

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Yip *et al* [179]	2006	Cross-sectional	Didn't mention	Unused samples from hospital	499	Age-groups	Discrete RIs	Chauvenet's criteria	Didn't mention	Nonparametric	No	Didn't mention
Vietri *et al* [180]	2006	Cross-sectional	Healthy children	Communities or school	2288	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention
Ritchie *et al* [181]	2006	Cross-sectional	Healthy children	Available test results from Laboratory	27635	As a continuous variable	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Didn't mention
Quinto *et al* [182]	2006	Cross-sectional	Healthy children	Available test results from Laboratory	2226	Integer ages and as a continuous variable	Both discrete and continuous Ris	Dixon/Reed method	Didn't specify	Nonparametric & quantile regression (using parametric function of age)	No	Statistical methods paper
Monagle *et al* [183]	2006	Cross-sectional	Healthy children	Hospitals	400	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn't mention
Berrahmoune *et al* [184]	2006	Cross-sectional	Healthy children	From another study	340	Age-groups	Discrete RIs	Didn't mention	Visually/graphically	Nonparametric	No	Didn't mention
Kempers *et al* [185]	2005	Cross-sectional	Children with a particular condition	Hospitals	142	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric	No	Didn't mention
Chailurkit *et al* [186]	2005	Cross-sectional	Healthy children	Communities or school	600	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric method but median instead of arithmetic mean	No	Didn't mention
Lugada *et al* [187]	2004	Cross-sectional	Didn't mention	Communities or school	3421	Age-groups	Discrete RIs	Didn't mention	Both visually and using format tests	Nonparametric	No	Didn't mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Elmlinger *et al* [188]	2004	Cross-sectional	Healthy children	Didn't specify	1584	Integer ages	Continuous RIs	Didn't mention	Didn't mention	Local regression analysis (using S-Plus)	No	Didn’t mention
Brinkworth *et al* [189]	2004	Cross-sectional	Healthy children	Available test results from Laboratory	2965	Integer ages	Discrete RIs	Didn't mention	Formal tests	Both nonparametric and parametric	No	Didn’t mention
Schauer *et al* [190]	2003	Cross-sectional	Healthy children	Didn't specify	235	Age-groups	Continuous RIs	Didn't mention	Didn't mention	Box-Cox transformation, mean +/- SD and polynomial regression for centile	No	Other published paper
Perrone *et al* [191]	2003	Cross-sectional	Children with a particular condition	Hospitals	101	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Najam *et al* [192]	2003	Cross-sectional	Didn’t mention	Available test results from Laboratory	1148	Age-groups	Discrete RIs	Did not specify	Didn't specify	Both nonparametric and parametric	No	Didn’t mention
de paula *et al* [193]	2003	Cross-sectional	Healthy children	Hospitals	463	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Statistical methods paper
Brabant [194]	2003	Cross-sectional	Healthy children	Available test results from Laboratory	1469	As a continuous variable	Continuous RIs	Didn't mention	Visually/graphically	Regression based method using parametric function of age	No	Didn’t mention
Blomme *et al* [195]	2003	Cross-sectional	Children with a particular condition	Available test results from Laboratory	151	Age-groups	Discrete RIs	Clinically implausible	Didn't mention	Nonparametric	No	CLSI/IFCC

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Schlebusch[196]	2002	Cross-sectional	Didn’t mention	Hospitals	521	Integer ages	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Muntau[197]	2002	Cross-sectional	Healthy children	Hospitals	1010	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Kawahara [198]	2002	Cross-sectional	Healthy children	Hospitals	228	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric	No	Didn’t mention
Hubner [199]	2002	Cross-sectional	Healthy children	Hospitals	460	As a continuous variable	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Statistical methods paper
Crofton *et al* [200]	2002	Cross-sectional	Didn’t mention	Unused samples from hospital	542	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Both nonparametric and parametric	No	Didn’t mention
Lagerstedt *et al* [201]	2001	Cross-sectional	Didn’t mention	Unused samples from hospital	196	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Not specified	No	Didn’t mention
Georges-Gobinet *et al* [202]	2001	Cross-sectional	Healthy children	Hospitals	242	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Nonparametric	No	Didn’t mention
Elmlinger *et al* [203]	2001	Cross-sectional	Healthy children	From another study	742	Integer ages (with a few integer ages combined)	Both discrete and continuous Ris	Didn't mention	Didn't specify	Nonparametric and regression based method using parametric function of age	No	CLSI/IFCC
Clark *et al* [204]	2001	Cross-sectional	Healthy children	Hospitals	120	Age-groups	Discrete RIs	Grubb's Test	Formal tests	Parametric	No	Didn’t mention
Ahmed *et al* [205]	2001	Cross-sectional	Healthy children	Communities or school	261	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Harmoinen *et al* [206]	2000	Cross-sectional	Healthy children	Hospitals	407	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	Yes	Other published paper

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Haraldsson *et al* [207]	2000	Didn't mention	Healthy children	Hospitals	184	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Nonlinear regression by the SPSS Statistical package	No	Didn’t mention
Finney *et al* [208]	2000	Cross-sectional		Available test results from Laboratory	291	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric	No	Didn’t mention
Sherriff *et al* [209]	1999	Cross-sectional	Didn’t mention	Communities or school	1213	Age-groups	Discrete RIs	Didn't mention	Both visually and using format tests	Nonparametric	No	Didn’t mention
Shahabuddin *et al* [210]	1998	Cross-sectional	Healthy children	Hospitals	102	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Ngamphaiboon *et al* [211]	1998	Cross-sectional	Healthy children	Hospitals	195	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Jagarinec [212]	1998	Cross-sectional	Healthy children	Communities or school	996	Age-groups	Discrete RIs	more than 3sd	Didn't mention	Nonparametric	No	Other published paper
Cullen *et al* [213]	1998	Cross-sectional		Communities or school	100	Age-groups	Discrete RIs	did not specify	Visually/graphically	Parametric	No	Didn’t mention
Bokenkamp *et al* [214]	1998	Cross-sectional	Didn’t mention	Available test results from Laboratory	258	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	CLSI/IFCC
Blohm *et al* [215]	1998	Cross-sectional	Didn’t mention	Unused samples from hospital	390	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Regression to estimate the mean with 95% CI	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Winklhofer-Roob *et al* [216]	1997	Cross-sectional	Healthy children	Communities or school	208	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Regression based method using parametric function of age	No	Statistical methods paper
Rukgauer *et al* [217]	1997	Cross-sectional	Didn’t mention	Hospitals	137	Age-groups	Discrete RIs	Didn't mention	Formal tests	Parametric	No	Didn’t mention
Khan *et al* [218]	1997	Cross-sectional	Didn’t mention	Communities or school	460	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Blum *et al* [219]	1997	Cross-sectional	Healthy children	Didn't specify	713	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Parametric	No	Didn’t mention
Rauch *et al* [220]	1996	Cross-sectional	Children with a particular condition	Didn't specify	133	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Nonparametric	No	Didn’t mention
Jensen *et al* [221]	1995	Cross-sectional	Children with a particular condition	Communities or school	6688	As a continuous variable	Continuous RIs	Didn't mention	Didn't specify	Regression based method using parametric function of age	No	Statistical methods paper
Adams *et al* [222]	1995	Cross-sectional	Didn’t mention	Hospitals	170	Age-groups	Discrete RIs	Grubb's Test	Formal tests	Parametric	No	CLSI/IFCC
Miles *et al* [223]	1994	Cross-sectional	Children with a particular condition	Hospitals	180	As a continuous variable	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Statistical methods paper
Miles *et al* [224]	1994	Didn't mention		Communities or school	180	Age-groups	Discrete RIs	Didn't mention	Didn't specify	Nonparametric	No	Didn’t mention
Lau *et al* [225]	1993	Cross-sectional	Healthy children	Communities or school, Hospitals	350	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	LMS	No	Statistical methods paper
Jonetz-Mentzel *et al* [226]	1993	Cross-sectional	Didn’t mention	Didn't specify	687	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Kumar *et al* [227]	2003	Cross-sectional	Didn’t mention	Communities or school	586	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Nonparametric	No	Didn’t mention
Elmlinger *et al* [228]	1999	Cross-sectional	Healthy children	Didn't specify	145	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn’t mention
Juul *et al* [229]	1995	Cross-sectional	Healthy children	Communities or school	921	As a continuous variable	Continuous RIs	Didn't mention	Didn't mention	Mean and variance curve was estimated with a smoothing spline	No	Didn’t mention
wood *et al* [230]	2004	Cross-sectional	Didn’t mention	Available test results from Laboratory	115755	Age-groups	Discrete RIs	Didn't mention	Formal tests	Nonparametric	No	Didn’t mention
Genc *et al* [231]	1997	Cross-sectional	Healthy children	Communities or school	1967	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Didn’t mention
Custer *et al* [232]	1995	Cross-sectional	Didn’t mention	Available test results from Laboratory	22464	Integer ages	Continuous RIs	Didn't mention	Didn't specify	LMS	No	Statistical methods paper
Berkel *et al* [233]	1994	Cross-sectional	Healthy children	Hospitals	329	Age-groups and as a continuous variable	Both discrete and continuous Ris	Didn't mention	Didn't mention	Parametric and regression based method using parametric function of age	No	CLSI/IFCC
Theunis *et al* [234]	1993	Cross-sectional	Healthy children	Hospitals	181	Age-groups	Discrete RIs	Didn't mention	Visually/graphically	Nonparametric	No	CLSI/IFCC

First author	Year	Study design	Study pop^{n}	Source	No. of sample	Age as a categorical or continuous variable	Type of RIs	Identification of outliers	Assessment of distribution	Method	95 CI of the RI	Provided justification or reference to the Statistical method
Tommasi *et al* [235]	1996	Cross-sectional	Healthy children	Hospitals	241	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Parametric	No	Didn’t mention
Wiedenann *et al* [236]	1993	Cross-sectional	Healthy children	Didn't specify	714	Age-groups	Discrete RIs	Obvious	Formal tests	Nonparametric		CLSI/IFCC
Brown *et al* [237]	1994	Longitudinal	Children with a particular condition	Hospitals	694	Integer ages	Both discrete and continuous RIs	Didn't mention	Didn't mention	Nonparametric and regression based method using parametric function of age	No	Statistical methods paper
Kogner *et al* [238]	1994	Cross-sectional	Healthy children	Hospitals	112	Age-groups	Discrete RIs	Didn't mention	Didn't mention	Both nonparametric and parametric	No	Statistical methods paper

Reference

[1] M. Garcia-Prat, D. Alvarez-Sierra, A. Aguilo-Cucurull, S. Salgado-Perandres, S. Briongos-Sebastian, C. Franco-Jarava, A. Martin-Nalda, R. Colobran, I. Montserrat, M. Hernandez-Gonzalez, R. Pujol-Borrell, P. Soler-Palacin, M. Martinez-Gallo, Extended immunophenotyping reference values in a healthy pediatric population, Cytometry Part B, Clinical Cytometry 17 (2018) 17.

[2] O. Grunewald, B. Lopez, S. Brabant, S. Rogeau, A. Deschildre, V. Phrommavanh, M. Lefort, E. Moitrot, D. Gyselinckx, A.S. Deleplancque, G. Lefevre, M. Labalette, S. Dubucquoi, Immunoglobulin G (IgG) and IgG subclass reference intervals in children, using Optilite reagents, Clinical Chemistry & Laboratory Medicine 56(8) (2018) 1319-1327.

[3] M.S. Jayasuriya, K.W. Choy, L.K. Chin, J. Doery, A. Stewart, P. Bergman, Z.X. Lu, Reference intervals for neonatal thyroid function tests in the first 7 days of life, Journal of Pediatric Endocrinology & Metabolism 31(10) (2018) 1113-1116.

[4] H. Jopling, A. Yates, N. Burgoyne, K. Hayden, C. Chaloner, L. Tetlow, Paediatric Anti-Mullerian Hormone measurement: Male and female reference intervals established using the automated Beckman Coulter Access AMH assay, Endocrinology, Diabetes and Metabolism 1 (4) (no pagination)(e00021) (2018).

[5] S. Karn, M. Bhattarai, R. Rauniyar, A. Adhikari, P. Karna, B.P. Upadhyay, B.P. Gupta, Determination of CD4+ T- Lymphocytes in Healthy Children of Kathmandu, Journal of Nepal Health Research Council 16(3) (2018) 325-329.

[6] X. Li, D. Wang, C. Yang, Q. Zhou, S.L. Zhuoga, L.Q. Wang, H.X. Yao, Q. Zhang, Q. Ai, C.X. Yang, J.C. Xu, Establishment of age- and gender-specific pediatric reference intervals for liver function tests in healthy Han children, World Journal of Pediatrics 14(2) (2018) 151-159.

[7] A. Lofving, M. Domellof, L. Hellstrom-Westas, O. Andersson, Reference intervals for reticulocyte hemoglobin content in healthy infants, Pediatric Research 23 (2018) 23.
[8] H.L. Oh, J.A. Lee, D.H. Kim, J.S. Lim, Reference values for serum ferritin and percentage of transferrin saturation in Korean children and adolescents, Blood Research 53(1) (2018) 18-24.
[9] X. Peng, Y. Lv, G. Feng, Y. Peng, Q. Li, W. Song, X. Ni, Algorithm on age partitioning for estimation of reference intervals using clinical laboratory database exemplified with plasma creatinine, Clinical Chemistry & Laboratory Medicine 56(9) (2018) 1514-1523.
[10] P. Ridefelt, L. Hilsted, A. Juul, D. Hellberg, P. Rustad, Pediatric reference intervals for general clinical chemistry components - merging of studies from Denmark and Sweden, Scandinavian Journal of Clinical & Laboratory Investigation 78(5) (2018) 365-372.
[11] M.F. Strand, P.M. Fredriksen, O.P. Hjelle, M. Lindberg, Reference intervals for serum lipids and prevalence of dyslipidaemia in 6-12-year-old children: The Health Oriented Pedagogical Project (HOPP), Scandinavian Journal of Public Health 46(21_suppl) (2018) 21-27.
[12] J. Uaariyapanichkul, S. Chomtho, K. Suphapeetiporn, V. Shotelersuk, S. Punnahitananda, P. Chinjarernpan, O. Suteerojntrakool, Age-Related Reference Intervals for Blood Amino Acids in Thai Pediatric Population Measured by Liquid Chromatography Tandem Mass Spectrometry, Journal of Nutrition and Metabolism 2018 (2018) 5124035.
[13] L.L. Wang, L.Y. Zhang, H.Y. Wang, H.K. Wen, H.Q. Tao, X.W. Zhao, Chinese Pediatric Reference Intervals for Serum Cortisol on IMMULITE 2000, Annals of Laboratory Medicine 38(1) (2018) 59-62.
[14] C. Weidhofer, E. Meyer, R. Ristl, H. Wiedemann, J. Cadamuro, U. Kipman, J. Zierk, C. Male, P. Quehenberger, E. Haschke-Becher, E. Einwallner, Dynamic reference intervals for

coagulation parameters from infancy to adolescence, Clinica Chimica Acta 482 (2018) 124-135.

[15] M. Welter, K.C. Boritza, M.I. Anghebem-Oliveira, R. Henneberg, A.B. Hauser, F.G.M. Rego, G. Picheth, Reference intervals for serum 1,5-anhydroglucitol in children, adolescents, adults, and pregnant women, Clinica Chimica Acta 486 (2018) 54-58.

[16] N. Ziegelasch, M. Vogel, E. Muller, N. Tremel, A. Jurkutat, M. Loffler, N. Terliesner, J. Thiery, A. Willenberg, W. Kiess, K. Dittrich, Cystatin C serum levels in healthy children are related to age, gender, and pubertal stage, Pediatric Nephrology 20 (2018) 20.

[17] N. Ata, B. Maguire, D.C. Hamilton, S. Kuhle, Percentile curves for cardiometabolic disease markers in Canadian children and youth: a cross-sectional study, BMC Pediatrics 18(1) (2018) 314.

[18] V.C.P. de Melo, P.R.S. Ferreira, L.O. Ricardi, M.C. Batista, C.N. Franca, C. Ferreira, Definition of reference ranges for beta-isomerized carboxy-terminal telopeptide collagen type I for children and adolescents, Journal of Pediatric Endocrinology & Metabolism 31(6) (2018) 637-640.

[19] V.L. Fulgoni, 3rd, S. Agarwal, M.D. Kellogg, H.R. Lieberman, Establishing Pediatric and Adult RBC Reference Intervals With NHANES Data Using Piecewise Regression, American Journal of Clinical Pathology 04 (2018) 04.

[20] H. Aral, O.F. Ozer, H. Onur, A.M. Cilingirturk, I.T. Ozgen, M. Ayyildiz, Pediatric reference intervals of free thyroxine and thyroid stimulating hormone in three different hospitals, Turkish Journal of Biochemistry 43(5) (2018) 530-539.

[21] K. Adeli, V. Higgins, K. Trajcevski, N. White-Al Habeeb, The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper, Critical Reviews in Clinical Laboratory Sciences 54(6) (2017) 358-413.

[22] M.H. Gannage-Yared, N. Balech, V. Farah, M. Antar, R. Saliba, E. Chahine, Pediatric TSH Reference Intervals and Prevalence of High Thyroid Antibodies in the Lebanese Population, International Journal of Endocrinology Print 2017 (2017) 6372964.

[23] J. Gitaka, C. Ogwang, M. Ngari, P. Akoo, A. Olotu, C. Kerubo, G. Fegan, P. Njuguna, G. Nyakaya, T. Otieno, G. Mwambingu, K. Awuondo, B. Lowe, R. Chilengi, J.A. Berkley, Clinical laboratory reference values amongst children aged 4 weeks to 17 months in Kilifi, Kenya: A cross sectional observational study, PLoS ONE [Electronic Resource] 12(5) (2017) e0177382.

[24] A.E. Kulle, T. Reinehr, G. Simic-Schleicher, N.C. Hornig, P.M. Holterhus, Determination of 17OHPreg and DHEAS by LC-MS/MS: Impact of age, sex, pubertal stage, and BMI on the D5 steroid pathway, Journal of Clinical Endocrinology and Metabolism 102(1) (2017) 232-241.

[25] L. Lorenz, A. Peter, J. Arand, F. Springer, C.F. Poets, A.R. Franz, Reference Ranges of Reticulocyte Haemoglobin Content in Preterm and Term Infants: A Retrospective Analysis, Neonatology 111(3) (2017) 189-194.

[26] B.C. MacQueen, R.D. Christensen, E. Henry, A.M. Romrell, T.J. Pysher, S.T. Bennett, M.C. Sola-Visner, The immature platelet fraction: creating neonatal reference intervals and using these to categorize neonatal thrombocytopenias, Journal of Perinatology 37(7) (2017) 834-838.

[27] G. Narula, S. Khodaiji, A. Bableshwar, M.S. Bindra, Age-related reference intervals for immunoglobulin levels and lymphocyte subsets in Indian children, Indian Journal of Pathology & Microbiology 60(3) (2017) 360-364.

[28] P.C. Parkin, J. Hamid, C.M. Borkhoff, K. Abdullah, E.G. Atenafu, C.S. Birken, J.L. Maguire, A. Azad, V. Higgins, K. Adeli, Laboratory reference intervals in the assessment of iron status in young children, BMJ Paediatrics Open 1(1) (2017) e000074.

[29] G. Saravanabhavan, K. Werry, M. Walker, D. Haines, M. Malowany, C. Khoury, Human biomonitoring reference values for metals and trace elements in blood and urine derived from the Canadian Health Measures Survey 2007-2013, International Journal of Hygiene & Environmental Health 220(2 Pt A) (2017) 189-200.

[30] R. Sokou, L. Foudoulaki-Paparizos, T. Lytras, A. Konstantinidi, M. Theodoraki, I. Lambadaridis, A. Gounaris, S. Valsami, M. Politou, A. Gialeraki, G.K. Nikolopoulos, N. Iacovidou, S. Bonovas, A.E. Tsantes, Reference ranges of thromboelastometry in healthy full-term and pre-term neonates, Clinical Chemistry & Laboratory Medicine 55(10) (2017) 1592-1597.

[31] M. Supriya, T. De, R. Christopher, Age and gender-specific reference intervals for lysosomal enzymes in dried blood spot samples: A study in Indian population, Clinical Biochemistry 50(15) (2017) 858-863.

[32] G. Wanjian, H. Jie, G. Liang, W. Cheng, X. Tian, S. Jianjiang, Z. Chunni, Establishment of Reference Interval for Alkaline Phosphatase in Healthy Children of Various Ethnicities, Aged 0-12 Years, Laboratory Medicine 48(2) (2017) 166-171.

[33] R. Zhai, M. Zhang, J. Liu, H. Guang, B. Li, D. Chen, S. Zhang, Reference Intervals of and Relationships among Essential Trace Elements in Whole Blood of Children Aged 0-14 years, Journal of Clinical Laboratory Analysis 31(2) (2017).

[34] J. Zierk, F. Arzideh, R. Haeckel, H. Cario, M.C. Fruhwald, H.J. Gros, T. Gscheidmeier, R. Hoffmann, A. Krebs, R. Lichtinghagen, M. Neumann, H.G. Ruf, U. Steigerwald, T. Streichert, W. Rascher, M. Metzler, M. Rauh, Pediatric reference intervals for alkaline phosphatase, Clinical Chemistry & Laboratory Medicine 55(1) (2017) 102-110.

[35] F.D. Arslan, M. Serdar, E. Merve Ari, M. Onur Oztan, S. Hikmet Kozcu, H. Tarhan, O. Cakmak, M. Zeytinli, H. Yasar Ellidag, Determination of Age-Dependent Reference Ranges

for Coagulation Tests Performed Using Destiny Plus, Iranian Journal of Pediatrics 26(3) (2016) e6177.

[36] L. Berron-Ruiz, G. Lopez-Herrera, C.E. Avalos-Martinez, C. Valenzuela-Ponce, E. Ramirez-SanJuan, G. Santoyo-Sanchez, F. Mujica Guzman, F.J. Espinosa-Rosales, L. Santos-Argumedo, Variations of B cell subpopulations in peripheral blood of healthy Mexican population according to age: Relevance for diagnosis of primary immunodeficiencies, Allergologia et Immunopathologia 44(6) (2016) 571-579.

[37] M.G. Choi, M.J. Park, S.H. Kim, Reference values of lead in blood and related factors among Korean adolescents: the Korean National Health and Nutrition Examination Survey 2010-2013, Korean Journal of Pediatrics 59(3) (2016) 114-9.

[38] C. Cuerq, L. Restier, J. Drai, E. Blond, A. Roux, S. Charriere, M.C. Michalski, M. Di Filippo, E. Levy, A. Lachaux, N. Peretti, Establishment of reference values of alpha-tocopherol in plasma, red blood cells and adipose tissue in healthy children to improve the management of chylomicron retention disease, a rare genetic hypocholesterolemia, Orphanet Journal Of Rare Diseases 11(1) (2016) 114.

[39] G. Demirdjian, S. Bord, C. Lejeune, R. Masica, D. Riviere, L. Nicouleau, P. Denizot, P.Y. Marquet, Performance characteristics of the Access AMH assay for the quantitative determination of anti-Mullerian hormone (AMH) levels on the Access family of automated immunoassay systems, Clinical Biochemistry 49(16-17) (2016) 1267-1273.

[40] N. Fukuzumi, K. Osawa, I. Sato, S. Iwatani, R. Ishino, N. Hayashi, K. Iijima, J. Saegusa, I. Morioka, Age-specific percentile-based reference curve of serum procalcitonin concentrations in Japanese preterm infants, Scientific Reports 6 (2016) 23871.

[41] I. Gennai, N. Di Iorgi, G. Reggiardo, C. Gatti, E. Bertelli, A.E.M. Allegri, S. Barco, M. Maghnie, G. Tripodi, G. Cangemi, Age- and sex-matched reference curves for serum collagen type I C-telopeptides and bone alkaline phosphatase in children and adolescents: An

alternative multivariate statistical analysis approach, Clinical Biochemistry 49(10-11) (2016) 802-807.

[42] J.Y. Kim, Y.R. Shin, H.K. Kil, M.R. Park, J.W. Lee, Reference Intervals of Thromboelastometric Evaluation of Coagulation in Pediatric Patients with Congenital Heart Diseases: A Retrospective Investigation, Medical Science Monitor 22 (2016) 3576-3587.

[43] C.S. Kira, A.M. Sakuma, E.M. De Capitani, C.U. de Freitas, M.R. Cardoso, N. Gouveia, Associated factors for higher lead and cadmium blood levels, and reference values derived from general population of Sao Paulo, Brazil, Science of the Total Environment 543(Pt A) (2016) 628-35.

[44] S.L. La'ulu, K.J. Rasmussen, J.A. Straseski, Pediatric Reference Intervals for Free Thyroxine and Free Triiodothyronine by Equilibrium Dialysis-Liquid Chromatography-Tandem Mass Spectrometry, Journal of clinical research in pediatric endocrinology 8(1) (2016) 26-31.

[45] U. Lausten-Thomsen, M. Christiansen, P. Louise Hedley, C. Esmann Fonvig, T. Stjernholm, O. Pedersen, T. Hansen, J.C. Holm, Reference values for serum leptin in healthy non-obese children and adolescents, Scandinavian Journal of Clinical & Laboratory Investigation 76(7) (2016) 561-567.

[46] E. Lopez-Ruzafa, M.A. Vazquez-Lopez, F. Lendinez-Molinos, J. Poveda-Gonzalez, R. Galera-Martinez, A. Bonillo-Perales, M. Martin-Gonzalez, Reference Values of Reticulocyte Hemoglobin Content and Their Relation With Other Indicators of Iron Status in Healthy Children, Journal of Pediatric Hematology/Oncology 38(7) (2016) e207-12.

[47] B.C. MacQueen, R.D. Christensen, B.A. Yoder, E. Henry, V.L. Baer, S.T. Bennett, H.M. Yaish, Comparing automated vs manual leukocyte differential counts for quantifying the 'left shift' in the blood of neonates, Journal of Perinatology 36(10) (2016) 843-8.

[48] J. Okebe, J. Mwesigwa, S.C. Agbla, F. Sanya-Isijola, I. Abubakar, U. D'Alessandro, A. Jaye, K. Bojang, Seasonal variation in haematological and biochemical reference values for healthy young children in The Gambia, BMC Pediatrics 16 (2016) 5.

[49] B. Shashaj, R. Luciano, B. Contoli, G.S. Morino, M.R. Spreghini, C. Rustico, R.W. Sforza, B. Dallapiccola, M. Manco, Reference ranges of HOMA-IR in normal-weight and obese young Caucasians, Acta Diabetologica 53(2) (2016) 251-60.

[50] P. Toulon, M. Berruyer, M. Brionne-Francois, F. Grand, D. Lasne, C. Telion, J. Arcizet, R. Giacomello, N. De Pooter, Age dependency for coagulation parameters in paediatric populations. Results of a multicentre study aimed at defining the age-specific reference ranges, Thrombosis & Haemostasis 116(1) (2016) 9-16.

[51] (!!! INVALID CITATION !!! [51]).

[52] M.G. Castagna, L. Fugazzola, F. Maino, D. Covelli, S. Memmo, F. Sestini, C. Fioravanti, C. Ferraris Fusarini, C. Scapellato, F. Macchini, G. Cevenini, F. Pacini, Reference range of serum calcitonin in pediatric population, Journal of Clinical Endocrinology & Metabolism 100(5) (2015) 1780-4.

[53] L.C. Franscini, M. Vazquez-Montes, T. Buclin, R. Perera, M. Dunand, E. Grouzmann, M. Beck-Popovic, Pediatric reference intervals for plasma free and total metanephrines established with a parametric approach: relevance to the diagnosis of neuroblastoma, Pediatric Blood & Cancer 62(4) (2015) 587-93.

[54] A. Ghasemi, I. Azimzadeh, M. Afghan, A.A. Momenan, F. Bagheripour, F. Azizi, Pediatric reference values for serum creatinine and estimated glomerular filtration rate in Iranians: Tehran Lipid and Glucose Study, Archives of Iranian Medicine 18(11) (2015) 753-9.

[55] L. Jia, J. Li, Y. Zhang, Y. Shi, E. Yuan, J. Liu, P. Wang, S. Rong, J. Xing, Y. Tian, J. Li, Age- and sex-related reference intervals of lymphocyte subsets in healthy ethnic Han Chinese

children, Cytometry Part A: The Journal of the International Society for Analytical Cytology 87(12) (2015) 1116-26.

[56] U. Lausten-Thomsen, M. Christiansen, C.E. Fonvig, C. Trier, O. Pedersen, T. Hansen, J.C. Holm, Reference values for serum total adiponectin in healthy non-obese children and adolescents, Clinica Chimica Acta 450 (2015) 11-4.

[57] F. Perez-de-Heredia, S. Gomez-Martinez, L.E. Diaz, A.M. Veses, E. Nova, J. Warnberg, I. Huybrechts, K. Vyncke, O. Androutsos, M. Ferrari, G. Palacios, A. Wastlund, E. Kovacs, F. Gottrand, M. Gonzalez-Gross, M.J. Castillo, M. Sjostrom, Y. Manios, A. Kafatos, D. Molnar, K. Widhalm, L.A. Moreno, A. Marcos, H.S. Group, Influence of sex, age, pubertal maturation and body mass index on circulating white blood cell counts in healthy European adolescents-the HELENA study, European Journal of Pediatrics 174(8) (2015) 999-1014.

[58] L.J. Schlapbach, E. Giannoni, S. Wellmann, M. Stocker, R.A. Ammann, R. Graf, Normal values for pancreatic stone protein in different age groups, BMC Anesthesiology 15 (2015) 168.

[59] N. Slhessarenko, C.M. Jacob, R.S. Azevedo, C.J. Fontes, G.V. Novak, A. Andriolo, Serum lipids in Brazilian children and adolescents: determining their reference intervals, BMC Public Health 15 (2015) 18.

[60] K. Soeby, P.B. Jensen, T. Werge, S. Sorensen, Mining of hospital laboratory information systems: a model study defining age- and gender-specific reference intervals and trajectories for plasma creatinine in a pediatric population, Clinical Chemistry & Laboratory Medicine 53(10) (2015) 1621-30.

[61] C. Teixeira, J. Barbot, M.I. Freitas, Reference values for reticulocyte parameters and hypochromic RBC in healthy children, International Journal of Laboratory Hematology 37(5) (2015) 626-30.

[62] G.O. Temel, E. Arzu Kanik, D.D. Yildirim, H. Didem Ovla, H. Kalafat, Determination of sub-groups in reference interval calculations and an application, Turkish Journal of Biochemistry 40(3) (2015) 197-204.
[63] F. Tosato, G. Bucciol, G. Pantano, M.C. Putti, M.C. Sanzari, G. Basso, M. Plebani, Lymphocytes subsets reference values in childhood, Cytometry Part A: The Journal of the International Society for Analytical Cytology 87(1) (2015) 81-5.
[64] O. Uemura, T. Nagai, K. Ishikura, S. Ito, H. Hataya, Y. Gotoh, N. Fujita, Y. Akioka, T. Kaneko, M. Honda, Reference glomerular filtration rate levels in Japanese children: using the creatinine and cystatin C based estimated glomerular filtration rate, Clinical and Experimental Nephrology 19(4) (2015) 683-687.
[65] K. Yamamoto, N. Kitagawa, T. Takamasu, Standard values of rapid turnover proteins and zinc in Japanese children, Asia Pacific Journal of Clinical Nutrition 24(3) (2015) 504-8.
[66] J. Zierk, F. Arzideh, T. Rechenauer, R. Haeckel, W. Rascher, M. Metzler, M. Rauh, Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence, Clinical Chemistry 61(7) (2015) 964-73.
[67] A.J. Zwiers, S.N. de Wildt, Y.B. de Rijke, S.P. Willemsen, N.S. Abdullahi, D. Tibboel, K. Cransberg, Reference intervals for renal injury biomarkers neutrophil gelatinase-associated lipocalin and kidney injury molecule-1 in young infants, Clin Chem Lab Med 53(8) (2015) 1279-89.
[68] S.M. Cho, S.G. Lee, H.S. Kim, J.H. Kim, Establishing pediatric reference intervals for 13 biochemical analytes derived from normal subjects in a pediatric endocrinology clinic in Korea, Clinical Biochemistry 47(18) (2014) 268-71.
[69] M. Bidlingmaier, N. Friedrich, R.T. Emeny, J. Spranger, O.D. Wolthers, J. Roswall, A. Korner, B. Obermayer-Pietsch, C. Hubener, J. Dahlgren, J. Frystyk, A.F. Pfeiffer, A. Doering, M. Bielohuby, H. Wallaschofski, A.M. Arafat, Reference intervals for insulin-like

growth factor-1 (igf-i) from birth to senescence: results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations, Journal of Clinical Endocrinology & Metabolism 99(5) (2014) 1712-21.
[70] R.D. Christensen, V.L. Baer, D.K. Lambert, E. Henry, S.J. Ilstrup, S.T. Bennett, Reference intervals for common coagulation tests of preterm infants (CME), Transfusion 54(3) (2014) 627-32:quiz 626.
[71] M. Duchamp, D. Sterlin, A. Diabate, B. Uring-Lambert, V. Guerin-El Khourouj, B. Le Mauff, D. Monnier, C. Malcus, M. Labalette, C. Picard, B-cell subpopulations in children: National reference values, Immunity, Inflammation and Disease 2(3) (2014) 131-40.
[72] E. Erhardt, R. Foraita, I. Pigeot, G. Barba, T. Veidebaum, M. Tornaritis, N. Michels, G. Eiben, W. Ahrens, L.A. Moreno, E. Kovacs, D. Molnar, I. consortium, Reference values for leptin and adiponectin in children below the age of 10 based on the IDEFICS cohort, International Journal of Obesity 38 Suppl 2 (2014) S32-8.
[73] N. Friedrich, O.D. Wolthers, A.M. Arafat, R.T. Emeny, J. Spranger, J. Roswall, J. Kratzsch, H.J. Grabe, C. Hubener, A.F.H. Pfeiffer, A. Doring, M. Bielohuby, J. Dahlgren, J. Frystyk, H. Wallaschofski, M. Bidlingmaier, Age- and sex-specific reference intervals across life span for insulin-like growth factor binding protein 3 (igfbp-3) and the igf-i to igfbp-3 ratio measured by new automated chemiluminescence assays, Journal of Clinical Endocrinology and Metabolism 99(5) (2014) 1675-1686.
[74] R.F. Greaves, M.R. Zacharin, S.M. Donath, T.E. Inder, L.W. Doyle, R.W. Hunt, Establishment of hormone reference intervals for infants born < 30 weeks' gestation, Clinical Biochemistry 47(15) (2014) 101-8.
[75] W.H. Hahn, C.W. Bae, Reference intervals of serum cystatin C/creatinine ratio of 30 postnatal days in neonates, Pediatric Nephrology 29(2) (2014) 311-4.

[76] I.M. Hirfanoglu, S. Unal, E.E. Onal, S. Beken, C. Turkyilmaz, H. Pasaoglu, E. Koc, E. Ergenekon, Y. Atalay, Analysis of serum gamma-glutamyl transferase levels in neonatal intensive care unit patients, Journal of Pediatric Gastroenterology & Nutrition 58(1) (2014) 99-101.

[77] S. Hu, J. Wu, D. Lu, Y. Dai, M. Mai, R. Wei, Q. Liu, W. Jiang, Y. Zhong, Reference values of fetal serum beta2-microglobulin in the Chinese: evaluation of its clinical usefulness, Clinical Chemistry & Laboratory Medicine 52(3) (2014) 413-7.

[78] T. Koester-Weber, J. Valtuena, C. Breidenassel, L. Beghin, M. Plada, S. Moreno, I. Huybrechts, G. Palacios, S. Gomez-Martinez, U. Albers, S. De Henauw, G. Maiani, A. Kafatos, D. Molnar, M. Sjostrom, K. Widhalm, Y. Manios, L.A. Moreno, A. Marcos, M.J. Castillo, P. Stehle, M. Gonzalez-Gross, Reference values for leptin, cortisol, insulin and glucose, among European adolescents and their association with adiposity: the HELENA study, Nutricion Hospitalaria 30(5) (2014) 1181-90.

[79] T.P. Loh, G. Antoniou, P. Baghurst, M.P. Metz, Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals, Pathology 46(4) (2014) 336-43.

[80] E. Macit, M. Kizilgun, E. Cakir, A. Karaoglu, E.O. Akgul, M. Oztosun, I. Aydin, F.N. Aydin, M. Agilli, T. Turker, R. Ogur, Y. Gulcan Kurt, H. Gul, T. Cayci, E. Ozkan, Pediatric reference intervals for plasma and urine essential amino acids in a Turkish population, Turkish Journal of Medical Sciences 44(2) (2014) 323-9.

[81] M.I. Moraes-Pinto, E. Ono, E.C. Santos-Valente, L.C. Almeida, P.R. Andrade, M.I. Dinelli, A.M. Santos, R. Salomao, Lymphocyte subsets in human immunodeficiency virus-unexposed Brazilian individuals from birth to adulthood, Memorias do Instituto Oswaldo Cruz 109(8) (2014) 989-98.

[82] A.A. Odutola, M.O. Afolabi, J. Jafali, I. Baldeh, O.A. Owolabi, P. Owiafe, G. Bah, B. Jaiteh, N.I. Mohammed, S.A. Donkor, J. Ndure, J. Adetifa, K.L. Flanagan, M.O.C. Ota, Haematological and biochemical reference values of Gambian infants, Tropical Medicine & International Health 19(3) (2014) 275-283.

[83] K.W. Oh, M.S. Koo, H.W. Park, M.L. Chung, M.H. Kim, G. Lim, Establishing a reference range for triiodothyronine levels in preterm infants, Early Human Development 90(10) (2014) 621-4.

[84] F. Savino, L. Rossi, S. Benetti, E. Petrucci, M. Sorrenti, L. Silvestro, Serum reference values for leptin in healthy infants.[Erratum appears in PLoS One. 2015;10(1):e0117622; PMID: 25574666], [Erratum appears in PLoS One. 2015;10(2):e0118046; PMID: 25658819], PLoS ONE [Electronic Resource] 9(11) (2014) e113024.

[85] H. Schlenz, T. Intemann, M. Wolters, E.M. Gonzalez-Gil, A. Nappo, A. Fraterman, T. Veidebaum, D. Molnar, M. Tornaritis, I. Sioen, S. Marild, L. Iacoviello, W. Ahrens, I. consortium, C-reactive protein reference percentiles among pre-adolescent children in Europe based on the IDEFICS study population, International Journal of Obesity 38 Suppl 2 (2014) S26-31.

[86] J.L. Shaw, A. Cohen, D. Konforte, T. Binesh-Marvasti, D.A. Colantonio, K. Adeli, Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children, Clinical Biochemistry 47(3) (2014) 166-72.

[87] S. Sheikhbahaei, B. Mahdaviani, A. Abdollahi, F. Nayeri, Serum thyroid stimulating hormone, total and free T4 during the neonatal period: Establishing regional reference intervals, Indian Journal of Endocrinology and Metabolism 18(1) (2014) 39-43.

[88] X. Sun, B. Lemyre, X. Nan, J. Harrold, S.L. Perkins, S.E. Lawrence, N. Barrowman, Free thyroxine and thyroid-stimulating hormone reference intervals in very low birth weight

infants at 3-6 weeks of life with the Beckman Coulter Unicel DxI 800, Clinical Biochemistry 47(1-2) (2014) 16-8.

[89] S.C. Voss, E. Varamenti, M. Elzain Elgingo, P.C. Bourdon, New parameters and reference values for monitoring iron status in Middle Eastern adolescent male athletes, Journal of Sports Medicine & Physical Fitness 54(2) (2014) 179-85.

[90] M. Wolters, H. Schlenz, R. Foraita, C. Galli, P. Rise, L.A. Moreno, D. Molnar, P. Russo, T. Veidebaum, M. Tornaritis, K. Vyncke, G. Eiben, L. Iacoviello, W. Ahrens, I. consortium, Reference values of whole-blood fatty acids by age and sex from European children aged 3-8 years, International Journal of Obesity 38 Suppl 2 (2014) S86-98.

[91] K.H. Yi, J.S. Hwang, E.Y. Kim, S.H. Lee, D.H. Kim, J.S. Lim, Prevalence of insulin resistance and cardiometabolic risk in Korean children and adolescents: a population-based study, Diabetes Research & Clinical Practice 103(1) (2014) 106-13.

[92] M.E. Mortensen, S.P. Caudill, K.L. Caldwell, C.D. Ward, R.L. Jones, Total and methyl mercury in whole blood measured for the first time in the U.S. population: NHANES 2011-2012, Environmental Research 134 (2014) 257-64.

[93] M. Aldrimer, P. Ridefelt, P. Rodoo, F. Niklasson, J. Gustafsson, D. Hellberg, Population-based pediatric reference intervals for hematology, iron and transferrin, Scandinavian Journal of Clinical & Laboratory Investigation 73(3) (2013) 253-61.

[94] H. Aitkenhead, S.J. Heales, Establishment of paediatric age-related reference intervals for serum prolactin to aid in the diagnosis of neurometabolic conditions affecting dopamine metabolism, Annals of Clinical Biochemistry 50(Pt 2) (2013) 156-8.

[95] G. Biino, I. Santimone, C. Minelli, R. Sorice, B. Frongia, M. Traglia, S. Ulivi, A. Di Castelnuovo, M. Gogele, T. Nutile, M. Francavilla, C. Sala, N. Pirastu, C. Cerletti, L. Iacoviello, P. Gasparini, D. Toniolo, M. Ciullo, P. Pramstaller, M. Pirastu, G. de Gaetano, C.L. Balduini, Age- and sex-related variations in platelet count in Italy: a proposal of

reference ranges based on 40987 subjects' data, PLoS ONE [Electronic Resource] 8(1) (2013) e54289.

[96] M. Can, E. Piskin, B. Guven, S. Acikgoz, G. Mungan, Evaluation of serum lipid levels in children, Pediatric Cardiology 34(3) (2013) 566-9.

[97] B. Guven, M. Can, G. Mungan, S. Acikgoz, Reference values for serum levels of insulin-like growth factor 1 (IGF-1) and IGF-binding protein 3 (IGFBP-3) in the West Black Sea region of Turkey, Scandinavian Journal of Clinical & Laboratory Investigation 73(2) (2013) 135-40.

[98] A. Heinze, M.C. Elze, S. Kloess, O. Ciocarlie, C. Konigs, S. Betz, M. Bremm, R. Esser, T. Klingebiel, M. Serban, J.L. Hutton, U. Koehl, Age-matched dendritic cell subpopulations reference values in childhood, Scandinavian Journal of Immunology 77(3) (2013) 213-20.

[99] L. Hilsted, P. Rustad, L. Aksglaede, K. Sorensen, A. Juul, Recommended Nordic paediatric reference intervals for 21 common biochemical properties.[Erratum appears in Scand J Clin Lab Invest. 2013 Dec;73(8):661], Scandinavian Journal of Clinical & Laboratory Investigation 73(1) (2013) 1-9.

[100] M. Joncquel-Chevalier Curt, D. Cheillan, G. Briand, G.S. Salomons, K. Mention-Mulliez, D. Dobbelaere, J.M. Cuisset, L. Lion-Francois, V. Des Portes, A. Chabli, V. Valayannopoulos, J.F. Benoist, J.M. Pinard, G. Simard, O. Douay, K. Deiva, M. Tardieu, A. Afenjar, D. Heron, F. Rivier, B. Chabrol, F. Prieur, F. Cartault, G. Pitelet, A. Goldenberg, S. Bekri, M. Gerard, R. Delorme, N. Porchet, C. Vianey-Saban, J. Vamecq, Creatine and guanidinoacetate reference values in a French population, Molecular Genetics & Metabolism 110(3) (2013) 263-7.

[101] A.E. Kulle, M. Welzel, P.M. Holterhus, F.G. Riepe, Implementation of a liquid chromatography tandem mass spectrometry assay for eight adrenal C-21 steroids and pediatric reference data, Hormone research in paediatrics 79(1) (2013) 22-31.

[102] J.A. Mendez, A.L. Fernandez-Sanmamed, M.S. Gomez-Holgado, F. Fernandez-Rodriguez, Age-related reference values for plasma amino acids in a Spanish population measured by gas chromatography-mass spectrometry, Journal of Pediatric Endocrinology & Metabolism 26(3-4) (2013) 333-41.

[103] A. Morovat, A. Catchpole, A. Meurisse, A. Carlisi, A.C. Bekaert, O. Rousselle, M. Paddon, T. James, E. Cavalier, IDS iSYS automated intact procollagen-1-N-terminus pro-peptide assay: method evaluation and reference intervals in adults and children, Clinical Chemistry & Laboratory Medicine 51(10) (2013) 2009-18.

[104] V. Osta, M.S. Caldirola, M. Fernandez, M.I. Marcone, G. Tissera, S. Pennesi, C. Ayuso, Utility of new mature erythrocyte and reticulocyte indices in screening for iron-deficiency anemia in a pediatric population, International Journal of Laboratory Hematology 35(4) (2013) 400-5.

[105] A.F. Radicioni, N. Tahani, M. Spaziani, A. Anzuini, C. Piccheri, A. Semeraro, L. Tarani, A. Lenzi, Reference ranges for thyroid hormones in normal Italian children and adolescents and overweight adolescents, Journal of Endocrinological Investigation 36(5) (2013) 326-30.

[106] K. Tauseef Bukhari, H. Zafar, Reference range variation in haematological indices amongst five different age groups of less than one year in Islamabad, Pakistan, Pakistan Journal of Medical Sciences 29(2) (2013) 577-80.

[107] L. Zhu, X. Zhang, X. He, X. Yang, Y. Wang, C. Wang, Z. Feng, Reference intervals for serum thyroid hormones in preterm hospitalized infants, Journal of Pediatric Endocrinology & Metabolism 26(5-6) (2013) 463-7.

[108] G. Eisenhofer, P. Lattke, M. Herberg, G. Siegert, N. Qin, R. Darr, J. Hoyer, A. Villringer, A. Prejbisz, A. Januszewicz, A. Remaley, V. Martucci, K. Pacak, H.A. Ross, F.C. Sweep, J.W. Lenders, Reference intervals for plasma free metanephrines with an age

adjustment for normetanephrine for optimized laboratory testing of phaeochromocytoma, Annals of Clinical Biochemistry 50(Pt 1) (2013) 62-9.

[109] I. Zec, I. Kucak, I. Begcevic, A.M. Simundic, D. Tislaric-Medenjak, Z.B. Megla, N. Vrkic, Reference intervals for reproductive hormones in prepubertal children on the automated Roche cobas e 411 analyzer, Clinical Biochemistry 45(15) (2012) 1206-12.

[110] S.B. Troy, A. Rowhani-Rahbar, L. Dyner, G. Musingwini, A.K. Shetty, G. Woelk, L. Stranix-Chibanda, K. Nathoo, Y.A. Maldonado, Hematologic and immunologic parameters in Zimbabwean infants: a case for using local reference intervals to monitor toxicities in clinical trials, Journal of Tropical Pediatrics 58(1) (2012) 59-62.

[111] I.F. Shatat, R.T. Abdallah, D.J. Sas, S.M. Hailpern, Serum uric acid in U.S. adolescents: distribution and relationship to demographic characteristics and cardiovascular risk factors, Pediatric Research 72(1) (2012) 95-100.

[112] P. Shahabi, G. Siest, B. Herbeth, N.C. Ndiaye, S. Visvikis-Siest, Clinical necessity of partitioning of human plasma haptoglobin reference intervals by recently-discovered rs2000999, Clinica Chimica Acta 413(19-20) (2012) 1618-24.

[113] P. Ridefelt, M. Aldrimer, P.O. Rodoo, F. Niklasson, L. Jansson, J. Gustafsson, D. Hellberg, Population-based pediatric reference intervals for general clinical chemistry analytes on the Abbott Architect ci8200 instrument, Clinical Chemistry & Laboratory Medicine 50(5) (2012) 845-51.

[114] R.C. Olney, J.W. Permuy, T.C. Prickett, J.C. Han, E.A. Espiner, Amino-terminal propeptide of C-type natriuretic peptide (NTproCNP) predicts height velocity in healthy children, Clinical Endocrinology 77(3) (2012) 416-22.

[115] J. Liu, E. Yuan, Z. Zhang, L. Jia, Z. Yin, X. Meng, H. Du, Age- and sex-specific reference intervals for blood copper, zinc, calcium, magnesium, iron, lead, and cadmium in infants and children, Clinical Biochemistry 45(6) (2012) 416-9.

[116] C.N. Lin, A. Wilson, B.B. Church, S. Ehman, W.L. Roberts, G.A. McMillin, Pediatric reference intervals for serum copper and zinc, Clinica Chimica Acta 413(5-6) (2012) 612-5.
[117] A.J. Lem, Y.B. de Rijke, H. van Toor, M.A. de Ridder, T.J. Visser, A.C. Hokken-Koelega, Serum thyroid hormone levels in healthy children from birth to adulthood and in short children born small for gestational age, Journal of Clinical Endocrinology & Metabolism 97(9) (2012) 3170-8.
[118] G. Kardar, M. Oraei, M. Shahsavani, Z. Namdar, G. Kazemisefat, M. Haghi Ashtiani, S. Shams, Z. Pourpak, M. Moin, Reference Intervals for Serum Immunoglobulins IgG, IgA, IgM and Complements C3 and C4 in Iranian Healthy Children, Iranian Journal of Public Health 41(7) (2012) 59-63.
[119] T. Isojima, A. Shimatsu, S. Yokoya, K. Chihara, T. Tanaka, N. Hizuka, A. Teramoto, K. Tatsumi, K. Tachibana, N. Katsumata, R. Horikawa, Standardized centile curves and reference intervals of serum insulin-like growth factor-I (IGF-I) levels in a normal Japanese population using the LMS method, Endocrine Journal 59(9) (2012) 771-780.
[120] S.E. Hyun, B.C. Lee, B.K. Suh, S.C. Chung, C.W. Ko, H.S. Kim, K.H. Lee, S.W. Yang, C.H. Shin, J.S. Hwang, D.H. Kim, B.K. Lim, J.D. Kim, H.W. Yoo, H.S. Kim, W.Y. Chung, M.J. Park, Y.J. Woo, C.J. Kim, D.Y. Lee, E.Y. Kim, J.H. Choi, H.S. Han, I.T. Hwang, H.S. Kim, Reference values for serum levels of insulin-like growth factor-I and insulin-like growth factor binding protein-3 in Korean children and adolescents, Clinical Biochemistry 45(1-2) (2012) 16-21.
[121] M. Gonzalez-Gross, J. Benser, C. Breidenassel, U. Albers, I. Huybrechts, J. Valtuena, A. Spinneker, M. Segoviano, K. Widhalm, D. Molnar, L.A. Moreno, P. Stehle, K. Pietrzik, Gender and age influence blood folate, vitamin B<inf>12</inf>, vitamin B<inf>6</inf>, and homocysteine levels in European adolescents: The Helena Study, Nutrition Research 32(11) (2012) 817-826.

[122] O. D'Apolito, D. Garofalo, G. la Marca, A. Dello Russo, G. Corso, Reference intervals for orotic acid in urine, plasma and dried blood spot using hydrophilic interaction liquid chromatography-tandem mass spectrometry, Journal of Chromatography B: Analytical Technologies in the Biomedical & Life Sciences 883-884 (2012) 155-60.

[123] E.A. Chaler, R. Fiorenzano, C. Chilelli, V. Llinares, G. Areny, V. Herzovich, M. Maceiras, J.M. Lazzati, M. Mendioroz, M.A. Rivarola, A. Belgorosky, Age-specific thyroid hormone and thyrotropin reference intervals for a pediatric and adolescent population, Clinical Chemistry & Laboratory Medicine 50(5) (2012) 885-90.

[124] G. Cangemi, N. Di Iorgi, S. Barco, G. Reggiardo, M. Maghnie, G. Melioli, Plasma total adiponectin levels in pediatrics: reference intervals calculated as a continuous variable of age, Clinical Biochemistry 45(18) (2012) 1703-5.

[125] G. Bedogni, G. Giannone, M. Maghnie, C. Giacomozzi, N. Di Iorgi, S. Pedicelli, E. Peschiaroli, G. Melioli, M. Muraca, M. Cappa, S. Cianfarani, Serum insulin-like growth factor-I (IGF-I) reference ranges for chemiluminescence assay in childhood and adolescence. Data from a population of in- and out-patients, Growth Hormone & Igf Research 22(3-4) (2012) 134-8.

[126] I.M. Appel, B. Grimminck, J. Geerts, R. Stigter, M.H. Cnossen, A. Beishuizen, Age dependency of coagulation parameters during childhood and puberty, Journal of Thrombosis & Haemostasis 10(11) (2012) 2254-63.

[127] B. Yuksel, M.N. Ozbek, N.O. Mungan, F. Darendeliler, B. Budan, A. Bideci, E. Cetinkaya, M. Berberoglu, O. Evliyaoglu, E. Yesilkaya, I. Arslanoglu, S. Darcan, R. Bundak, O. Ercan, Serum IGF-1 and IGFBP-3 levels in healthy children between 0 and 6 years of age, Journal of clinical research in pediatric endocrinology 3(2) (2011) 84-8.

[128] P. Yi, L. Liu, H. Mei, F. Zeng, Z. Huang, H. Niu, Establishment of reference range of plasma amino acids for younger Chinese children by reverse phase HPLC, Journal of Pediatric Endocrinology & Metabolism 24(9-10) (2011) 733-8.

[129] O. Uemura, M. Honda, T. Matsuyama, K. Ishikura, H. Hataya, N. Yata, T. Nagai, Y. Ikezumi, N. Fujita, S. Ito, K. Iijima, T. Kitagawa, Age, gender, and body length effects on reference serum creatinine levels determined by an enzymatic method in Japanese children: a multicenter study, Clinical & Experimental Nephrology 15(5) (2011) 694-699.

[130] S. Turan, B. Topcu, I. Gokce, T. Guran, Z. Atay, A. Omar, T. Akcay, A. Bereket, Serum alkaline phosphatase levels in healthy children and evaluation of alkaline phosphatase z-scores in different types of rickets, Journal of clinical research in pediatric endocrinology 3(1) (2011) 7-11.

[131] W. Tamimi, H. Tamim, N. Felimban, A. Almutair, Y. Altwaijri, I. Alalwan, Age- and gender-specific reference intervals for serum lipid levels (measured with an Advia 1650 analyzer) in school children, Pediatrics International 53(6) (2011) 814-9.

[132] M.H. Roehrl, D. Lantz, C. Sylvester, J.Y. Wang, Age-dependent reference ranges for automated assessment of immature granulocytes and clinical significance in an outpatient setting, Archives of Pathology & Laboratory Medicine 135(4) (2011) 471-7.

[133] R.K. Marwaha, R. Khadgawat, N. Tandon, R. Kanwar, A. Narang, A. Sastry, K. Bhadra, Reference intervals of serum lipid profile in healthy Indian school children and adolescents, Clinical Biochemistry 44(10-11) (2011) 760-6.

[134] Y. Huang, E. Eapen, S. Steele, V. Grey, Establishment of reference intervals for bone markers in children and adolescents, Clinical Biochemistry 44(10-11) (2011) 771-8.

[135] F.A. Verburg, C. Kirchgassner, H. Hebestreit, U. Steigerwald, E.G. Lentjes, K. Ergezinger, I. Grelle, C. Reiners, M. Luster, Reference ranges for analytes of thyroid function in children, Hormone & Metabolic Research 43(6) (2011) 422-6.

[136] S.M. Clifford, A.M. Bunker, J.R. Jacobsen, W.L. Roberts, Age and gender specific pediatric reference intervals for aldolase, amylase, ceruloplasmin, creatine kinase, pancreatic amylase, prealbumin, and uric acid, Clinica Chimica Acta 412(9-10) (2011) 788-90.

[137] D. Bouvier, C. Castellani, M. Fournier, J.B. Dauphin, S. Ughetto, M. Breton, A. Labbe, A.M. Weinberg, V. Sapin, Reference ranges for serum S100B protein during the first three years of life, Clinical Biochemistry 44(10-11) (2011) 927-9.

[138] C. Alberti, D. Chevenne, I. Mercat, E. Josserand, P. Armoogum-Boizeau, J. Tichet, J. Leger, Serum concentrations of insulin-like growth factor (IGF)-1 and IGF binding protein-3 (IGFBP-3), IGF-1/IGFBP-3 ratio, and markers of bone turnover: reference values for French children and adolescents and z-score comparability with other references, Clinical Chemistry 57(10) (2011) 1424-35.

[139] C. Langer, B. Tambyrayah, S. Thedieck, U. Nowak-Gottl, Testing for lipoprotein(a) concentration and apolipoprotein(a) phenotypes: method standardization and pediatric reference values, Seminars in Thrombosis & Hemostasis 37(7) (2011) 810-3.

[140] S. Xu, X. Gu, H. Pan, H. Zhu, F. Gong, Y. Li, Y. Xing, Reference ranges for serum IGF-1 and IGFBP-3 levels in Chinese children during childhood and adolescence, Endocrine Journal 57(3) (2010) 221-8.

[141] T.I. Takala, E. Makela, P. Suominen, J. Matomaki, H. Lapinleimu, L. Lehtonen, A. Rajamaki, K. Irjala, P.M. Lahteenmaki, Blood cell and iron status analytes of preterm and full-term infants from 20 weeks onwards during the first year of life, Clinical Chemistry & Laboratory Medicine 48(9) (2010) 1295-301.

[142] E.K. Southcott, J.L. Kerrigan, J.M. Potter, R.D. Telford, P. Waring, G.J. Reynolds, A.R. Lafferty, P.E. Hickman, Establishment of pediatric reference intervals on a large cohort of healthy children, Clinica Chimica Acta 411(19-20) (2010) 1421-7.

[143] W.A. Salameh, M.M. Redor-Goldman, N.J. Clarke, R.E. Reitz, M.P. Caulfield, Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges, Steroids 75(2) (2010) 169-75.

[144] B. Piatosa, B. Wolska-Kusnierz, K. Siewiera, H. Grzduk, E. Galkowska, E. Bernatowska, Distribution of leukocyte and lymphocyte subsets in peripheral blood. Age related normal values for preliminary evaluation of the immune status in Polish children, Central-European Journal of Immunology 35(3) (2010) 168-175.

[145] E. Oswald, B. Stalzer, E. Heitz, M. Weiss, M. Schmugge, A. Strasak, P. Innerhofer, T. Haas, Thromboelastometry (ROTEM) in children: age-related reference ranges and correlations with standard coagulation tests, British Journal of Anaesthesia 105(6) (2010) 827-35.

[146] H. Morbach, E.M. Eichhorn, J.G. Liese, H.J. Girschick, Reference values for B cell subpopulations from infancy to adulthood, Clinical & Experimental Immunology 162(2) (2010) 271-9.

[147] R.K. Marwaha, R. Khadgawat, N. Tandon, R. Kanwar, A. Narang, A. Sastry, K. Bhadra, M. Kalaivani, Reference intervals of serum calcium, ionized calcium, phosphate and alkaline phosphatase in healthy Indian school children and adolescents, Clinical Biochemistry 43(15) (2010) 1216-9.

[148] M.M. Kushnir, T. Blamires, A.L. Rockwood, W.L. Roberts, B. Yue, E. Erdogan, A.M. Bunker, A.W. Meikle, Liquid chromatography-tandem mass spectrometry assay for androstenedione, dehydroepiandrosterone, and testosterone with pediatric and adult reference intervals.[Erratum appears in Clin Chem. 2011 Jul;57(7):1084], Clinical Chemistry 56(7) (2010) 1138-47.

[149] D. Klarmann, C. Eggert, C. Geisen, S. Becker, E. Seifried, T. Klingebiel, W. Kreuz, Association of ABO(H) and I blood group system development with von Willebrand factor and Factor VIII plasma levels in children and adolescents, Transfusion 50(7) (2010) 1571-80.
[150] G. Fideleff, M.G. Suarez, P.G. Sobrado, M.L. Celadilla, M. Mardyks, C.C. Rodriguez, A. Jelen, H.R. Boquete, H.L. Fideleff, Confirmation of neonatal screening: reference intervals and evaluation of methodological changes in TSH measurement, Journal of Pediatric Endocrinology & Metabolism 23(8) (2010) 759-64.
[151] J.A. Dawson, C.O. Kamlin, M. Vento, C. Wong, T.J. Cole, S.M. Donath, P.G. Davis, C.J. Morley, Defining the reference range for oxygen saturation for infants after birth, Pediatrics 125(6) (2010) e1340-7.
[152] R.D. Christensen, J. Jensen, A. Maheshwari, E. Henry, Reference ranges for blood concentrations of eosinophils and monocytes during the neonatal period defined from over 63 000 records in a multihospital health-care system, Journal of Perinatology 30(8) (2010) 540-5.
[153] A.M. Buchanan, F.J. Muro, J. Gratz, J.A. Crump, A.M. Musyoka, M.W. Sichangi, A.B. Morrissey, K. M'Rimberia J, B.N. Njau, L.J. Msuya, J.A. Bartlett, C.K. Cunningham, Establishment of haematological and immunological reference values for healthy Tanzanian children in Kilimanjaro Region, Tropical Medicine & International Health 15(9) (2010) 1011-21.
[154] D.P. Boer, Y.B. de Rijke, W.C. Hop, K. Cransberg, E.M. Dorresteijn, Reference values for serum creatinine in children younger than 1 year of age, Pediatric Nephrology 25(10) (2010) 2107-13.
[155] J.H. Wu, H.C. Chou, P.C. Chen, S.F. Jeng, C.Y. Chen, P.N. Tsao, C.J. Hsieh, H.M. Huang, W.S. Hsieh, Impact of delivery mode and gestational age on haematological

parameters in Taiwanese preterm infants, Journal of Paediatrics & Child Health 45(6) (2009) 332-6.

[156] S.E. Wiedmeier, E. Henry, M.C. Sola-Visner, R.D. Christensen, Platelet reference ranges for neonates, defined using data from over 47,000 patients in a multihospital healthcare system, Journal of Perinatology 29(2) (2009) 130-6.

[157] S. Sfar, M. Zaouali Ajina, S. Ben Mdalla, A. Baklouti, H. Knani, Z. Tabka, A. Zbidi, Establishment of reference serum insulin-like growth factor-I and insulin-like growth factor binding protein-3 levels of Tunisian children during childhood and adolescence, Journal of Endocrinological Investigation 32(8) (2009) 671-6.

[158] C.L. Ooi, N. Lepage, E. Nieuwenhuys, A.P. Sharma, G. Filler, Pediatric reference intervals for soluble transferrin receptor and transferrin receptor-ferritin index, World Journal of Pediatrics 5(2) (2009) 122-6.

[159] D.S. Lai, S.C. Chen, Y.H. Chang, C.Y. Chen, J.B. Lin, Y.J. Lin, S.F. Yang, C.C. Yang, W.K. Chen, D.B. Lin, Pediatric reference intervals for several biochemical analytes in school children in Central Taiwan.[Erratum appears in J Formos Med Assoc. 2010 Jan;109(1):89 Note: Chen, Shivan-Chih [corrected to Chen, Shiuan-Chih]], Journal of the Formosan Medical Association 108(12) (2009) 957-63.

[160] J. Jopling, E. Henry, S.E. Wiedmeier, R.D. Christensen, Reference ranges for hematocrit and blood hemoglobin concentration during the neonatal period: data from a multihospital health care system, Pediatrics 123(2) (2009) e333-7.

[161] K.L. Johnson-Davis, S.J. Moore, W.E. Owen, J.M. Cutler, E.L. Frank, A rapid HPLC method used to establish pediatric reference intervals for vitamins A and E, Clinica Chimica Acta 405(1-2) (2009) 35-8.

[162] M. Heiduk, I. Page, C. Kliem, K. Abicht, G. Klein, Pediatric reference intervals determined in ambulatory and hospitalized children and juveniles, Clinica Chimica Acta 406(1-2) (2009) 156-161.

[163] I.M. Adetifa, P.C. Hill, D.J. Jeffries, D. Jackson-Sillah, H.B. Ibanga, G. Bah, S. Donkor, T. Corrah, R.A. Adegbola, Haematological values from a Gambian cohort--possible reference range for a West African population, International Journal of Laboratory Hematology 31(6) (2009) 615-22.

[164] K. Abrahamsson, U. Jodal, U. Sillen, Reference interval for serum creatinine in children and adolescents with myelomeningocele, Scand J Urol Nephrol 43(3) (2009) 233-5.

[165] S. Thayyil, S. Sheik, S.T. Kempley, A. Sinha, A gestation- and postnatal age-based reference chart for assessing renal function in extremely premature infants, Journal of Perinatology 28(3) (2008) 226-9.

[166] O.P. Soldin, J.R. Dahlin, E.G. Gresham, J. King, S.J. Soldin, IMMULITE 2000 age and sex-specific reference intervals for alpha fetoprotein, homocysteine, insulin, insulin-like growth factor-1, insulin-like growth factor binding protein-3, C-peptide, immunoglobulin E and intact parathyroid hormone, Clinical Biochemistry 41(12) (2008) 937-42.

[167] B. Shine, Use of routine clinical laboratory data to define reference intervals, Annals of Clinical Biochemistry 45(Pt 5) (2008) 467-75.

[168] M.M. Kushnir, A.L. Rockwood, J. Bergquist, M. Varshavsky, W.L. Roberts, B. Yue, A.M. Bunker, A.W. Meikle, High-sensitivity tandem mass spectrometry assay for serum estrone and estradiol, American Journal of Clinical Pathology 129(4) (2008) 530-9.

[169] J. Kratzsch, G. Schubert, F. Pulzer, R. Pfaeffle, A. Koerner, A. Dietz, M. Rauh, W. Kiess, J. Thiery, Reference intervals for TSH and thyroid hormones are mainly affected by age, body mass index and number of blood leucocytes, but hardly by gender and thyroid autoantibodies during the first decades of life, Clinical Biochemistry 41(13) (2008) 1091-8.

[170] G. Sobrero, L. Munoz, L. Bazzara, S. Martin, L. Silvano, S. Iorkansky, L. Bergoglio, C. Spencer, M. Miras, Thyroglobulin reference values in a pediatric infant population, Thyroid 17(11) (2007) 1049-54.

[171] M. Rauchenzauner, A. Schmid, P. Heinz-Erian, K. Kapelari, G. Falkensammer, A. Griesmacher, G. Finkenstedt, W. Hogler, Sex- and age-specific reference curves for serum markers of bone turnover in healthy children from 2 months to 18 years, J Clin Endocrinol Metab 92(2) (2007) 443-9.

[172] G. Lippi, G.L. Salvagno, S. Rugolotto, G.P. Chiaffoni, E.M. Padovani, M. Franchini, G.C. Guidi, Routine coagulation tests in newborn and young infants, J Thromb Thrombolysis 24(2) (2007) 153-5.

[173] T. Koskas, K. Souare, T. Ouahabi, D. Porquet, D. Chevenne, Reference intervals for follicle-stimulating hormone, luteinizing hormone and prolactin in children and young adults on the bioMerieux Mini-Vidas system, Clin Chem Lab Med 45(4) (2007) 541-5.

[174] A.P. Kong, G.W. Wong, K.C. Choi, C.S. Ho, M.H. Chan, C.W. Lam, P.C. Tong, R. Ozaki, W.Y. So, R.C. Ma, J.T. Lau, J.C. Chan, Reference values for serum levels of insulin-like growth factor (IGF-1) and IGF-binding protein 3 (IGFBP-3) and their ratio in Chinese adolescents, Clin Biochem 40(15) (2007) 1093-9.

[175] T. Arndt, S. Stanzel, A.C. Sewell, Paediatric age-dependent serum transferrin isoform distribution studied by HPLC, Clinical Laboratory 53(9-12) (2007) 575-82.

[176] S.A. Akuyam, H.S. Isah, W.N. Ogala, Evaluation of serum lipid profile of under-five Nigerian children, Ann Afr Med 6(3) (2007) 119-23.

[177] K. Kapelari, C. Kirchlechner, W. Hogler, K. Schweitzer, I. Virgolini, R. Moncayo, Pediatric reference intervals for thyroid hormone levels from birth to adulthood: a retrospective study, BMC Endocrine Disorders 8 (2008) 15.

[178] C. Castellani, T. Stojakovic, M. Cichocki, H. Scharnagl, W. Erwa, A. Gutmann, A.M. Weinberg, Reference ranges for neuroprotein S-100B: from infants to adolescents, Clinical Chemistry & Laboratory Medicine 46(9) (2008) 1296-9.

[179] P.M. Yip, M.K. Chan, J. Nelken, N. Lepage, G. Brotea, K. Adeli, Pediatric reference intervals for lipids and apolipoproteins on the VITROS 5,1 FS Chemistry System, Clin Biochem 39(10) (2006) 978-83.

[180] M.T. Vietri, M. Sessa, P. Pilla, M. Misso, D. Di Troia, A. Sorriento, N. Parente, A.M. Molinari, M. Cioffi, Serum osteocalcin and parathyroid hormone in healthy children assessed with two new automated assays, J Pediatr Endocrinol Metab 19(12) (2006) 1413-9.

[181] R.F. Ritchie, G.E. Palomaki, L.M. Neveux, T.B. Ledue, W.Y. Craig, S. Marcovina, O. Navolotskaia, Reference distributions for apolipoproteins AI and B and the apolipoprotein B/AI ratios: A practical and clinically relevant approach in a large cohort, Journal of Clinical Laboratory Analysis 20(5) (2006) 209-217.

[182] L. Quinto, J.J. Aponte, J. Sacarlal, M. Espasa, P. Aide, I. Mandomando, C. Guinovart, E. Macete, M.M. Navia, R. Thompson, C. Menendez, P.L. Alonso, Haematological and biochemical indices in young African children: in search of reference intervals, Trop Med Int Health 11(11) (2006) 1741-8.

[183] P. Monagle, C. Barnes, V. Ignjatovic, J. Furmedge, F. Newall, A. Chan, L. De Rosa, S. Hamilton, P. Ragg, S. Robinson, A. Auldist, C. Crock, N. Roy, S. Rowlands, Developmental haemostasis. Impact for clinical haemostasis laboratories, Thromb Haemost 95(2) (2006) 362-72.

[184] H. Berrahmoune, J.V. Lamont, B. Herbeth, P.S. FitzGerald, S. Visvikis-Siest, Biological determinants of and reference values for plasma interleukin-8, monocyte chemoattractant protein-1, epidermal growth factor, and vascular endothelial growth factor: Results from the STANISLAS cohort, Clinical Chemistry 52(3) (2006) 504-10.

[185] M.J. Kempers, A.S. van Trotsenburg, D.A. van Tijn, E. Bakker, B.M. Wiedijk, E. Endert, J.J. de Vijlder, T. Vulsma, Disturbance of the fetal thyroid hormone state has long-term consequences for treatment of thyroidal and central congenital hypothyroidism, J Clin Endocrinol Metab 90(7) (2005) 4094-100.

[186] L.O. Chailurkit, U. Suthutvoravut, P. Mahachoklertwattana, S. Charoenkiatkul, R. Rajatanavin, Biochemical markers of bone formation in Thai children and adolescents, Endocr Res 31(3) (2005) 159-69.

[187] E.S. Lugada, J. Mermin, F. Kaharuza, E. Ulvestad, W. Were, N. Langeland, B. Asjo, S. Malamba, R. Downing, Population-based hematologic and immunologic reference values for a healthy Ugandan population, Clin Diagn Lab Immunol 11(1) (2004) 29-34.

[188] M.W. Elmlinger, W. Kuhnel, M.M. Weber, M.B. Ranke, Reference ranges for two automated chemiluminescent assays for serum insulin-like growth factor I (IGF-I) and IGF-binding protein 3 (IGFBP-3), Clin Chem Lab Med 42(6) (2004) 654-64.

[189] R.S. Brinkworth, E. Whitham, H. Nazeran, Establishment of paediatric biochemical reference intervals, Ann Clin Biochem 41(Pt 4) (2004) 321-9.

[190] U. Schauer, F. Stemberg, C.H. Rieger, M. Borte, S. Schubert, F. Riedel, U. Herz, H. Renz, M. Wick, W. Herzog, Establishment of age-dependent reference values for IgA subclasses, Clin Chim Acta 328(1-2) (2003) 129-33.

[191] L. Perrone, S. Perrotta, P. Raimondo, J. Mucerino, C. De Rosa, M.C. Siciliani, N. Santoro, E. Miraglia del Giudice, Inappropriate leptin secretion in thalassemia: a potential cofactor of pubertal timing derangement, J Pediatr Endocrinol Metab 16(6) (2003) 877-81.

[192] Y. Najam, M. Khan, F. Ilahi, A. Alam, Distribution of T4 TSH values in children--the Shifa experience, JPMA J Pak Med Assoc 53(1) (2003) 26-8.

[193] P.F. de Paula, J.E. Barbosa, P.R. Junior, V.P. Ferriani, M.R. Latorre, V. Nudelman, L. Isaac, Ontogeny of complement regulatory proteins - concentrations of factor h, factor I, c4b-

binding protein, properdin and vitronectin in healthy children of different ages and in adults, Scand J Immunol 58(5) (2003) 572-7.

[194] G. Brabant, A. von zur Muhlen, C. Wuster, M.B. Ranke, J. Kratzsch, W. Kiess, J.M. Ketelslegers, L. Wilhelmsen, L. Hulthen, B. Saller, A. Mattsson, J. Wilde, R. Schemer, P. Kann, K.B. German, Serum insulin-like growth factor I reference values for an automated chemiluminescence immunoassay system: results from a multicenter study, Horm Res 60(2) (2003) 53-60.

[195] B. Blomme, E. Gerlo, B. Hauser, Y. Vandenplas, Disaccharidase activities in Belgian children: reference intervals and comparison with non-Belgian Caucasian children, Acta Paediatr 92(7) (2003) 806-10.

[196] H. Schlebusch, N. Liappis, E. Kalina, C. Klein, High sensitive CRP and creatinine: Reference intervals from infancy to childhood, LaboratoriumsMedizin 26(5-6) (2002) 341-346.

[197] A.C. Muntau, M. Streiter, M. Kappler, W. Roschinger, I. Schmid, A. Rehnert, P. Schramel, A.A. Roscher, Age-related reference values for serum selenium concentrations in infants and children, Clinical Chemistry 48(3) (2002) 555-60.

[198] K. Kawahara, S. Yokoya, Establishment of reference intervals of thyrotropin and free thyroid hormones during the first week of life, Clinical Pediatric Endocrinology 11(1) (2002) 1-9.

[199] U. Hubner, C. Englisch, H. Werkmann, H. Butz, T. Georg, S. Zabransky, W. Herrmann, Continuous age-dependent reference ranges for thyroid hormones in neonates, infants, children and adolescents established using the ADVIA centaurTM analyzer, Clinical Chemistry and Laboratory Medicine 40(10) (2002) 1040-1047.

[200] P.M. Crofton, A.E.M. Evans, N.P. Groome, M.R.H. Taylor, C.V. Holland, C.J.H. Kelnar, Inhibin B in boys from birth to adulthood: Relationship with age, pubertal stage, FSH and testosterone, Clinical Endocrinology 56(2) (2002) 215-221.

[201] S.A. Lagerstedt, D.R. Hinrichs, S.M. Batt, M.J. Magera, P. Rinaldo, J.P. McConnell, Quantitative determination of plasma c8-c26 total fatty acids for the biochemical diagnosis of nutritional and metabolic disorders, Mol Genet Metab 73(1) (2001) 38-45.

[202] A. Georges-Gobinet, S. Pedeboscq, H. Malanda, L. Bordenave, Assessment of thyrotropin, free thyroxine and free triiodothyronine concentrations in healthy children measured by radioisotopic assays, Immuno-Analyse et Biologie Specialisee 16(2) (2001) 119-121.

[203] M.W. Elmlinger, W. Kuhnel, H.G. Lambrecht, M.B. Ranke, Reference intervals from birth to adulthood for serum thyroxine (T4), triiodothyronine (T3), free T3, free T4, thyroxine binding globulin (TBG) and thyrotropin (TSH), Clin Chem Lab Med 39(10) (2001) 973-9.

[204] S.J. Clark, D.D. Deming, J.R. Emery, L.M. Adams, E.I. Carlton, J.C. Nelson, Reference ranges for thyroid function tests in premature infants beyond the first week of life, J Perinatol 21(8) (2001) 531-6.

[205] J. Ahmed, M.M. Zaman, M.A. Rouf, M.M. Hassan, S. Zareen, Reference value of immunoglobulins in healthy school children of Bangladesh, J Epidemiol 11(6) (2001) 263-5.

[206] A. Harmoinen, E. Ylinen, M. Ala-Houhala, M. Janas, M. Kaila, T. Kouri, Reference intervals for cystatin C in pre- and full-term infants and children, Pediatric Nephrology 15(1-2) (2000) 105-8.

[207] A. Haraldsson, C.M. Weemaes, S. Jonasdottir, O. Olafsson, G. Van De Wiel, J. Goertz, I. Klasen, Serum immunoglobulinD in infants and children, Scand J Immunol 51(4) (2000) 415-8.

[208] H. Finney, D.J. Newman, H. Thakkar, J.M. Fell, C.P. Price, Reference ranges for plasma cystatin C and creatinine measurements in premature infants, neonates, and older children, Arch Dis Child 82(1) (2000) 71-5.

[209] A. Sherriff, A. Emond, N. Hawkins, J. Golding, Haemoglobin and ferritin concentrations in children aged 12 and 18 months. ALSPAC Children in Focus Study Team, Arch Dis Child 80(2) (1999) 153-7.

[210] S. Shahabuddin, I.H. al Ayed, M.O. el-Rad, M.I. Qureshi, Lymphocyte subset reference ranges in healthy Saudi Arabian children, Pediatr Allergy Immunol 9(1) (1998) 44-8.

[211] J. Ngamphaiboon, A. Theamboonlert, Y. Poovorawan, Serum IgG subclass levels in a group of healthy Thai children, Asian Pac J Allergy Immunol 16(1) (1998) 49-55.

[212] N. Jagarinec, Z. Flegar-Mestric, B. Surina, D. Vrhovski-Hebrang, V. Preden-Kerekovic, Pediatric reference intervals for 34 biochemical analytes in urban school children and adolescents, Clin Chem Lab Med 36(5) (1998) 327-37.

[213] A. Cullen, B. Kiberd, T. Matthews, P. Mayne, H.T. Delves, M. O'Regan, Antimony in blood and urine of infants, J Clin Pathol 51(3) (1998) 238-40.

[214] A. Bokenkamp, M. Domanetzki, R. Zinck, G. Schumann, J. Brodehl, Reference values for cystatin C serum concentrations in children, Pediatric Nephrology 12(2) (1998) 125-9.

[215] M.E. Blohm, D. Vesterling-Horner, G. Calaminus, U. Gobel, Alpha 1-fetoprotein (AFP) reference values in infants up to 2 years of age, Pediatr Hematol Oncol 15(2) (1998) 135-42.

[216] B.M. Winklhofer-Roob, M.A. van't Hof, D.H. Shmerling, Reference values for plasma concentrations of vitamin E and A and carotenoids in a Swiss population from infancy to adulthood, adjusted for seasonal influences, Clinical Chemistry 43(1) (1997) 146-53.

[217] M. Rukgauer, J. Klein, J.D. Kruse-Jarres, Reference values for the trace elements copper, manganese, selenium, and zinc in the serum/plasma of children, adolescents, and adults, J Trace Elem Med Biol 11(2) (1997) 92-8.

[218] F.A. Khan, M. Dilawar, D.A. Khan, Reference values of common blood chemistry analytes in healthy population of Rawalpindi-Islamabad area, JPMA J Pak Med Assoc 47(6) (1997) 156-9.

[219] W.F. Blum, P. Englaro, S. Hanitsch, A. Juul, N.T. Hertel, J. Muller, N.E. Skakkebaek, M.L. Heiman, M. Birkett, A.M. Attanasio, W. Kiess, W. Rascher, Plasma leptin levels in healthy children and adolescents: dependence on body mass index, body fat mass, gender, pubertal stage, and testosterone, J Clin Endocrinol Metab 82(9) (1997) 2904-10.

[220] F. Rauch, B. Middelmann, F. Rosmalen, M.J. Seibel, E. Schonau, Free deoxypyridinoline in urine and serum--results in children and adolescents, Exp Clin Endocrinol Diabetes 104(5) (1996) 396-9.

[221] F.S. Jensen, L.T. Skovgaard, J. Viby-Mogensen, Identification of human plasma cholinesterase variants in 6,688 individuals using biochemical analysis, Acta Anaesthesiol Scand 39(2) (1995) 157-62.

[222] L.M. Adams, J.R. Emery, S.J. Clark, E.I. Carlton, J.C. Nelson, Reference ranges for newer thyroid function tests in premature infants, J Pediatr 126(1) (1995) 122-7.

[223] A. Mouzinho, C.R. Rosenfeld, P.J. Sanchez, R. Risser, Revised reference ranges for circulating neutrophils in very-low-birth-weight neonates, Pediatrics 94(1) (1994) 76-82.

[224] J. Miles, P. Riches, The determination of IgG subclass concentrations in serum by enzyme-linked immunosorbent assay: establishment of age-related reference ranges for cord blood samples, children aged 5-13 years and adults, Ann Clin Biochem 31(Pt 3) (1994) 245-8.

[225] Y.L. Lau, B.M. Jones, K.W. Ng, C.Y. Yeung, Percentile ranges for serum IgG subclass concentrations in healthy Chinese children, Clin Exp Immunol 91(2) (1993) 337-41.

[226] L. Jonetz-Mentzel, G. Wiedemann, Establishment of reference ranges for cortisol in neonates, infants, children and adolescents, Eur J Clin Chem Clin Biochem 31(8) (1993) 525-9.

[227] S. Kumar, S. Mazumder, S. Banerjee, A. Mukherjee, R. Lahiri, D.K. Mukherjee, Lipid profile of Indian children and adolescents, J Indian Med Assoc 101(7) (2003) 403-4, 406.

[228] M.W. Elmlinger, H.G. Lambrecht, W. Kuhnel, Evaluation of an automated chemiluminescence assay to measure serum erythropoietin and determination of age-dependent reference ranges, LaboratoriumsMedizin 23(5) (1999) 289-294.

[229] A. Juul, P. Dalgaard, W.F. Blum, P. Bang, K. Hall, K.F. Michaelsen, J. Muller, N.E. Skakkebaek, Serum levels of insulin-like growth factor (IGF)-binding protein-3 (IGFBP-3) in healthy infants, children, and adolescents: the relation to IGF-I, IGF-II, IGFBP-1, IGFBP-2, age, sex, body mass index, and pubertal maturation, J Clin Endocrinol Metab 80(8) (1995) 2534-42.

[230] W.G. Wood, J. Singer, H. Schmidt-Gayk, Establishment of reference intervals for enzyme catalytic activity concentration measurements at 37 degrees C--a practical approach, Clinical Laboratory 50(7-8) (2004) 455-75.

[231] S. Genc, F. Gurdol, S. Guvenc, Y. Kargi, Variations in serum cholinesterase activity in different age and sex groups, Eur J Clin Chem Clin Biochem 35(3) (1997) 239-40.

[232] E.M. Custer, C.A. Finch, R.E. Sobel, A. Zettner, Population norms for serum ferritin, Journal of Laboratory and Clinical Medicine 126(1) (1995) 88-94.

[233] A.I. Berkel, I. Tezcan, F. Ersoy, O. Sanal, Serum immunoglobulin G subclass values in healthy Turkish children and adults, Turk J Pediatr 36(3) (1994) 197-204.

[234] L. Theunis, H. Neels, F. Fischbach, E. Miravalles, Age-dependence, sex-dependence and reference values of HbA(1c) determined by a new method on Boehringer Mannheim/Hitachi instruments, Klinisches Labor 39(12) (1993) 1026-1028.

[235] M. Tommasi, L. Bacciottini, A. Benucci, A. Brocchi, A. Passeri, D. Saracini, A. D'Agata, G. Cappelli, Serum biochemical markers of bone turnover in healthy infants and children, Int J Biol Markers 11(3) (1996) 159-64.

[236] G. Wiedenann, L. Jonetz-Mentzel, R. Panse, Establishment of reference ranges for thyrotropin, triiodothyronine, thyroxine and free thyroxine in neonates, infants, children and adolescents, European Journal of Clinical Chemistry and Clinical Biochemistry 31(5) (1993) 277-288.

[237] A.K. Brown, L.A. Sleeper, S.T. Miller, C.H. Pegelow, F.M. Gill, M.A. Waclawiw, Reference values and hematologic changes from birth to 5 years in patients with sickle cell disease. Cooperative Study of Sickle Cell Disease, Arch Pediatr Adolesc Med 148(8) (1994) 796-804.

[238] P. Kogner, O. Bjork, E. Theodorsson, Plasma neuropeptide Y in healthy children: influence of age, anaesthesia and the establishment of an age-adjusted reference interval, Acta Paediatr 83(4) (1994) 423-7.

Appendix E: Supplementary documents of the publication included in Chapter 6 (Hoq *el al*. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. Biometrical Journal. 2020. (Submitted))

Supplementary Table 1: Average coverage across integer age and sex for Hoq *et al* and Royston's methods, for five scenarios and four sample sizes

Performance Measure		**Coverage**							
Limit		**Lower limit**				**Upper limit**			
Sample size		**100**	**200**	**400**	**1000**	**100**	**200**	**400**	**1000**
Scenario	**Method**								
1[a]	Hoq *et al*	63.0	84.1	**93.8**	**93.9**	65.3	83.5	**95.0**	**93.4**
	Royston	84.5	87.8	83.7	65.6	66.0	60.4	47.9	25.5
2[b]	Hoq *et al*	58.4	81.7	**95.1**	**95.5**	49.4	79.6	**95.1**	**94.5**
	Royston	76.3	74.4	81.4	86.1	63.6	69.4	72.5	67.8
3[c]	Hoq *et al*	79.7	**94.7**	**95.1**	**95.7**	79.4	**95.3**	**95.0**	**95.9**
	Royston	**85.1**	**85.0**	**89.4**	**90.5**	**85.5**	**89.1**	**89.1**	**91.5**
4[d]	Hoq *et al*	77.1	**92.7**	88.4	77.3	76.2	**90.8**	87.2	70.9
	Royston	76.3	84.0	83.4	75.3	73.8	76.0	73.6	60.3
5[e]	Hoq *et al*	68.6	87.4	84.5	67.2	68.2	83.4	82.1	67.1
	Royston	69.5	63.7	49.5	30.6	70.0	61.5	49.3	34.7

Coverage more than 90% are highlighted with bold font.

a. Linear in age (separately by sex), SD increasing linearly with age e.g. Creatinine.
b. Nonlinear in age (separately by sex), SD increasing nonlinearly with age e.g. Alkaline Phosphatase.
c. Linear in age, constant variance, no sex difference e.g. Potassium.
d. Nonlinear in age, SD increasing linearly with age, sex difference e.g. Total Protein.
e. Nonlinear in age, constant variance, no sex difference e.g. Phosphate.

Supplementary Table 2: Average 95% confidence interval width for lower and upper reference limits for Hoq *et al* and Royston's methods across integer age and sex, for five scenarios and four sample sizes

Performance Measure		**Average confidence interval width**							
Limit		**Lower limit**				**Upper limit**			
Sample size		**100**	**200**	**400**	**1000**	**100**	**200**	**400**	**1000**
Scenario	**Method**								
1[a]	Hoq *et al*	7.6	8.8	8.6	5.0	8.0	8.6	8.6	4.9
	Royston	8.3	**6.2**	**4.4**	**2.8**	8.3	**6.2**	**4.4**	**2.8**
2[b]	Hoq *et al*	58.9	74.9	81.7	49.2	69.5	81.3	88.3	52.8
	Royston	68.3	**55.2**	**42.1**	**27.3**	**75.0**	**57.7**	**43.0**	**27.4**
3[c]	Hoq *et al*	0.35	0.38	0.26	0.15	0.35	0.39	0.26	0.15
	Royston	**0.22**	**0.16**	**0.12**	**0.08**	**0.22**	**0.16**	**0.12**	**0.08**
4[d]	Hoq *et al*	5.17	5.72	3.61	2.18	5.11	5.48	3.54	2.15
	Royston	**3.73**	**2.76**	**1.98**	**1.25**	**3.75**	**2.75**	**1.98**	**1.25**
5[e]	Hoq *et al*	0.17	0.20	0.14	0.09	0.18	0.19	0.14	0.08
	Royston	**0.13**	**0.09**	**0.07**	**0.04**	**0.13**	**0.09**	**0.07**	**0.04**

Lowest confidence interval width between Hoq *et al* and Royston's method was highlighted with bold font.

a. Linear in age (separately by sex), SD increasing linearly with age e.g. Creatinine.
b. Nonlinear in age (separately by sex), SD increasing nonlinearly with age e.g. Alkaline Phosphatase.
c. Linear in age, constant variance, no sex difference e.g. Potassium.
d. Nonlinear in age, SD increasing linearly with age, sex difference e.g. Total Protein.
e. Nonlinear in age, constant variance, no sex difference e.g. Phosphate.

Supplementary Figures 1 to 8

FIGURE S-1: Average estimated and true RIs by age, sex, sample size and method for scenario 1

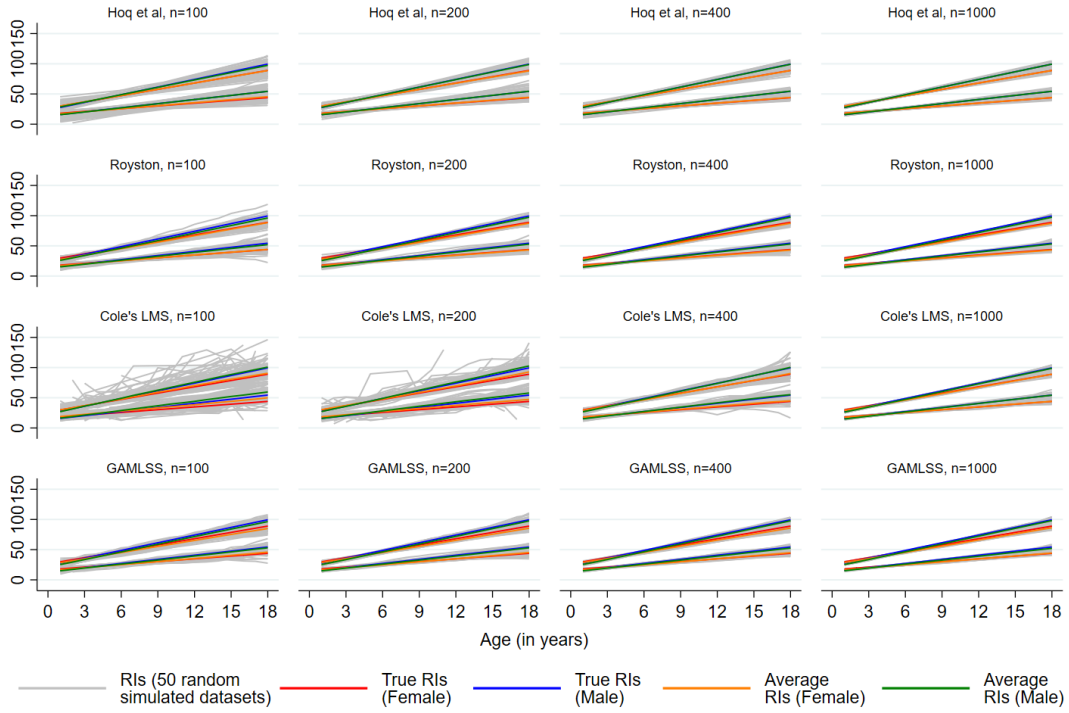


FIGURE S-2: Average estimated and true RIs by age, sex, sample size and method for scenario 2

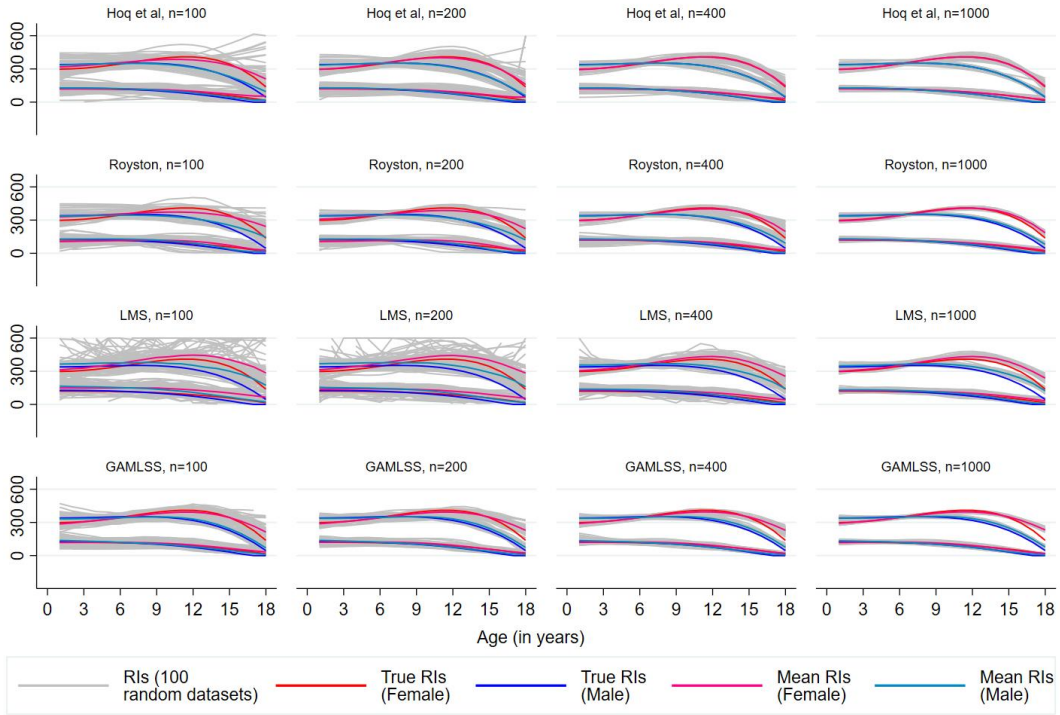


FIGURE S-3: Average estimated and true RIs by age, sex, sample size and method for scenario 3

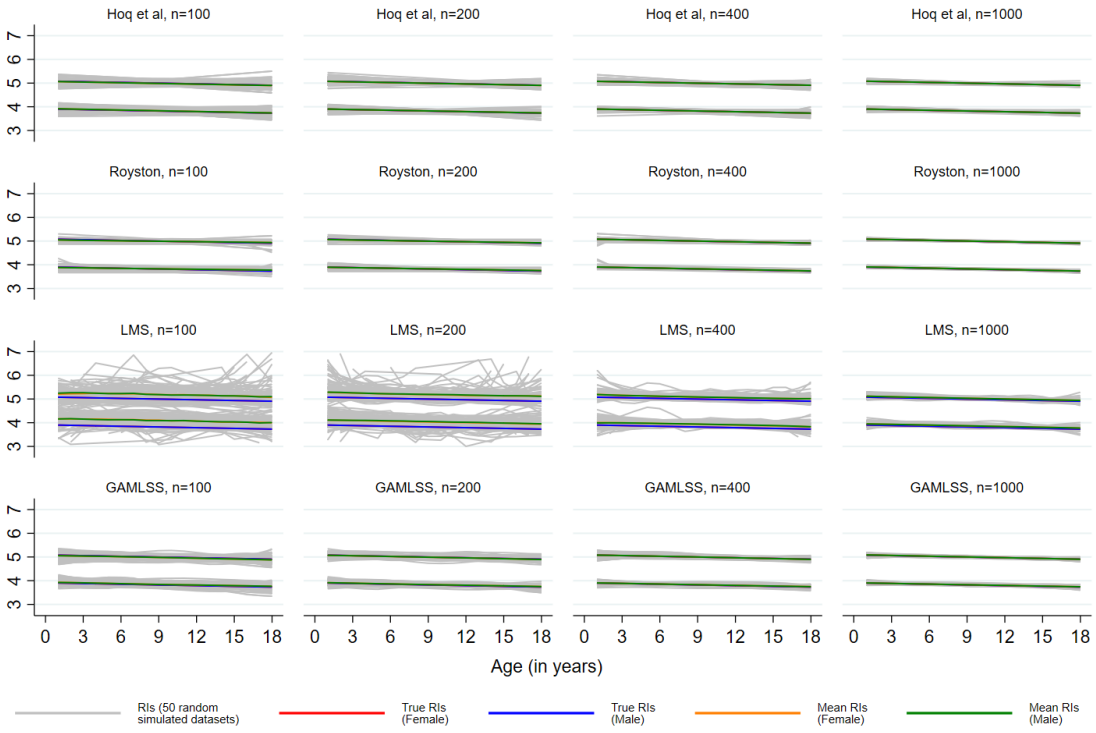


FIGURE S-4: Average estimated and true RIs by age, sex, sample size and method for scenario 4

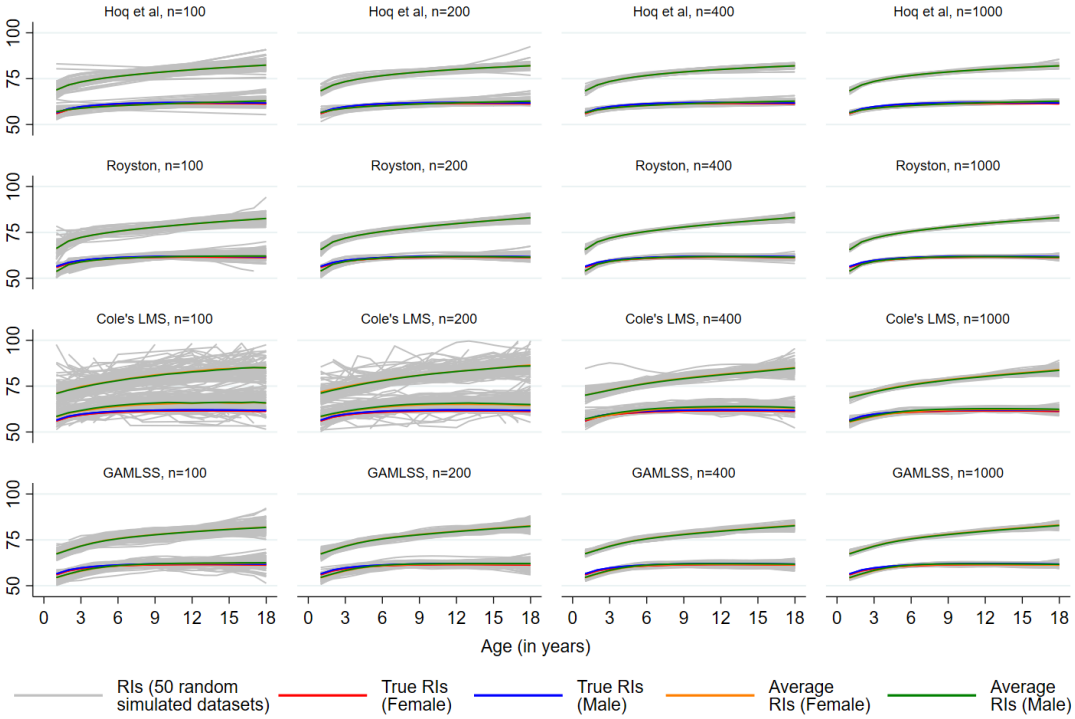


FIGURE S-5: Average estimated and true RIs by age, sex, sample size and method for scenario 5

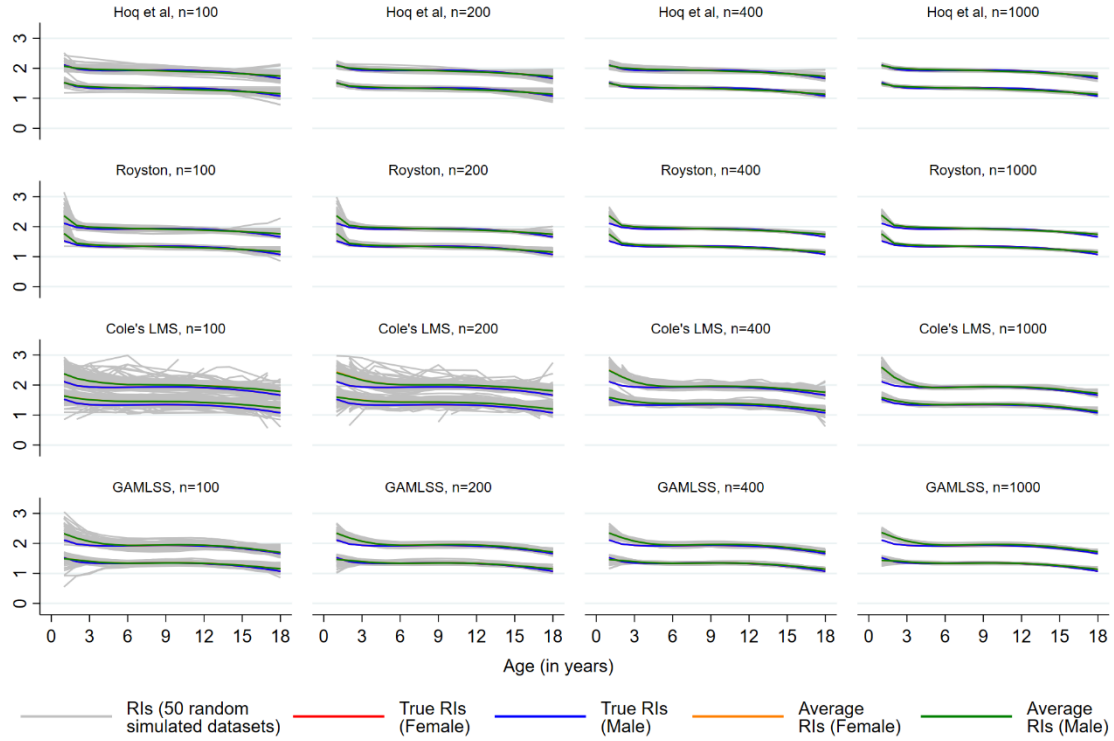
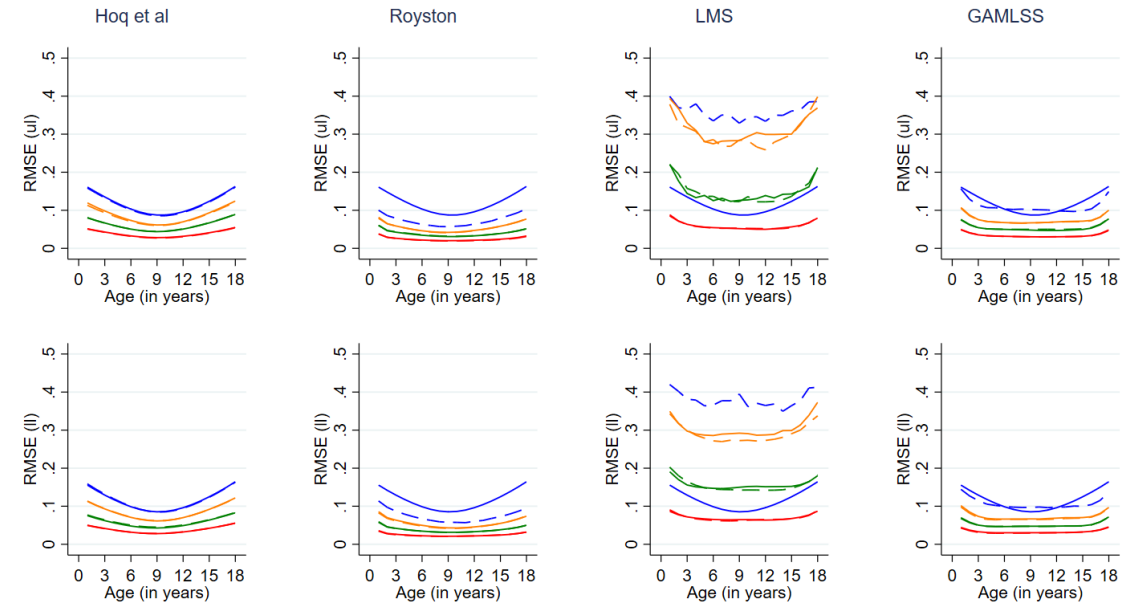
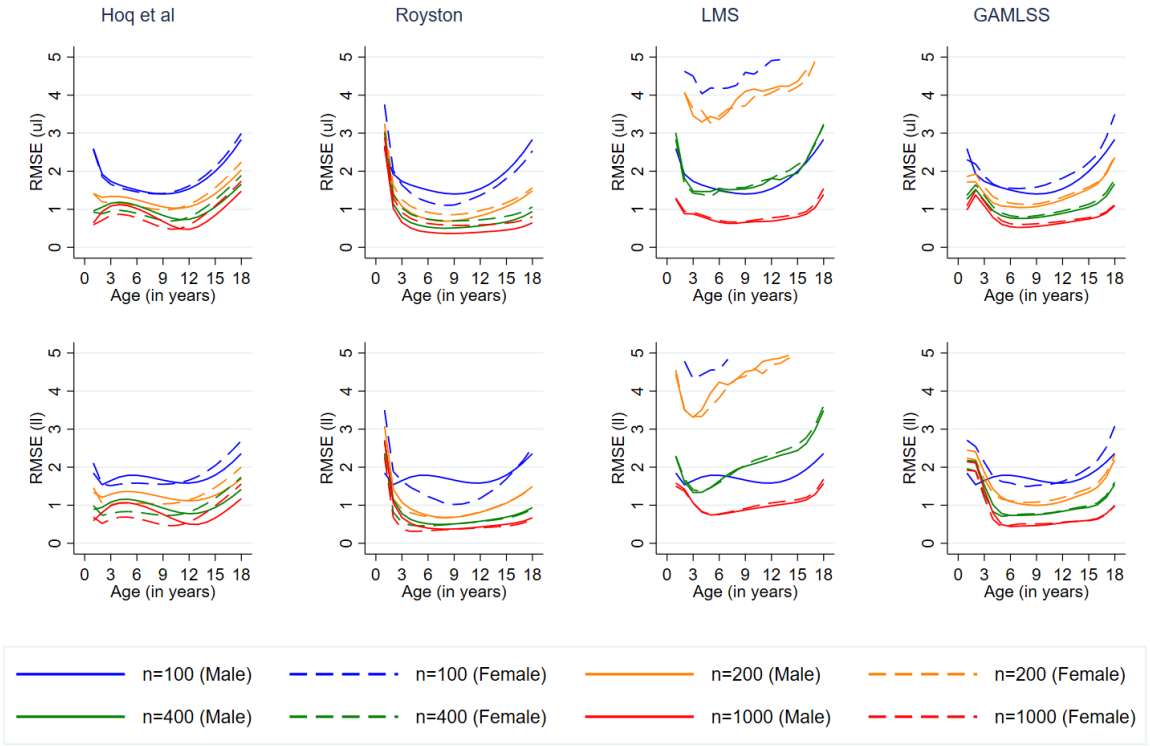


FIGURE S-6: Root Mean Square Error by age, sex, sample size and method for scenario 3



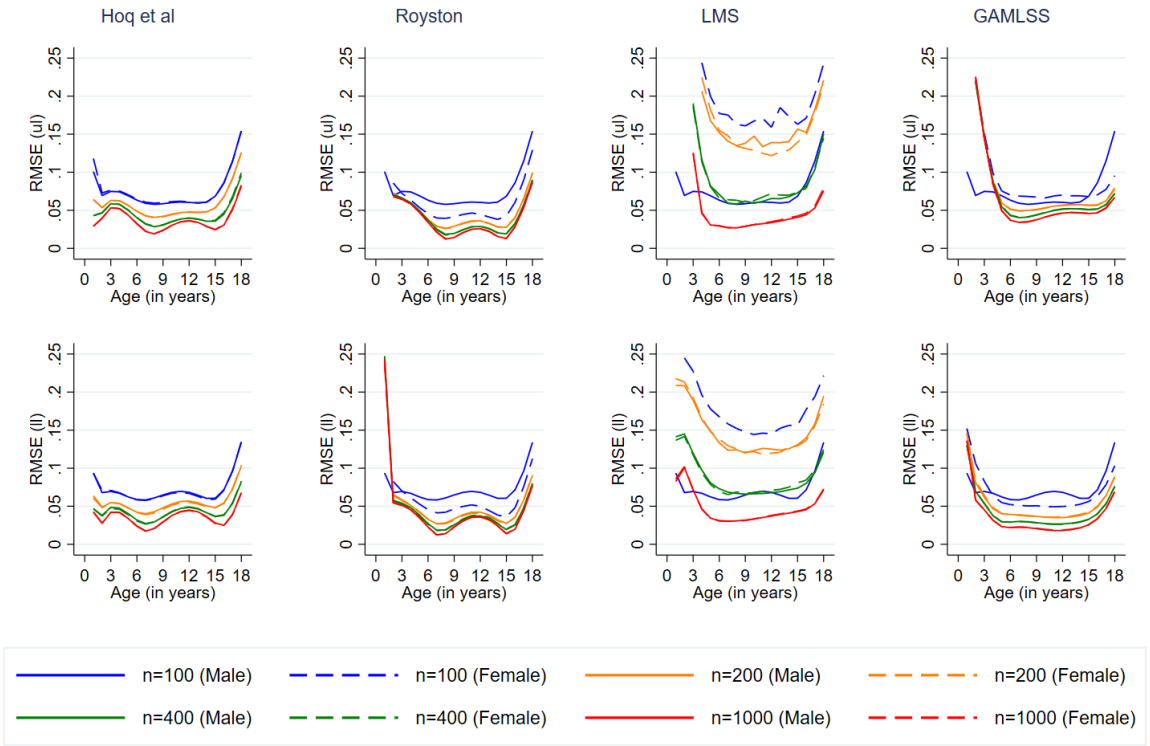
Values of RMSE > 0.5 not shown

FIGURE S-7: Root Mean Square Error by age, sex, sample size and method for scenario 4



Values of RMSE > 5 were not shown

FIGURE S-8: Root Mean Square Error by age, sex, sample sizes and methods for scenario 5



Values of RMSE > 0.25 were not shown

Supplemental Document 1: Stata and R codes of all four methods

```
A: Stata Code for Royston's method
* Developed by Monsurul Hoq for the purpose of the comparing four methods for estimating continuous RI.

* The following program is applicable for estimating RIs (2.5th - 97.5th) using age and sex (where 1 = Male and 0 = Female).
* A brief discussion of the method is include in the masnucript.
* Further details including "xrigls" and "xriml" can be found in Royston et al, JRSS 1998. 161(1): p. 79-101.


clear
set more off
cap log close
version 15.1



cap program drop rri
program rri
       * step 1: identify the best fitting model for mean and SD
       * For the entire code 1=analyte, 2=age and 3=sex
              xrigls `1' `2', nograph fp(m:df 4, s:df 4) centile(2.5 97.5) // xrigls is a user-defined package in Stata develped
by P. Royston (JRSS 1998. 161(1): p. 79-101)
              local m_fp `r(mpow)' // store the polynomial powers of age for the mean
              local s_fp `r(spow)' // store the polynomial powers of age for the standard deviation

       if "`m_fp'"=="" { // If no or linear association with age and constant SD
              xriml `1' `2', dist(n) se centile(2.5 97.5) nograph noscaling // xriml is a user-defined package in Stata develped
by P. Royston (JRSS 1998. 161(1): p. 79-101)
               }

       else {

       * Step 2: Assess the difference in mean response by sex for the model
       * using the best fitting power variables of age in the previous step

              if strlen("`m_fp'") <= 2 { // if the 1-degree polynomials were identified for the mean
                     fp <`2'>, replace powers(`m_fp') dim(1): reg `1' `3' <`2'>
                     mat b  = e(b) // store the estimates of b coefficients
                     mat se = e(V) // store the variance-covarience matrix
                     }

              else if strlen("`m_fp'") > 2 { // if the 2-degree polynomials were identified for the mean
```

```
                fp <`2'>, replace powers(`m_fp') dim(2): reg `1' `3' <`2'>
                mat b  = e(b)
                mat se = e(V)
                }

        estimate store a // store the findings from the regression y = a + b1*age+b2*sex

        * store the p-value corresponding to the b-coefficient of sex
        local t_sex  = b[1,1] / sqrt(se[1,1])
        local p_sex  = tprob(e(N)-2,`t_sex')

        di `p_pex'

* Step 3: Is there an interaction between the age (selected degree of FP models) and sex ?
*(so far the power variables selected in the -mfpi- were consistent with -mfp-)

reg `1' c.`2'_?##i.`3'  // here `2'_? will recall the power variables of age i.e `m_fp'
estimate store b // store the findings from the regression y = a + b1*age+b2*sex+b3*age*sex
lrtest a b, stats // compare the two models with or without interaction term
local p_int `r(p)' // Ho: No difference in deviance between the model significant. Store the p-value

di `p_int'

if `p_sex' > 0.05 & `p_int' > 0.05 { // if both sex and interaction with sex is insignificant

        if "`s_fp'"==""  { // if SD is constant
                xriml `1' `2', dist(n) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling // normal distribution
                local dev_n = 2*e(ll)
                xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling // exponential normal
distribution
                local dev_en = 2*e(ll)
                xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling // modulus exponential normal
distribution
                local dev_men = 2*e(ll)

                *Comparison of deviance between the 3 modeles
                local diff1 = abs(`dev_n' - `dev_en')
                local diff2 = abs(`dev_en' - `dev_men')
                local p1 = 1-chi2(1,`diff1')
                local p2 = 1-chi2(1,`diff2')

                * Selection of the final model
                if `p1' > 0.05 {
                        xriml `1' `2', dist(n) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
```

```
                }
        else if (`p1' <.05) & (`p2' > .05) {
                xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
                }
        else {
                xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
                }
        }

    else { // if s is a function of age
        xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
        local dev_n = 2*e(ll)
        xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
        local dev_en = 2*e(ll)
        xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
        local dev_men = 2*e(ll)
        local diff1 = abs(`dev_n' - `dev_en')
        local diff2 = abs(`dev_en' - `dev_men')
        local p1 = 1-chi2(1,`diff1')
        local p2 = 1-chi2(1,`diff2')
        if `p1' > 0.05 {
                xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
                }
        else if (`p1' <.05) & (`p2' > .05) {
                xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
                }
        else {
                xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
                }
        }
    }

else if `p_sex' < 0.05 & `p_int' > 0.05 { // there is a difference by sex but not seperate by sex

    if "`s_fp'"=="" { // if S is constant
        xriml `1' `2', dist(n) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
        local dev_n = 2*e(ll)
        xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
        local dev_en = 2*e(ll)
        xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
        local dev_men = 2*e(ll)
        local diff1 = abs(`dev_n' - `dev_en')
        local diff2 = abs(`dev_en' - `dev_men')
        local p1 = 1-chi2(1,`diff1')
```

```
                local p2 = 1-chi2(1,`diff2')
                if `p1' > 0.05 {
                        xriml `1' `2', dist(n) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
                        }
                else if (`p1' <.05) & (`p2' > .05) {
                        xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling  covar(m:`3', s:`3')
                        }
                else {
                        xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling  covar(m:`3', s:`3')
                        }
                }

        else { // if s is a function of age
                xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
                local dev_n = 2*e(ll)
                xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
                local dev_en = 2*e(ll)
                xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling covar(m:`3', s:`3')
                local dev_men = 2*e(ll)
                local diff1 = abs(`dev_n' - `dev_en')
                local diff2 = abs(`dev_en' - `dev_men')
                local p1 = 1-chi2(1,`diff1')
                local p2 = 1-chi2(1,`diff2')
                if `p1' > 0.05 {
                        xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling covar(m:`3',
s:`3')
                        }
                else if (`p1' <.05) & (`p2' > .05) {
                        xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling  covar(m:`3',
s:`3')
                        }
                else {
                        xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling  covar(m:`3',
s:`3')
                        }
                }
        }

    else if `p_int' < 0.05 { // seperately by sex
        *since its not possible to include interaction term in the model, RIs are estimated seperately by sex
        if "`s_fp'"=="" { // if S is consistent
                tempfile male female
                preserve
                keep if `3'==1
```

```
xriml `1' `2', dist(n) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling
local dev_n = 2*e(ll)
xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
local dev_en = 2*e(ll)
xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
local dev_men = 2*e(ll)
local diff1 = abs(`dev_n' - `dev_en')
local diff2 = abs(`dev_en' - `dev_men')
local p1 = 1-chi2(1,`diff1')
local p2 = 1-chi2(1,`diff2')
if `p1' > 0.05 {
       xriml `1' `2', dist(n) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
       }
else if (`p1' <.05) & (`p2' > .05) {
       xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
       }
else {
       xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
       }
save `male', replace
restore
keep if `3'==0
xriml `1' `2', dist(n) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling
local dev_n = 2*e(ll)
xriml `1' `2', dist(en) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
local dev_en = 2*e(ll)
xriml `1' `2', dist(men) fp(m:`m_fp')  se centile(2.5 97.5) nograph noscaling
local dev_men = 2*e(ll)
local diff1 = abs(`dev_n' - `dev_en')
local diff2 = abs(`dev_en' - `dev_men')
local p1 = 1-chi2(1,`diff1')
local p2 = 1-chi2(1,`diff2')
if `p1' > 0.05 {
       xriml `1' `2', dist(n) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling
       }
else if (`p1' <.05) & (`p2' > .05) {
       xriml `1' `2', dist(en) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling
       }
else {
       xriml `1' `2', dist(men) fp(m:`m_fp') se centile(2.5 97.5) nograph noscaling
       }
append using `male'
}
```

```
else { // if s is a function of age
       tempfile male female
       preserve
       keep if `3'==1
       xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
       local dev_n = 2*e(ll)
       xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
       local dev_en = 2*e(ll)
       xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
       local dev_men = 2*e(ll)
       local diff1 = abs(`dev_n' - `dev_en')
       local diff2 = abs(`dev_en' - `dev_men')
       local p1 = 1-chi2(1,`diff1')
       local p2 = 1-chi2(1,`diff2')
       if `p1' > 0.05 {
              xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
              }
       else if (`p1' <.05) & (`p2' > .05) {
              xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
              }
       else {
              xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
              }
       save `male', replace
       restore
       keep if `3'==0
       xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
       local dev_n = 2*e(ll)
       xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
       local dev_en = 2*e(ll)
       xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp')  se centile(2.5 97.5) nograph noscaling
       local dev_men = 2*e(ll)
       local diff1 = abs(`dev_n' - `dev_en')
       local diff2 = abs(`dev_en' - `dev_men')
       local p1 = 1-chi2(1,`diff1')
       local p2 = 1-chi2(1,`diff2')
       if `p1' > 0.05 {
              xriml `1' `2', dist(n) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
              }
       else if (`p1' <.05) & (`p2' > .05) {
              xriml `1' `2', dist(en) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
              }
       else {
              xriml `1' `2', dist(men) fp(m:`m_fp', s: `s_fp') se centile(2.5 97.5) nograph noscaling
```

```
                        }
                    append using `male'
                    }
            }

      }

end
```

B: Stata codes for Hoq *et al*'s method

```
* Developed by Monsurul Hoq for the purpose of the comparing four methods for estimating continuous RI.

* The following program is applicable for estimating RIs (2.5th - 97.5th) using age and sex (where 1 = Male and 0 = Female).
* A brief discussion of the method is include in the masnucript.
* Further details can be found in Hoq et al, Clin Chem 2019 65(10).


clear
set more off
cap log close
version 15.1

cap program drop ri
program ri

            * Regress the mean of y on age and sex to identify the best fitting power variable of age
            * For the entire code 1=analyte, 2=age and 3=sex
            mfp, center(no, `2':mean): reg `1' `2' `3'

            local pw `e(fp_k1)' // store selected powers of age for the mean

            * Is the difference in mean response by sex significant ?
            mat b  = e(b) // store the estimates of b coefficients
            mat se = e(V) // store the variance-covarience matrix
            local col = `e(rank)'-1 // to understand the position of the sex variable in the model

            * store the p-value corresponding to the b-coefficient of sex
            local t_sex  = b[1,`col'] / sqrt(se[`col',`col'])
            local p_sex  = tprob(e(N)-2,`t_sex')

            * Is there an interaction between the age (selected powers) and sex ?
```

```
            * For a linear model
            if "`pw'"=="1" {  // `pw' stores the power variables as string
                  mfpi reg `1' `2', showmodel with(`3') linear(`2')
                  local p_int `r(Plin)'  // Ho: No difference in deviance between the model significant. Store the p-value
                  }
            * For a 1-Degree Fractional Polynomial model
            else if strlen("`pw'") <= 2 {
                  mfpi reg `1' `2', showmodel with(`3') fp1(`2')
                  local p_int `r(Pfp1)'
                  }
            * For a 2-Degree Fractional Polynomial model
            else if strlen("`pw'") > 2 {
                  mfpi reg `1' `2', showmodel with(`3') fp2(`2')
                  local p_int `r(Pfp2)'
                  }

            if `p_sex' > 0.05 & `p_int' > 0.05 { // both sex and interaction with sex is insignificant

                  fp <`2'>, fp("`pw'") replace all : reg `1' <`2'> // This will generate the power variables of age (age_?)

                  *estimate the RIs (i.e. upper and lower limits) applying quantile regression where power variables of age are
the covriates
                  qreg `1' `2'_?, quantile(2.5) vce(,bof)
                  predict `1'_ll
                  predict `1'_ll_se1, stdp

                  qreg `1' `2'_? , quantile(97.5) vce(,bof)
                  predict `1'_ul
                  predict `1'_ul_se1, stdp
                        }

            else if `p_sex' < 0.05 & `p_int' > 0.05 { // there is a difference by sex but not seperate by sex

                  fp <`2'>, fp("`pw'") replace all : reg `1' i.`3' <`2'>

                  qreg `1' `2'_? i.`3', quantile(2.5) vce(,bof)
                  predict `1'_ll
                  predict `1'_ll_se1, stdp

                  qreg `1' `2'_? i.`3', quantile(97.5) vce(,bof)
                  predict `1'_ul
                  predict `1'_ul_se1, stdp
                        }
```

```
            else if `p_int' < 0.05 { // seperate by sexage

                  fp <`2'>, fp("`pw'") replace all : reg `1' i.`3' <`2'>

                  qreg `1' c.`2'_?##i.`3', quantile(2.5) vce(,bof)
                  predict `1'_ll
                  predict `1'_ll_se1, stdp

                  qreg `1' c.`2'_?##i.`3', quantile(97.5) vce(,bof)
                  predict `1'_ul
                  predict `1'_ul_se1, stdp

                  }

      end

C: State code for Cole's LMS method
* The following code from ZWANG was update by Monsurul Hoq to select the EDF based on a cross-validation.

*! version 0.1 ZWANG Nov 1998, Menzies School of Health Research
*! version 0.2 M.S. Pearce Nov 1998, Child Health, University of Newcastle
*  upon Tyne
cap program drop colelms2_cx
program define colelms2_cx
      version 5.0
      local varlist "req ex min(1) max(1)"
      #delimit ;
      local options "AGEgroup(string) SEX(string) df(string)
      NOSmooth CENtile(string) save(string) ";
      #delimit cr
      parse "`*'"
      parse "`varlist'", parse(" ")
      tempvar xlog xrec mg mh ma sg sa sh A B sd sdlog sdr
      if "`centile'"=="" { local centile "2.5 10 25 50 75 90 97.5" ///
} age-sp
      gen `xlog'=ln(`1')
      gen `xrec'=1/`1'
      local GM "`1'"
      if "`save'"=="{local save "LMStemp" ///
}
      if "`sex'"!="{
            # delimit ;
            collapse (mean) "`1'" `xlog' `xrec' (sd) sd="`1'" `sdlog'=`xlog'
            `sdr'=`xrec' (count) num=`1', by(`sex' `agegroup');
```

```
       # delimit cr
       }
if "`sex'"==""{
       # delimit ;
       collapse (mean) "`1'" `xlog' `xrec' (sd) sd="`1'" `sdlog'=`xlog'
       `sdr'=`xrec' (count) num=`1', by(`agegroup');
       # delimit cr
       }
gen `mg'=exp(`xlog')
gen `mh'=1/`xrec'
gen `ma'=`1'
gen `sg'=`sdlog'
gen `sa' =sd/`mg'
gen `sh'=`sdr'*`mg'
gen `A'=ln(`sa' /`sh')
gen `B'=ln((`sa' *`sh')/(`sg'^2))
qui {
       gen L=-`A'/(2*`B')
       gen S=`sg'*exp(`A'*L/4)
       gen M=`mg'+(`ma'-`mh')*L/2+(`ma'-2*`mg'+`mg')*(L^2)/2

* Cross-validation to select the optimum EDF for fitting a model for mu
       forvalue i = 1/18 {
                     forvalue o = 0/1 {
                     forvalue b = 2/6 {
* Drop each interage age at a time and estimate the model using the reamining age and mean for df = 2 - 6

                     gam M `agegroup' if `sex'==`o' & `agegroup'!=`i', df(`b')
                     rename GAM_mu MS1m`i'`o'`b'
                     }
              }
       }

gen diff2 = .
gen diff3 = .
gen diff4 = .
gen diff5 = .
gen diff6 = .

* Calcuate the error in predicting the new data
forvalue i = 2/17 {
       forvalue o = 0/1 {
              forvalue b = 2/6 {
```

```
            local j = `i'-1
            local k = `i'+1
            sum MS1m`i'`o'`b' if inlist(age, `j' , `k') & sex==`o'
            replace MS1m`i'`o'`b' = r(mean) if age==`i' & sex==`o'
            replace diff`b'  = M - MS1m`i'`o'`b' if age==`i' & sex==`o'
            }
      }
}

* identify the EDF between 2 and 6 with minimum error in predicting the mean
forvalue i = 2/6 {
      gen diff_s`i' = sqrt((diff`i')^2)
      sum diff_s`i'
      }
preserve
collapse (mean) diff_s?
gen id=1
reshape long diff_s, i(id)
sort diff_s
local df = _j[1]
di "`df'"
restore

* use the the identified EDF for calculating the centiles
      if "`nosmooth'"=="{
            if "`sex'"=="{
                  gam L `agegroup', df(`df')
                  gen LS=GAM
                  gam M `agegroup', df(`df')
                  gen MS=GAM
                  gam S `agegroup', df(`df')
                  gen SS=GAM
                  }
            if "`sex'"!="{
                  quietly summ `sex'
                  local one=_result(5)
                  local two=_result(6)
                  gam L `agegroup' if `sex'==`one', df(`df')
                  gen LS=GAM if `sex'==`one'
                  gam L `agegroup' if `sex'==`two', df(`df')
                  replace LS=GAM if `sex'==`two'
                  gam M `agegroup' if `sex'==`one', df(`df')
                  gen MS=GAM if `sex'==`one'
                  gam M `agegroup' if `sex'==`two', df(`df')
```

```
                        replace MS=GAM if `sex'==`two'
                        gam S `agegroup' if `sex'==`one', df(`df')
                        gen SS=GAM if `sex'==`one'
                        gam S `agegroup' if `sex'==`two', df(`df')
                        replace SS=GAM if `sex'==`two'
                        }
                lab var LS "L smoothed"
                lab var MS "M smoothed"
                lab var SS "S smoothed"
                drop s_ r_ e_ GAM
                parse "`centile'", parse(" ")
                while "`1'" !="" {
                        local tcent = `1'/100
                        local new = subinstr("`1'", ".", "x",.)
                        local Z`new' = invnorm(`tcent')
                        gen C`new' = MS*((1+LS*SS*`Z`new'')^(1/LS))
                        mac shift
                        }
                }
        }
end
```

D: R code for GAMLSS

The following steps were applied to estimate RIs using GAMLSS method.

```
rm(list=ls())
#setwd("Y:/PhD/Simulation/ALP")
#call the packages
library(gamlss)
library(haven)
library(MASS)
source("extract_centiles.R")
SS=c(100,200,400,1000)
for (k in 1:4) {
  for (i in 1:1000) {
    S=Sys.time()
   tryCatch(
      {
        print(i)
# reading the data
        datafile1 <-read_stata(paste0("Data/",SS[k],"/size",SS[k],"_sim",i,".dta"))
```

```
        alp_all <- subset(datafile1, select=c(age,sex,alp))
        alp_all0 <- na.omit(alp_all)
        alp_all0$alp[alp_all0$alp<1]<-1
        alp_all0$id<-1:nrow(alp_all0)
# finding the hyper parameters for mean and SD i.e. edf for model including age and sex
        m1 <- quote(gamlss(alp~cs(age,df=p[1])+as.factor(sex), data=alp_all0, family=BCT(), control=gamlss.control(trace=FALSE)))
        op <-find.hyper(model=m1, par=c(3), lower=c(1), steps=c(.1), k=log(SS[k]))
        m2 <- gamlss(alp~cs(age,df=op$par[1])+as.factor(sex),data=alp_all0, family=BCT())
# assess the difference in mean by sex
        m3<-stepGAIC.VR(m2)
        m3$anova
# assess the interaction between age and sex
        m4<-stepGAIC.VR(m2, scope=list(lower=~cs(age,df=op$par[1]), upper=~(cs(age,df=op$par[1])+as.factor(sex))^2))
        m4$anova
# assess the difference in SD by sex
        m5<-stepGAIC.VR(m4, what="sigma",scope=~(cs(age,df=op$par[1])+as.factor(sex)))
        m5$anova
# Find the hyper parameter for mean and SD for the selected model
        m6 <- quote(gamlss(m4$mu.formula,sigma.fo=m5$sigma.formula, data=alp_all0, family=BCT(),
control=gamlss.control(trace=FALSE)))
        op <-find.hyper(model=m6, par=c(3,3), lower=c(1,1), steps=c(.1,.1), k=log(SS[k]))
        m7 <- gamlss(m4$mu.formula,sigma.fo=m5$sigma.formula,data=alp_all0, family=BCT())

# calculate the centile
        o1<-extract_centiles(m7,xvar=alp_all0$age,idvar=alp_all0$id,groupvar = alp_all0$sex, col.cent=1,
                             cent=c(2.5,97.5),
                             lty.centiles=c(3,3, 2,2,1,2, 2,3, 3),
                             lwd.cent=c(1,1,1.5,1.5,2,1.5,1.5,1,1),
                             ylab = "ALP",
                             xlab = "Age (years)",
                             legend = F,
                             xlim = c(min(alp_all0$age),18),
                             save=F,
                             las=1)
        results<-read.csv("Coordinates for selected percentiles.csv" )
        write.csv(results,paste0("Results/",SS[k],"/methods5_size",SS[k],"_sim",i,".csv"), row.names = FALSE)

      }, error=function(e){///
})
  }
}
```

Appendix F: Author accepted manuscript published in Clinical Biochemist reviews (Hoq *et al*. Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations. *Clinical Biochemist Review*. 2020; 41 (2).)

Review Article

Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations

Short title: Review of Paediatric Reference Intervals

Monsurul Hoq[1, 2, 3], Susan Matthews[4], Susan Donath[1, 3], John Carlin[1, 3], Vera Ignjatovic[2, 3], *Paul Monagle[2, 3, 5]

[1] Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia

[2] Haematology Research, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia

[3] Department of Paediatrics, The University of Melbourne, Parkville, VIC 3052, Australia

[4] International Centre for Point of Care Testing, Flinders University, Bedford Park, Australia.

[5] The Royal Children's Hospital, Parkville, VIC 3052, Australia.

* For correspondence:

Prof Paul Monagle

Address:

Haematology Research Laboratory,

Murdoch Children's Research Institute,

50 Flemington Road, Parkville, Victoria, 3052, Australia

Phone: +61 3 9345 5165

Email: paul.monagle@rch.org.au

Abstract:

Establishing paediatric reference intervals (RIs) is a challenging task due to difficulties in subject recruitment, collection of adequate blood volume and the inherent, physiological changes of many biomarkers with age. Despite these challenges, several national and international initiatives have demonstrated a) the feasibility of prospectively designed paediatric RI studies, b) the development of continuous RIs and c) the comparison of reference values across analyser types to harmonise paediatric RIs. Whilst these studies have improved the interpretation of paediatric test results and compliance with international accreditation (ISO15189) requirements, several gaps and challenges in translating current paediatric RIs into routine laboratory practice remain. Future priorities for paediatric RI studies include: a) determination of the impact of discrete versus continuous RIs; analyser-specific versus harmonised RIs, and prospective collection versus data mining on the proportion of results outside the RIs, b) understanding the clinical implications of analyser-to-analyser variation in reference values and use of evidence-based paediatric harmonised RIs, where applicable, c) adaptation of laboratory information systems to incorporate continuous RIs, d) further understanding of the biological variation in paediatric biomarkers, e) studies to address the paucity of accurate data for neonatal RI development, f) periodic demonstration of RIs being clinically 'fit-for purpose' and g) agreement and policy updates

for use of modern, best practice statistical methods in estimation of paediatric RIs. Furthermore, *in vitro* diagnostic (IVD) manufacturers may require incentivised paediatric RI studies and publications through co-ordinated grants and collaboration at end user sites to reduce the burden on sole users.

Introduction

Laboratory test results are critical to modern clinical medicine, with claims that nearly 70% of physicians' medical decisions based on information provided by laboratory test reports.[1] The importance of laboratory test results is further evident from the sheer number of pathology tests performed. According to the Royal College of Pathologists of Australasia (RCPA), approximately 500 million laboratory tests are performed each year in Australia and more than 11 million Australians have at least one test per year.[2]

A test result by itself is of little value unless it is reported with a reliable and appropriate clinically related interpretation. A reference interval (RI) is one of such clinical assessment tools used frequently to interpret test results. Statistically, a RI is defined as a range that includes α% ($0 < \alpha < 100$) of the reference population, where lower and upper limits are $(100-\alpha)/2$ and $(100+\alpha)/2$ centiles respectively.[3] The interval between the two extreme percentiles represents the central part of the population, and is frequently estimated by performing the relevant test/s on healthy individuals or a defined reference population. The basic concept of a RI is that observations that are outside the interval are considered unusual or abnormal, with the understanding that this will falsely classify (100-α) % of the reference population as "unusual".

The appropriateness of any RI is a critical factor in the clinical interpretation of a test result and subsequent patient diagnosis and treatment. For this reason, the International Organisation for Standardisation (ISO) have specified requirements for all aspects of laboratory testing, including pre-analytical, analytical and post-analytical (e.g. RIs) phases.[4, 5] According to these requirements, clinical laboratories need to periodically demonstrate the appropriateness of the RIs to the population served and the measurement system in use.[4, 5] The establishment of a RI is a complex and time-consuming process, especially for children.[6] The RI establishment process is broadly classified into four steps: (i) definition of a reference population; (ii) selection of the reference individuals, (iii) collection of reference samples; (iv) laboratory testing of the reference samples according to routine, standardised operating procedures; and (v) application of statistical methods to the laboratory results generated in step (iv).[6] For children, the determination of RIs is difficult because it requires recruitment of adequate numbers of healthy children, collection of relatively large blood volumes proportional to the patient's blood reserve and an understanding of the physiological changes in biomarkers associated with normal growth and development.[7-9]

One of the main steps in establishing paediatric RIs is recruiting healthy children representative of the community that a specific laboratory or hospital serves.[10] Ideally, each child in the community should have the same probability of being selected as a representative of the reference population. Hence, recruitment of children should occur prospectively from a community, having no current interaction with a health service. However, this is extremely difficult, both ethically and pragmatically. A reasonable compromise is to actively recruit otherwise heathy children who present to health care services for minor surgical procedures, who would usually not be having clinically investigative blood tests because they are assumed to be normal. From a neonatal perspective, healthy term newborns on the post-natal

wards of hospitals represent an ideal sampling population. However, very low consent rates, restrictions in the amount of blood that can be safely collected from children with respect to body weight, a requirement for specialist phlebotomy procedures and staff combined with other ethical (e.g. single collection attempt) and logistical challenges (e.g. storage, transportation and cost), limit the ease with which many laboratories worldwide can attain prospective collection of paediatric blood samples.[11, 12] This is particularly true for smaller, non-specialised laboratories that service regional health centres. An alternative approach has been to conduct retrospective studies, with data extraction, often referred to as "data mining", from existing laboratory information systems (LIS) or testing leftover samples from laboratory storage, have been used in the estimation of paediatric RIs.[13-16] Data mining has the advantage of generating large datasets compared to the direct approach of obtaining and testing samples for the single purpose of establishing RIs.[16] However, despite the large sample size and rigorous statistical analysis, the representativeness of the findings and thus generated RIs, is influenced by the fact that the majority of the blood tests were initially conducted for investigation of possible disease states, an issue that may affect paediatric testing more than adult.[15, 16]

In children, changes in specific biomarkers are associated with normal, physiological growth and development (e.g. creatinine and alkaline phosphatase (ALP)) and hence appropriate paediatric RIs must also reflect these changes from birth through adolescence.[13, 14, 17] Traditionally, RIs were estimated for different discrete age-groups to reflect these age-specific changes in biomarkers.[6, 18] However, several studies have identified that discrete age-group RIs are misleading, specifically when testing children whose age is close to the defined age-group cut-off values.[13, 14] For example, the continuous increase of creatinine concentration with increased muscle mass (and therefore age) or increase in ALP associated

with the osteoblastic activity during the pubertal growth spurt, make it evident that discrete age-group RIs alone are not sufficient to capture these physiologically based changes.[14, 19]

Despite recent activities to achieve standardisation of laboratory results, between-method differences remain for the same tests.[20-22] As sequential paediatric specimens on the same patient are often tested by different laboratories, the application of appropriate reagent and analyser-specific RIs for reliable clinical decision making is important. Most importantly, the clinical implication of any inter-analyser variation must be understood when making a decision based on an-analyser-specific RIs.[17] While there are initiatives to harmonise paediatric RIs used between different laboratories and/or different analysers, the decision to harmonise is often based on clinical consensus between pathologists in The United Kingdom or using Bhattacharya analysis in Australia, rather than clinical evaluation of test results that were outside the RIs.[22, 23]

Considering the challenges in establishing paediatric RIs, the process is likely to be practically limited to large scale, paediatric specialist centres with dedicated research resources or multi-national diagnostic manufacturers. Laboratories that analyse small number of paediatric specimens amongst predominantly adult cohorts and/or smaller regional laboratories may only be able to adopt previously published paediatric RIs. Hence the verification and validation of RIs in paediatric populations will play an important role in ensuring appropriate RIs are in use at local laboratory sites.

In this context, the aim of this review is to briefly outline the current status of paediatric RI studies, describe some of the existing gaps and challenges in paediatric RIs and highlight

important future priorities in the practical implementation of paediatric RIs, in order to improve clinical diagnosis and reduce patient risk.

Current Status

Population-based, prospective paediatric studies

The quality of RIs depends on the selection and recruitment of the individuals chosen to provide a specific RI, the control and quality of the laboratory's pre-analytical, analytical and post-analytical process and the statistical methods applied. While controlling all these aspects is challenging, several national and international initiatives have demonstrated that establishing paediatric RIs using samples collected prospectively from population-based studies is feasible. According to the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), several countries and regions have taken initiatives to establish paediatric RIs.[24] In Germany, the German Health Interview and Examination Survey for Children and Adolescents (KiGGS) presented the age and sex-specific percentile distribution of key laboratory parameters based on more than 14,000 blood and serum samples.[25] In the USA, the Children's Health Improvement through Laboratory Diagnostics (CHILDx) study is focusing on determining paediatric RIs for a number of clinical laboratory assays by collecting both blood and urine specimens from healthy subjects.[26] In Scandinavian countries, the Scandinavian Initiative for establishment of paediatric reference intervals (NORICHILD) has reported discrete age-group paediatric RIs based on blood collected from nation-wide community surveys.[27]

The Canadian Laboratory Initiative on Paediatric Reference Intervals (CALIPER) is one of the largest paediatric RIs projects completed to date and its numerous publications of paediatric RIs have contributed significantly to improving the clinical diagnosis and

interpretation of paediatric test results.[18] CALIPER has collected thousands of samples from healthy children in Canada across several ethnic groups and has established paediatric RIs on Abbott Architect analysers.[28] CALIPER has also mathematically transferred the established paediatric RIs on other analytical platforms, published a systematic review of the statistical methods applied in the estimation of RIs and a simulation study comparing statistical methods for estimation of age-group RIs.[29-31]

Australia, the Lifestyle Of Our Kids (LOOK) study, which is a community-based longitudinal study, followed a cohort of 852 healthy eight year old children and reported RIs for 37 analytes at age 8, 10 and 12 years.[32] Most recently, the Australian Harmonising Age Pathology Parameters in Kids (HAPPI Kids) study also reported continuous age-and sex-specific RIs as a function of age and sex for common biochemistry analytes across five commercially available analysers.[19, 33] The HAPPI Kids study was the first study to directly compare test results across analysers and provide scientific evidence in support of harmonising RIs for the majority of biochemistry analytes.[19, 33] The HAPPI Kids study has collected and tested samples from more than 5000 children with the aim to publish age and sex-specific RIs for common haematology and immunology analytes, as well as to investigate the impact of blood group on RIs.[19, 33] Furthermore, the HAPPI Kids team has conducted a systematic review of statistical methods, focussing on the choice between age as a discrete or continuous variable, and a simulation study comparing the statistical methods applied in the estimation of continuous RIs.[34, 35]

Addressing the lack of accurate paediatric RIs in China, the Pediatric Reference intervals in China (PRINCE) project aims to establish and verify paediatric RIs for 31 analytes using data collected from approximately 15,000 healthy children from 10 children's hospital.[36] To date

the project has evaluated the study feasibility and efficiency of the method using data from 602 children.[36]

Discrete age-group or continuous age-specific reference intervals

Considering the challenges in clinical diagnosis based on the discrete age-group RIs for biomarkers that change with age, several studies have established continuous age and sex-specific paediatric RIs.[13, 14, 33, 37, 38] Bussler *et al* established continuous RIs for alanine aminotransferase (ALT), aspartate aminotransferase (AST), and gamma-glutamyl transferase (GGT) as 3rd and 95th percentiles using a widely used semi-parametric statistical method ("LMS") based on data from 1746 healthy children (3131 cases) aged 11 months to 16 years.[37] Zierk *et al* reported continuous age and sex-specific reference intervals for 22 haematological and biochemical analytes also using an LMS-type method based on data available data from a hospital laboratory database for children aged 6 months to 18 years.[14] Recently, CALIPER reported continuous RIs for 38 biochemical markers using non-parametric quantile regression based on the data from the CALIPER database for children aged 6 months to < 19 years.[38] As mentioned, the HAPPI Kids study also established continuous age-specific RIs for 30 biochemistry analytes across five analysers using quantile regression where (fractional) power variables of age and sex were used as the covariates, using prospectively collected data from 360 healthy children aged 1 month to <18 years.[33] With this method, the HAPPI Kids study reported continuous age-specific RIs as a function of age and sex, enabling a RI to be calculated for any age.[33]

Analyser-specific or common paediatric reference intervals

Addressing inter-analyser variation, CALIPER study mathematically transferred and validated discrete age-grouped paediatric RIs from Abbott Architect analysers to other

systems including the Beckman Coulter DxC800, Ortho Vitros 5600, Roche Cobas 6000 and Siemens Vista 1500 analysers for common biochemical and immunological analytes.[28, 31, 39] The authors reported the transferred RIs were similar for most analytes to the RIs established on the Abbott Architect c800 while in some cases (e.g. lipase on Siemens Vista and apoB on Ortho Vitros) the reference limits were higher than those for the Abbott Architect.[31] The Australasian Association of Clinical Biochemists (AACB) and Royal College of Pathologists (RCPA) endorsed paediatric harmonised RIs for 9 biochemical analytes (i.e. sodium, serum and plasma potassium, chloride, bicarbonate, creatinine, calcium, phosphate, magnesium and ALP for children 0 week to <18 years), established based on the paediatric data from 15 laboratories (i.e. over 200,000 test results) using Bhattacharya analysis.[22] However, the comparison of analytical methods to estimate variability between manufacturers were performed using adult samples and may not be representative of paediatric population particularly for neonates.[22] Most recently, the HAPPI Kids study reported common paediatric RIs for 24 biochemical analytes in addition to analyser-specific RIs for 30 biochemical analytes on Ortho Vitros 5600, Abbott Architect c8000, Roche Cobas c701, Siemens ADVIA 1800, and Beckman Coulter AU5800.[19, 33] The HAPPI Kids study is the first study to provide scientific evidence of statistical and clinical significance of differences in age-specific mean reference values across analysers for the paediatric population[19]. This head-to-head comparison of reference values was possible as each blood sample was divided into five aliquots and tested directly on the 5 different analysers.[33]

Gaps and Challenges

Recruitment of healthy neonates and infants

Accurate and reliable paediatric RIs that reflect dynamic changes during the first few days, weeks and months of life, are still lacking. Difficulties in recruiting adequate numbers of

healthy term neonates and infants were acknowledged by all the major population-based paediatric RI studies.[18, 19, 40] The CALIPER paediatric RIs for neonates and infant (0 - < 1 year) are limited as children were recruited from outpatient clinics.[18] Similarly, studies from Denmark and Sweden addressed the limitation of small sample sizes (6 months to 18 years) by merging datasets to establish paediatric RIs.[40] The HAPPI Kids study successfully collected blood samples from term neonates within 72 hours of birth, albeit with low consent rates (16%) but was unsuccessful at collecting adequate samples from pre-term neonates born at 32- 36 weeks gestation (unpublished data). However, the HAPPI Kids study had no samples collected from children aged between 72 hours and one month of age.[19]

Pre-analytical and analytical processes for testing paediatric samples

Routine pre-analytical and analytical processes are important factors that require control when establishing paediatric RIs.[9] Following safe clinical practice, blood collection for the HAPPI Kids study was limited to less than 3% of the total blood volume for a child.[33] In addition, ethical consideration of younger children limited multiple or repeated blood collection attempts in this study.[19] Post-collection, standard operating procedures defined HAPPI Kids specimen processing, but nevertheless authors reported data exclusion due to pre-analytical interferences in certain subsets of samples (e.g. sodium, total protein (due to evaporations), and total bilirubin (due to light exposure)).[19] Adeli *et al* also raised concerns regarding the ability of the automated laboratory equipment to handle small volume specimens, leading to frequent manual processing and difficulties in standardising specimen processing.[17] This problem was also observed in the analysis of the HAPPI Kids specimens, where automated processing was overridden to prevent the analyser from aborting automated specimen analysis due to inadequate sample volumes.[19] In addition to operational pre-analytical and analytical concerns, intra-individual biological variation or total allowable

error for different analytes based on adult samples may not be applicable for the paediatric population, especially during the first few years of life.[17] Addressing the lack of data on biological variation specific to paediatric populations, the CALIPER study first published paediatric within-day biological variation and quality specification for 38 biomarkers using nested analysis of variation techniques.[41] Considering the challenges in recruiting healthing children using direct sampling approach, *Loh et al* proposed an indirect method for estimating within-individual biological variation and estimated between-individual biological variation for 22 biomarkers which were comparable to those reported in the CALIPER study.[42, 43]

Validation and assimilation of the continuous paediatric reference intervals in a routine laboratory

Whilst continuous age-specific RIs improve our understanding of the changes in biomarker concentrations with age, there is no current recommendation or endorsed guidelines for the validation of continuous age-specific RIs for use by a routine laboratory. In contrast, the CLSI published guidelines for validation is available for the validation of age-group RIs.[6] It is uncertain to what extent this guideline is applicable to continuous age-specific RIs. As such, the HAPPI Kids study developed a method to validate its published continuous age-specific RIs with some modification of the current guideline.[44] Publication and endorsement of a standard protocol for continuous RI validation would provide direction and support for the adoption of continuous paediatric RIs in local laboratories.

Even so, incorporating continuous RIs into existing LIS designed for discrete intervals remains challenging[45]. At present, LIS have limited capacity to visually display continuous reference curves or calculate continuous RIs from equations based on age and/or sex.[19] Moreover, some laboratories still report test results in a portable document format (PDF) or

hard copy, making it difficult to track sequential changes in test results appropriately over time. In addition to the challenges associated with the translation of continuous RIs into LIS, many pathology results are also viewed in other online systems, including electronic medical records and Medical Practitioner desktop or mobile software packages. In these situations, developments are not only required to the LIS, but also to prepare the RIs, display them if this is required in a rendered format (e.g. PDF or directly on-screen) and also systematically transfer the results to receiving systems that have the capacity to accurately and safely render the results. This poses additional complexity for the electronic management of patient results and applicable RIs, but as the diversity in receiving systems and software expands, standardisation is vital. Recent studies endorsing the use of continuous paediatric RIs provide impetus for the development of improved modern LIS functionality and associated integrated electronic medical records (EMR) and other mobile and desktop software used to view and interpret patient results at the medical practitioner interface.[13, 19, 37, 38]

Future Considerations

Further research addressing the gaps and challenges

Target studies to establish neonatal and infant reference intervals

Physiological change in neonates is very dynamic, with the first few days of life being noted as "the most dramatic period of change that occurs in human life".[46] As such, concentrations of respiratory and cardiovascular biomarkers can change within minutes after birth, whilst other biomarkers change more slowly as organ systems evolve to transition the neonate from the intrauterine environment to adult physiology.[46] These dramatic changes in biomarkers make the interpretation of pathological test results challenging, especially in the context of currently available paediatric RIs. As discussed previously, there is a further lack of reliable and accurate neonatal and infantile RIs. Targeted studies of children aged 0, 1, 2, 3, 7, 14 and

28 days, and during the 2nd, 3rd, 4th, 5th, 6th, 9th, and 12th months of life are required. In an effort to address the existing limitations of available data, the CALIPER study has undertaken the recruitment of expectant mothers to facilitate the monitoring of circulating biomarkers at birth and throughout the first year of life.[18] Hence, priority should be given to analytes with proven clinical utility during the neonatal and infancy periods.

Statistical methods applied in the estimation of RIs

There is a wide variety of statistical methods that have been applied in the estimation of age-specific RIs. These include Royston's parametric curves, Cole's LMS, generalised additive models for location, scale and shape (GAMLSS) and quantile regression.[3, 47-49] Underlying each method is a set of defined assumptions under which the method will provide optimum results. While there have been a few studies comparing the performance of different statistical methods based on available data, there is a lack of well-designed statistical simulation studies comparing the performance of these methods for different scenarios i.e. complexity of relationship between age and biomarkers, distribution of samples across age and sample size[48, 50]. The minimum sample size required to estimate age-specific paediatric RIs is also frequently not agreed upon. Royston proposed a formula for sample size calculation for estimating age-specific RIs using samples distributed uniformly across age and requiring the ratio of the standard error (SE) of the estimated limits to be no more than 10% of the standard deviation of the variation in the population.[3] A modification to the formula for various age distributions has also been proposed.[3, 51] However, this approach based on the SE of the limit is not applicable for LMS or GAMLSS. There is also no recommendation for sample size requirement when utilising these statistical methods. Hence, statistical simulation studies may provide guidance for minimum sample size required to minimise bias for different scenarios mentioned earlier.

Comparison of flagging rates using discrete versus continuous paediatric reference intervals

One justification for the use of continuous age-specific RIs is the improvement in clinical decision making, particularly in children with ages close to the age "cut-off" values used in discrete age-group RIs. As shown in **figure 1**, a girl aged 7 years with creatinine results of 55 µmol/L is within the discrete age-group RIs and outside the continuous RIs. In contrast, a girl aged 12 years with creatinine results of 65 µmol/L is outside the discrete age-group RIs and within the continuous RIs. Zierk *el al* compared the flagging rate of test results as lying outside the RIs in sick children and a reference population between discrete age-group RIs used in laboratories and published continuous age-specific RIs, and reported that the flagging rate of test results outside the RIs for continuous age-specific RIs in comparison to discrete age-group RIs reduced substantially for most blood count analytes and biochemical analytes, except for red cell distribution width, AST, ALT, potassium and lactate dehydrogenase (LDH).[14] However, the proportion of samples and patients considered pathologic based on the continuous RIs was more than 5% for all analytes.[14] Contrary to Zierk *et al*, the recent CALIPER study reported that the flagging rate of test results outside the continuous RIs was approximately 2.5% below or above the lower or upper limit respectively for most analytes.[38] The flagging rate did not reduce substantially when discrete age-group and continuous age-specific RIs using the same sample were compared[38]. However, the authors discussed that for continuous age-specific RIs the flagging rate of test results outside the RIs were consistent across age while discrete age-group RIs performed better near the centre of the age-group.[38] The difference in flagging rates between the study of Zierk *et al* and the CALIPER study is expected as one is an indirect study based on patient samples and the other is based on presumed healthy children.[14, 38] While the overall flagging rate can be made the same with both types of intervals, the continuous RI is more likely to correctly classify each patient and should be used as a gold standard to compare the discrete age-group RIs.

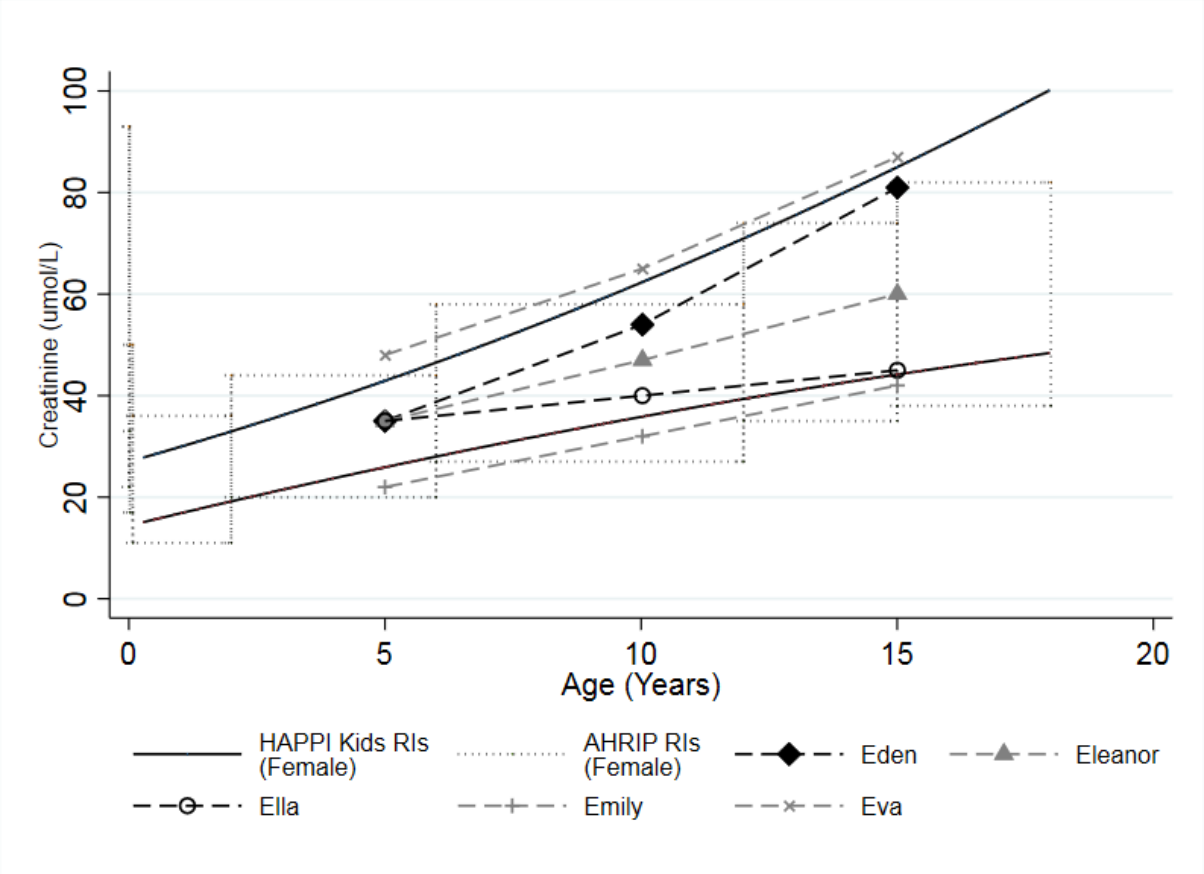


Figure 1: Visual comparison of trends in reference values over time between hypothetical individuals.

Explanation of Figure 1: Creatinine results (µmol/L) for hypothetical patients, Eden and Ella, are within the Harmonising Age Pathology Parameters in Kids (HAPPI Kids) Reference Intervals (RIs) and the Australasian Harmonised Reference Intervals for Paediatrics (AHRIP). However, longitudinal changes (e.g. increasing for Eden and decreasing for Ella) in their creatinine in comparison with the trend observed for the HAPPI Kids continuous RI may be of clinical concern and would allow the physician to better monitor their physiological condition**.**

Comparison of paediatric reference intervals between countries

Several countries have established age-specific continuous RIs and have harmonised RIs across laboratories or methods.[18, 19, 22] However, to date there has been no comparison of these RIs between countries. Considering CALIPER, Zierk *et al* and HAPPI Kids study have published RIs for common biochemistry analytes, a unique opportunity now exists to compare the documented paediatric RIs between countries using an indirect approach.[14, 19, 38] While this type of comparative study will be confounded by pre and post-analytical differences, and statistical methods applied, it may still assist in understanding ethnic and geographic differences in certain biomarkers after adjusting for confounding factors. The findings would significantly contribute to the establishment of global harmonisation guidelines and standards for paediatric RIs.

Translation of research findings into practice

Interpretation of pathology results obtained from multiple laboratories

Due to potential variation in pathology reference values derived from different analytical principles and/or analyser types (from the same or multiple laboratories), laboratories have a responsibility to report patient results with valid RIs for the method in use and the population they serve and also play an important role in the education of physicians for the interpretation of pathology test differences results obtained from multiple laboratories.[17, 19] At present, it remains extremely difficult for physicians to identify from the patient pathology report biomarkers with adopted harmonised or common RI's (e.g. results comparable between laboratories), versus laboratory specific RIs that cannot be compared. This issue is of particularly importance in health system models in which patients are diagnosed, managed and treated within both core and peripheral services and clinics. CALIPER has taken the initiative in addressing the assessment of paediatric laboratory test results for physicians,

healthcare worker, parents and patients worldwide by the development and launch of a mobile application. The CALIPER App is an easy and transportable tool for the assessment of the latest reference value database (CALIPER) based on healthy children and adolescents.[52] The development of mobile Apps and complexes LIS that enable the end user to access all RIs' relevant to the patient's assessment would enable better clinical use of our understanding of inter analysers differences.

Incorporating results in electronic databases

The inability of existing LIS to incorporate advanced mathematical functions or graphical representation of patient pathology data has been identified and remains a key challenge to incorporate continuous age-specific RIs. Overcoming these post-analytical reporting issues may enable easier visual recognition of biomarker trends over time and thus improved longitudinal follow up of paediatric patients. An example of a visual comparison of trends in reference values over time between hypothetical individuals is shown in **Figure 1**.

Beyond the LIS, the need for universal standards for patient data and RI data transmission becomes even more apparent when the wide variety of medical interfaces used for this purpose are considered. This extends to both larger electronic medical record systems and also desktop and mobile software applications used by medical practitioners to promptly view, interpret and action patient results. In these situations, further technological advancement is required not only to the LIS, but also to prepare the RIs, display them in a rendered format (e.g. PDF or directly on-screen) and also systematically transfer the results to receiving systems. All of these process require a capacity to accurately and safely render the results.

Multi-national LIS and/or EMR developers have an active role to play in the future development and delivery of end-user requested functionality to improve clinical care in paediatric patient cohorts. From a commercial perspective, these types of LIS and EMR advancements would be directly transferrable to a wide variety of different patient cohort sub-sets, where physiological or treatment-based changes in biomarkers are evident (e.g. obstetric patients, therapeutic drug monitoring regimes, oncology patients etc.).

Updating guidelines for newer statistical methods

As outlined, several studies recently published continuous age-specific RIs for the interpretation of pathological data in children.[13, 14, 19, 38] The statistical methods applied in estimation of continuous age-specific RIs are very different from those utilised in the derivation of discrete age-group RIs. Additionally, there is a variety of suitable methods for estimating continuous age-specific RIs, without any available consensus, guideline or standards from peak professional bodies (e.g. IFCC Reference Intervals and Decision Limits Committee (C-RIDL)) to assist medical professionals and biostatisticians in this task. Hence, further tasks that lie ahead for relevant professional bodies are the discussion of, consensus agreement on and guideline development of statistical best practice for the estimation of continuous RIs (or revision of guidelines for generation of discrete age-group RIs). It is likely that statistical simulation and comparative studies will be of methodological significance for guideline development.

Sustainability of paediatric reference intervals and clinical interpretation

Collaboration with diagnostic industry partners

International accreditation standards for medical laboratories (e.g. ISO15189) mandate individual laboratories as responsible for the periodic review of biological RIs and, where

applicable, the further investigation and corrective action of RIs that are: a) no longer appropriate for the reference population served or b) when pre-analytical or analytical laboratory procedures have changed (ISO 15189:2012).[4, 5] This requirement is aimed at maintaining the analytical and clinical interpretative quality of pathology results and acknowledges the relevance of changes in the ethnic distribution of local populations and advancements in automated analyser technologies, analytical methods, reagents and calibrators (as previously discussed). Depending on the origin of the RIs, the commutability of the RI values and the traceability of available metrological and pre-metrological information, the task of individual laboratories maintaining "accurate" RIs can range from onerous (e.g. production of reference values and estimation of the RI) to minimal (e.g. adoption of a published RI without validation). For paediatric patients, where reference values and published RIs have been historically sparse and laboratory testing is commonly performed by non-paediatric specialist laboratories, establishing or adopting paediatric RIs can be particularly difficult to attain. Even in laboratories with access to large numbers of paediatric specimens or major studies (e.g. CALIPER or HAPPI Kids), the cost and time associated with the continual review of and/or re-establishment of appropriate paediatric RIs for changed laboratory conditions, remain extremely challenging. The ethical and resource burden on sole laboratories to complete paediatric RI studies and publications may be improved if there was incentivised grants to facilitate further collaborative between vendors and end user groups, potentially coordinated by vendor agents. A further step would be if existing regulatory frameworks for IVDs were extended to specifically mandate IVD manufacturers for the provision of paediatric RIs for IVD products developed, marketed and sold as 'fit for purpose' for intended paediatric use. The ability of IVD manufacturers to meet such paediatric RI regulatory obligations would require continued collaborations and future development of collective commercial partnerships with major paediatric research teams.

Once established, the on-going verification of RIs can be achieved by the use of EQA to ensure analytical stability, with a focus on analyte concentrations found in paediatric samples, the application of an “average of normal” or alternative indirect techniques.

Test results presented as a percentile

The representation of an individual’s pathology test result outcome with appropriate RIs at discrete time-points (e.g. elevated biomarker at day x) can complicate clinical interpretation by masking of the trend in longitudinal test results that is not associated with intra-individual variation. This has been depicted in **Figure 1**, where 5, 10 and 15 year creatinine results for hypothetical patients, Eden and Ella, fell within the HAPPI Kids RIs and the Australasian Harmonised Reference Intervals for Paediatrics (AHRIP), but longitudinal trends in their respective creatinine levels (e.g. increasing for Eden and decreasing for Ella) in comparison with the trend observed for the HAPPI Kids continuous RI may be clinical concern. In contrast, Emily’s creatinine results of 30, 45 and 60 □mol/L (at 5, 10 and 15 years respectively) demonstrate her creatinine value remained stable on the 50th percentile. Future consideration should be given to the representation of paediatric pathology test results as percentiles, as commonly used for other child growth parameters (e.g. weight, height, and head circumference) and the potential advantages that this may have in assisting physicians in the clinical interpretation of individual data and thus monitoring of physiological conditions**.**

Conclusion:

Paediatric RIs appropriate for the population being served by the servicing laboratory and the pre-analytical, analytical and post-analytical procedures utilised in generation of test results are essential for improving clinical interpretation and subsequent clinical decisions. Despite the past and present challenges of establishing paediatric reference values and estimating

paediatric RIs, there have been new and significant developments published that offer exciting improvements to pathology result interpretation, patient safety and ease of clinical management. However, with rapid advancement in technologies, both testing processes of the laboratory and the ways of reporting patient results are changing quickly. Hence, long-term planning to direct future research, software development and best practice clinical and statistical guidelines for paediatric RI development, implementation, review and accreditation will be critical for sustainability of quality and acceptance into routine laboratory practise.

Reference:

1. Rohr U-P, Binder C, Dieterle T, Giusti F, Messina CGM, Toerien E, *et al*. The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. PloS one. 2016;11(3):e0149856-e.
2. The Royal College of Pathologists of Australasia. Pathology The Facts. RCPA; 2019.
3. Royston P. Constructing time-specific reference ranges. Statistics in Medicine. 1991;10(5):675-90.
4. Fuentes-Arderiu X. Biological reference intervals and ISO 15189. Clinica chimica acta; international journal of clinical chemistry. 2006;364(1-2):365-6.
5. ISO. 15189:2012, I. (2017). ISO 15189:2012 - Medical laboratories -- Requirements for quality and competence. . 2012.
6. CLSI. CLSI Releases Guidelines for Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory. 2008.
7. Ridefelt P, Hellberg D, Aldrimer M, Gustafsson J. Estimating reliable paediatric reference intervals in clinical chemistry and haematology. Acta paediatrica (Oslo, Norway : 1992). 2014;103(1):10-5.

8. Coffin CM, Hamilton MS, Pysher TJ, Bach P, Ashwood E, Schweiger J, *et al*. Pediatric laboratory medicine: current challenges and future opportunities. American Journal Of Clinical Pathology. 2002;117(5):683-90.

9. Ozarda Y. Reference intervals: current status, recent developments and future considerations. Biochemia medica. 2016;26(1):5-16.

10. Forsman R. Statistical Bases of Reference Values in Laboratory Medicine. Eugene K. Harris and James C. Boyd. New York; Marcel Dekker, Inc., 1995. Hardback $165.00. ISBN 0-8247-9339-0. Clinical chemistry. 1996;42(8):1303.

11. Katayev A, Balciza C, Seccombe DW. Establishing Reference Intervals for Clinical Laboratory Test ResultsIs There a Better Way? American Journal of Clinical Pathology. 2010;133(2):180-6.

12. Howie S. Blood sample volumes in child health research: review of safe limits. Bulletin of the World Health Organization. 2011;89.

13. Loh TP, Antoniou G, Baghurst P, Metz MP. Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. Pathology. 2014;46(4):336-43.

14. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, *et al*. Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clinical chemistry. 2015;61(7):964-73.

15. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, *et al*. Indirect methods for reference interval determination - review and recommendations. Clin Chem Lab Med. 2018;57(1):20-9.

16. Farrell C-JL, Nguyen L. Indirect Reference Intervals: Harnessing the Power of Stored Laboratory Data. The Clinical biochemist Reviews. 2019;40(2):99-111.

17. Adeli K, Higgins V, Trajcevski K, Palmert MR. Important considerations for interpreting biochemical tests in children. BMJ. 2018;361:k1950.

18. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. Critical reviews in clinical laboratory sciences. 2017;54(6):358-413.

19. Hoq M, Matthews S, Karlaftis V, Burgess J, Cowley J, Donath S, *et al*. Reference Values for 30 Common Biochemistry Analytes Across 5 Different Analyzers in Neonates and Children 30 Days to 18 Years of Age. Clinical chemistry. 2019:clinchem.2019.306431.

20. Delanghe JR, Cobbaert C, Galteau MM, Harmoinen A, Jansen R, Kruse R, *et al*. Trueness verification of actual creatinine assays in the European market demonstrates a disappointing variability that needs substantial improvement. An international study in the framework of the EC4 creatinine standardization working group. Clin Chem Lab Med. 2008;46(9):1319-25.

21. Steinmetz J, Schiele F, Gueguen R, Ferard G, Henny J. Standardization of gamma-glutamyltransferase assays by intermethod calibration. Effect on determining common reference limits. Clin Chem Lab Med. 2007;45(10):1373-80.

22. Tate JR, Sikaris KA, Jones GRD, Yen T, Koerbin G, Ryan J, *et al*. Harmonising Adult and Paediatric Reference Intervals in Australia and New Zealand: An Evidence-Based Approach for Establishing a First Panel of Chemistry Analytes. The Clinical Biochemist Reviews. 2014;35(4):213-35.

23. Pathology Harmony Group. Harmonisation of Reference Intervals: Clinical Biochemistry Outcomes; 2011 [Available from: http://www.acb.org.uk/docs/default-source/committees/scientific/guidelines/acb/pathology-harmony-ii-biochemistry.pdf?sfvrsn=2.

24. International Federation of Clinical Chemistry and Laboraotry Medicine. Paediatric Reference Range Initiatives 2019 [Available from: https://www.ifcc.org/c-etplm/paediatric-reference-range-initiatives/.

25. Kohse KP. KiGGS — The German survey on children's health as data base for reference intervals and beyond. Clinical Biochemistry. 2014;47(9):742-3.

26. Flanders MM, Crist RA, Roberts WL, Rodgers GM. Pediatric Reference Intervals for Seven Common Coagulation Assays. Clinical chemistry. 2005;51(9):1738-42.

27. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Martensson A, *et al*. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scandinavian journal of clinical and laboratory investigation. 2004;64(4):271-84.

28. Colantonio DA, Kyriakopoulou L, Chan MK, Daly CH, Brinc D, Venner AA, *et al*. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clinical chemistry. 2012;58(5):854-68.

29. Daly CH, Liu X, Grey VL, Hamid JS. A systematic review of statistical methods used in constructing pediatric reference intervals. Clin Biochem. 2013;46(13-14):1220-7.

30. Daly CH, Higgins V, Adeli K, Grey VL, Hamid JS. Reference interval estimation: Methodological comparison using extensive simulations and empirical data. Clin Biochem. 2017;50(18):1145-58.

31. Estey MP, Cohen AH, Colantonio DA, Chan MK, Marvasti TB, Randell E, *et al*. CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens Clinical Chemistry Assays: Direct validation using reference samples from the CALIPER cohort. Clinical Biochemistry. 2013;46(13–14):1197-219.

32. Southcott EK, Kerrigan JL, Potter JM, Telford RD, Waring P, Reynolds GJ, *et al*. Establishment of pediatric reference intervals on a large cohort of healthy children. Clinica Chimica Acta. 2010;411(19–20):1421-7.

33. Hoq M, Karlaftis V, Matthews S, Burgess J, Donath S, Carlin J, *et al*. A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the settings of clinical biochemistry, immunology and haematology: The HAPPI Kids study. BMJ Open. 2019.

34. Hoq M, Monagle P, Donath S, Carlin J, editors. Comparison of statistical methods for constructing age-specific paediatric reference intervals: A simulation study. Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference; 2019; Adelaide SA: Clinical Biochemist Review Supplement; 2019.

35. Hoq M, Canterford L, Khanom G, Monagle P, Donath S, Carlin J, editors. Statistical methods applied in the estimation of age-specific paediatric reference intervals for a laboraotry blood test: A systematic review. Australasian Association of Clinical Biochemists' 57th Annual Scientific Conference; 2019; Adelaide, SA: Clinical Biochemist Reviews Supplement; 2019.

36. Ni X, Song W, Peng X, Shen Y, Peng Y, Li Q, *et al*. Pediatric reference intervals in China (PRINCE): design and rationale for a large, multicenter collaborative cross-sectional study. Science Bulletin. 2018;63(24):1626-34.

37. Bussler S, Vogel M, Pietzner D, Harms K, Buzek T, Penke M, *et al*. New pediatric percentiles of liver enzyme serum levels (ALT, AST, GGT): Effects of age, sex, BMI and pubertal stage. Hepatology (Baltimore, Md). 2017.

38. Asgari S, Higgins V, McCudden C, Adeli K. Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals. Clin Biochem. 2019;73:82-9.

39. Araújo PAT, Thomas D, Sadeghieh T, Bevilacqua V, Chan MK, Chen Y, *et al*. CLSI-based transference of the CALIPER database of pediatric reference intervals to Beckman Coulter DxC biochemical assays. Clinical Biochemistry. 2015;48(13):870-80.

40. Ridefelt P, Hilsted L, Juul A, Hellberg D, Rustad P. Pediatric reference intervals for general clinical chemistry components - merging of studies from Denmark and Sweden. Scandinavian journal of clinical and laboratory investigation. 2018;78(5):365-72.

41. Bailey D, Bevilacqua V, Colantonio DA, Pasic MD, Perumal N, Chan MK, *et al*. Pediatric within-day biological variation and quality specifications for 38 biochemical markers in the CALIPER cohort. Clinical chemistry. 2014;60(3):518-29.

42. Loh TP, Metz MP. Indirect estimation of pediatric between-individual biological variation data for 22 common serum biochemistries. Am J Clin Pathol. 2015;143(5):683-93.

43. Loh TP, Ranieri E, Metz MP. Derivation of pediatric within-individual biological variation by indirect sampling method: an LMS approach. Am J Clin Pathol. 2014;142(5):657-63.

44. Hoq M, Matthews S, Karlaftis V, Burgess J, Donath S, Carlin J, *et al*. Validation of the HAPPI Kids continuous age-specific paediatric reference intervals. Journal of Applied Laboratory Medicine. 2020.

45. Higgins V, Adeli K. Advances in pediatric reference intervals: from discrete to continuous. Journal of Laboratory and Precision Medicine. 2018;3(1).

46. Doherty T, Salik I. Physiology, Neonatal. Treasure Island (FL): StatPearls Publishing; 2019 2019 Mar 12.

47. Cole TJ. The LMS method for constructing normalized growth standards. European journal of clinical nutrition. 1990;44(1):45-60.

48. Gannoun A, Girard S, Guinot C, Saracco J. Reference curves based on non-parametric quantile regression. Stat Med. 2002;21(20):3119-35.

49. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. 2007. 2007;23(7):46.

50. Wright EM, Royston P. A Comparison of Statistical Methods for Age-related Reference Intervals. Journal of the Royal Statistical Society: Series A (Statistics in Society). 1997;160(1):47-69.

51. Bellera CA, Hanley JA. Original Article: A method is presented to plan the required sample size when estimating regression-based reference limits. Journal of Clinical Epidemiology. 2007;60:610-5.

52. CALIPER. CALIPER Pediatric Reference Interval Database The Research Institute at The Hospital for Sick Children [Available from: https://caliper.research.sickkids.ca/#/.