

Comparison of Statistical Methods for Estimating Continuous Paediatric Reference Intervals: A Simulation Study

Monsurul Hoq (✉ monsurul.hoq@mcri.edu.au)

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, 48 Flemington Road, Parkville, Victoria - 3052, Australia

Susan Donath

Murdoch Children's Research Institute

Paul Monagle

Royal Children's Hospital

John Carlin

Murdoch Children's Research Institute

Research Article

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

Posted Date: January 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-144805/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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}

¹Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia.

²Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, Victoria, Australia

³The Royal Children's Hospital, Parkville, Victoria, 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 $\alpha/2^{\text{th}}$ and $(100 - \alpha/2)^{\text{th}}$ 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 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. These

parameters are then used to transform the data towards normality and centiles are calculated [10]. 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. 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 [18] study, 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 ($n = 100, 200, 400, \text{ and } 1000$) 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].

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 $\alpha = 2.5$ 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**.

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, similarly to the approach proposed by Cole and Green [9, 24]. 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), $-2l(\hat{\theta}) + (\xi \cdot df)$, where $l(\hat{\theta})$ 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. $\frac{\sum_{k=0}^1 \sum_{j=1}^{18} \sqrt{E(\hat{\theta}_{jk}^{(\alpha)} - \theta_{jk}^{(\alpha)})^2}}{(18*2)}$ 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 $E(\hat{\theta}_{jk}^{(\alpha)} - \theta_{jk}^{(\alpha)})^2$ 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], `mfp_i`[28] and `qreg` were used for Hoq *et al*'s method, and `fp`, `lrtest`,

`xrignls`[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 code for all methods are provided in the **supplementary document 1 A-D**.

Results

Results from the simulation 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**. 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**. 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 **Supplementary Table 1 and 2** 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.

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 [18] 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.

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 supported 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. Morris TP, White IR, Crowther MJ: Using simulation studies to evaluate statistical methods. *Statistics in Medicine* 2019, 38(11):2074-2102.
19. 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.
20. Cole TJ: The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990, 44(1):45-60.
21. 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.

22. Royston P, Ambler G: Multivariable fractional polynomials. *Stata Technical Bulletin* 1998, 43:24-32.
23. Hastie T, Tibshirani R: Generalized additive models for medical research. *Stat Methods Med Res* 1995, 4(3):187-196.
24. Rigby RA, Stasinopoulos DM: Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. *Statistical Modelling* 2006, 6(3):209-229.
25. Stasinopoulos M, Rigby B, Akantziliotou C: Instructions on how to use the gamlss package in R Second Edition. In.; 2008.
26. StataCorp: *Stata Statistical Software : Release 15*. In. College Station, TX: StataCorp LP; 2017.
27. Royston P, Sauerbrei W: Two techniques for investigating interactions between treatment and continuous covariates in clinical trials. *Stata Journal* 2009, 9(2):230-251.
28. 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.
29. 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.
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).

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	$54.0 + 0.0713 \times age^3 - 0.0268 \times age^3 \times \ln(age)$
		$209.064 + 0.301 \times age^3 - 0.112 \times age^3 \times \ln(age)$ 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

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 <i>et al</i>	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 <i>et al</i>	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 <i>et al</i>	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 <i>et al</i>	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 <i>et al</i>	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

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

Figure titles

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

Figure 2: Root Mean Square Error by age, sex, sample size and method for scenario 1

Figure 3: Root Mean Square Error by age, sex, sample size and method for scenario 2

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

Figure 5: RIs for creatinine from the HAPPI Kids study applying four methods for two sample sizes

Supplementary table 1 and 2

Table S-1: Average coverage across integer age and sex for Hoq *et al* and Royston's methods, for five scenarios and four sample sizes

Table S-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

Supplementary Figures 1 to 5: Average estimated and true RIs by age, sex and method for five scenarios

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

Supplementary Figures 6 – 8: Root mean square error by age, sex and method for three scenarios

Figure S-6: Root Mean Square Error by age, sex, sample size and method for scenario 3

Figure S-7: Root Mean Square Error by age, sex, sample size and method for scenario 4

Figure S-8: Root Mean Square Error by age, sex, sample size and method for scenario 5

Supplementary document 1: Stata and R codes of all four statistical methods applied in estimating age-specific paediatric RIs

1 A: Stata code for Cole's LMS method

1 B: Stata code for Royston's method

1 C: Stata code for Hoq *et al's* method

1 D: R code for GAMLSS

Figures

FIGURE 1: Distribution of measurement across age for five biomarkers from the HAPPI Kids Study

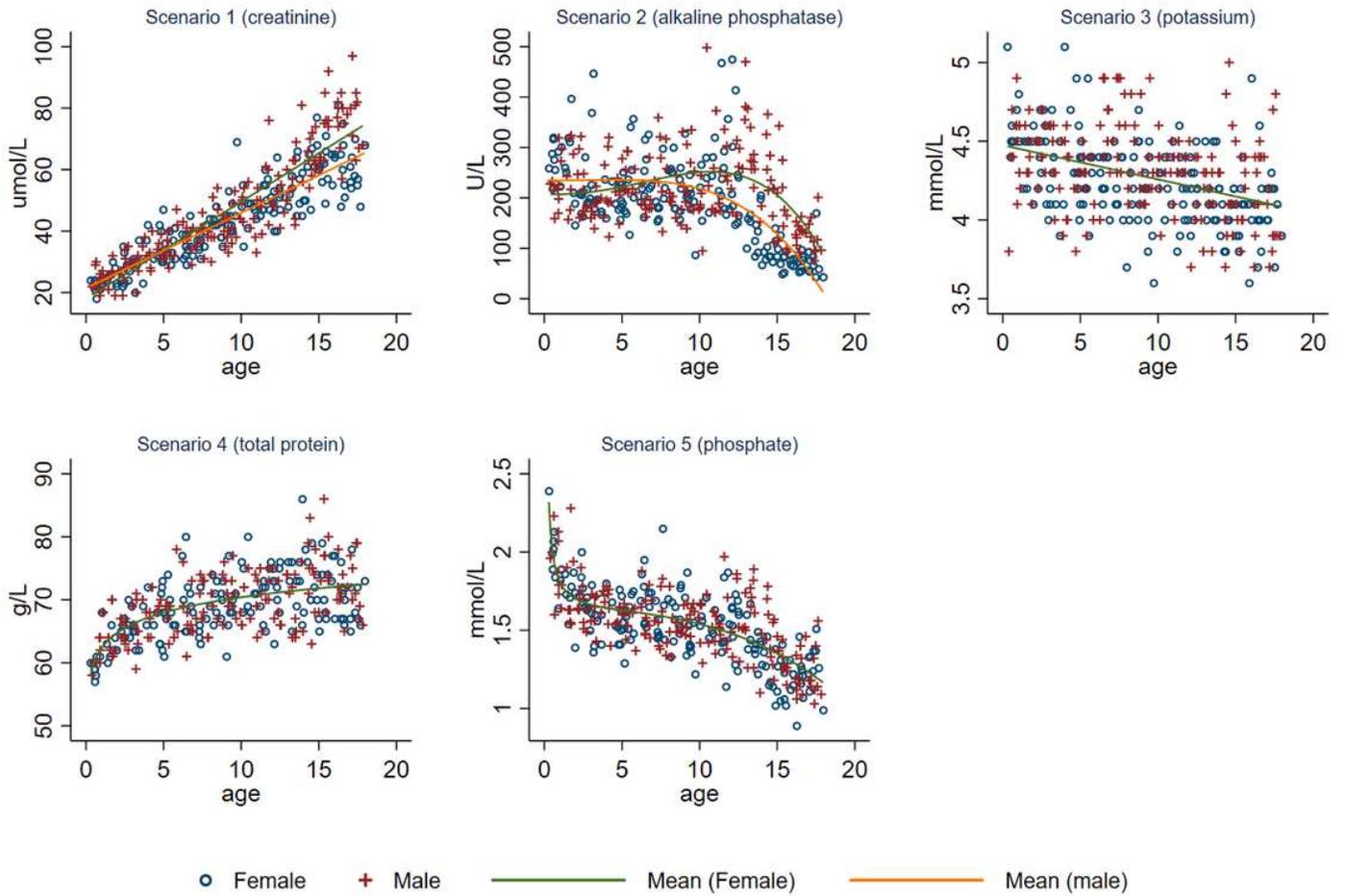
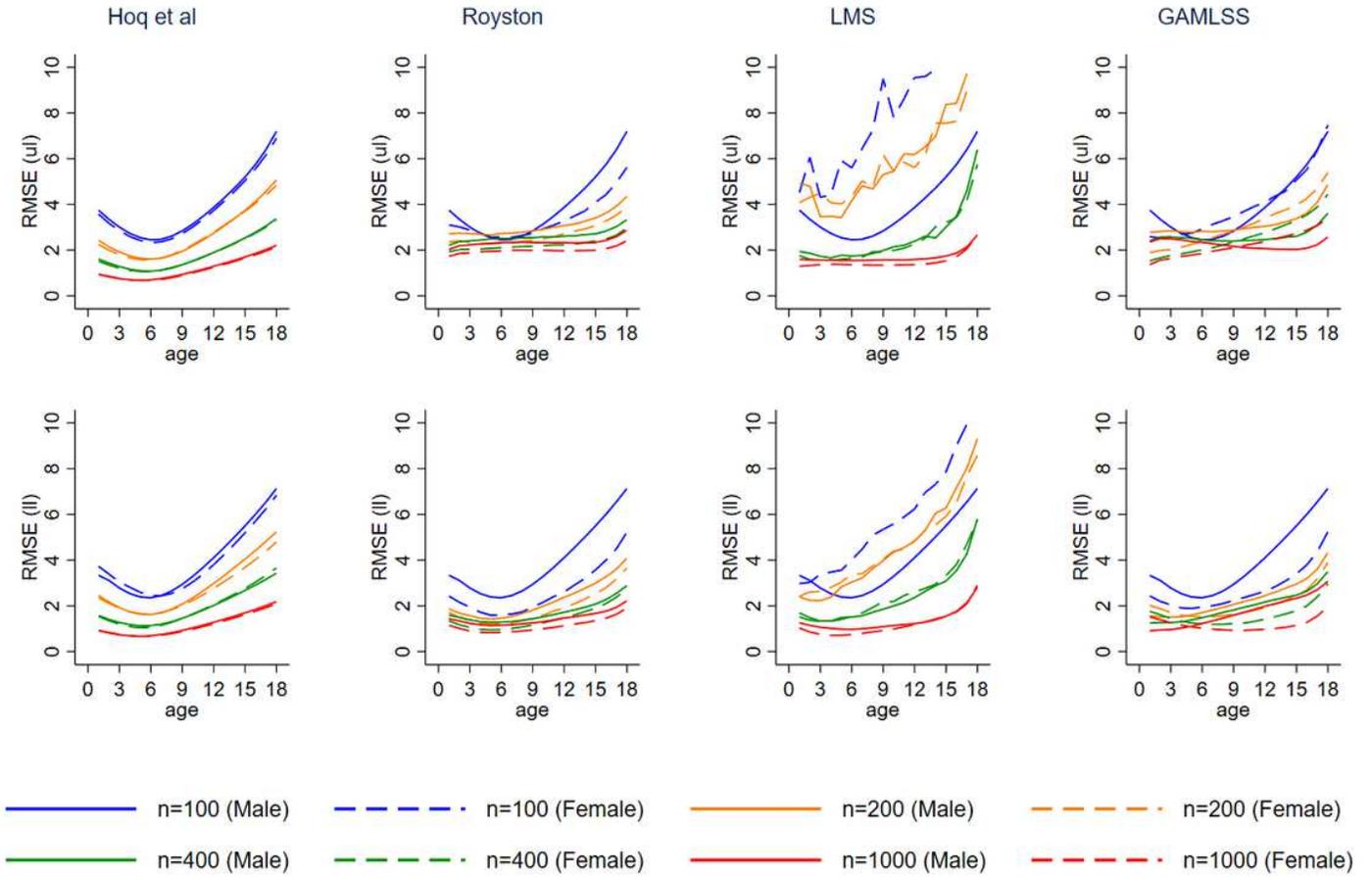


Figure 1

Distribution of measurement across age for five biomarkers from the HAPPI Kids Study

FIGURE 2: Root Mean Square Error by age, sex, sample size and method for scenario 1

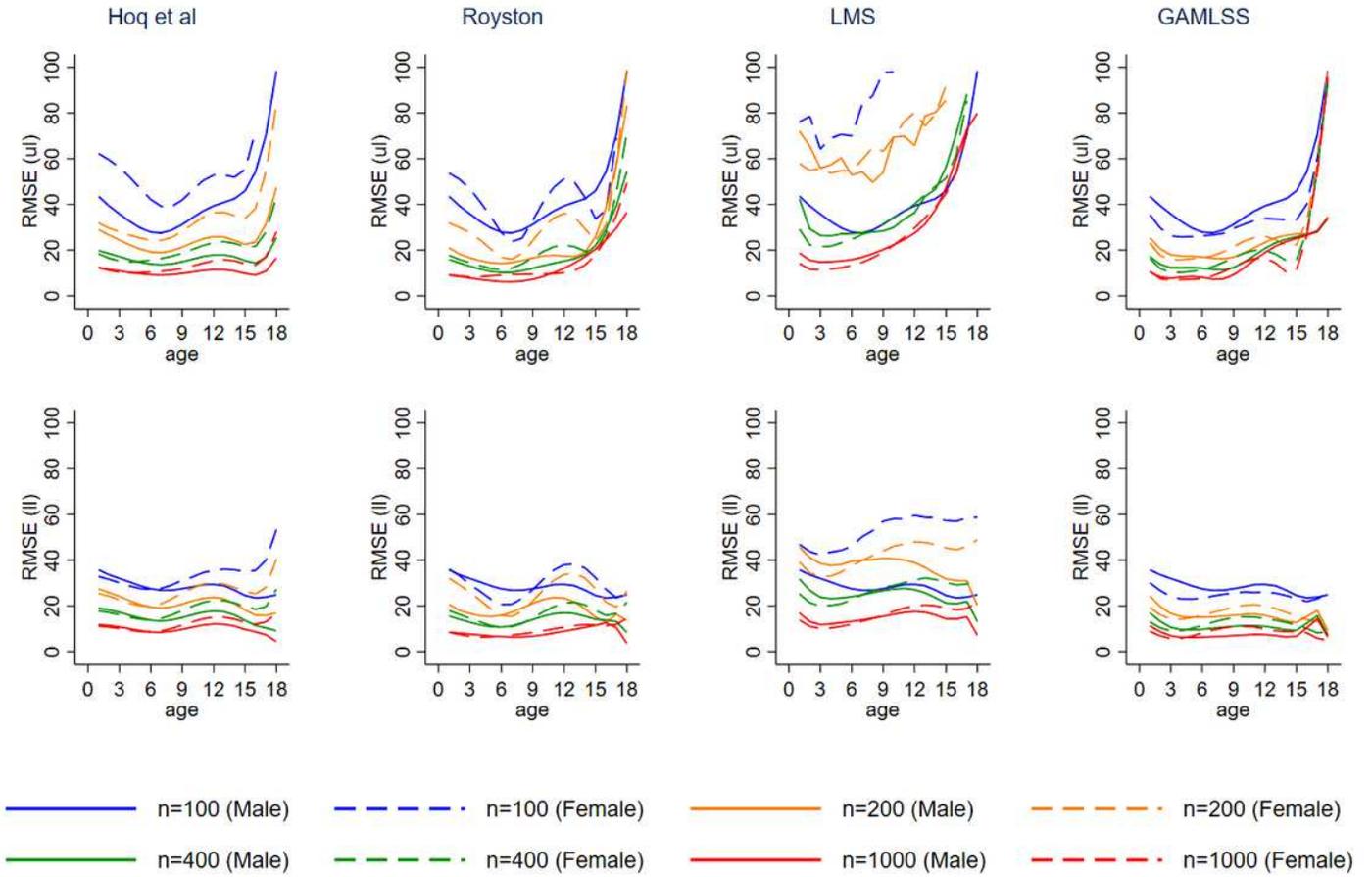


Values of RMSE > 10 not shown

Figure 2

Root Mean Square Error by age, sex, sample size and method for scenario 1

FIGURE 3: Root Mean Square Error by age, sex, sample size and method for scenario 2



Values of RMSE > 100 not shown

Figure 3

Root Mean Square Error by age, sex, sample size and method for scenario 2

FIGURE 4: RIs for creatinine from the HAPPI Kids study applying four methods for two sample sizes

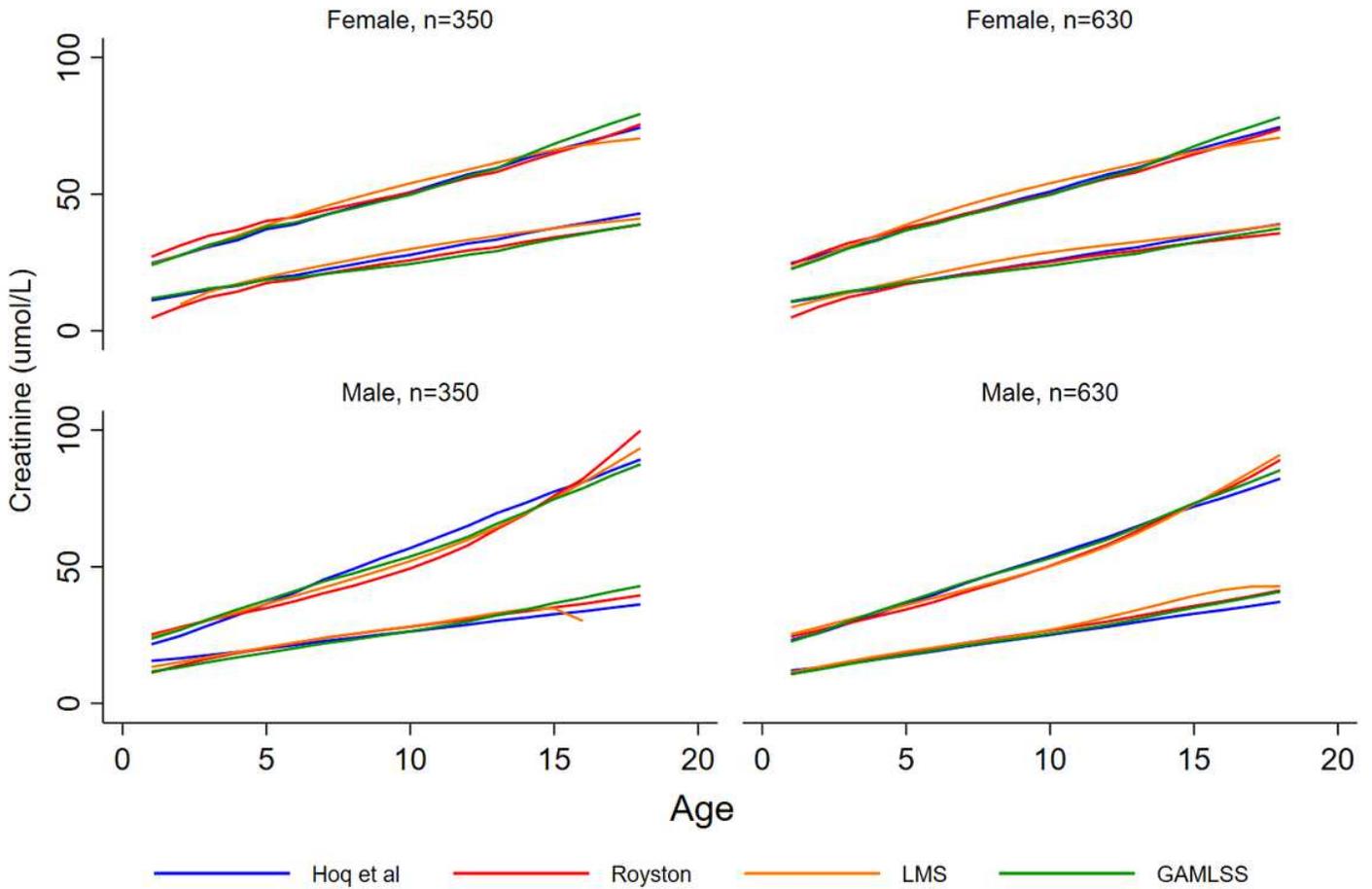


Figure 4

RIs for creatinine from the HAPPI Kids study applying four methods for two sample sizes

FIGURE 5: RIs for alkaline phosphatase from HAPPI Kids study applying four methods for two sample sizes

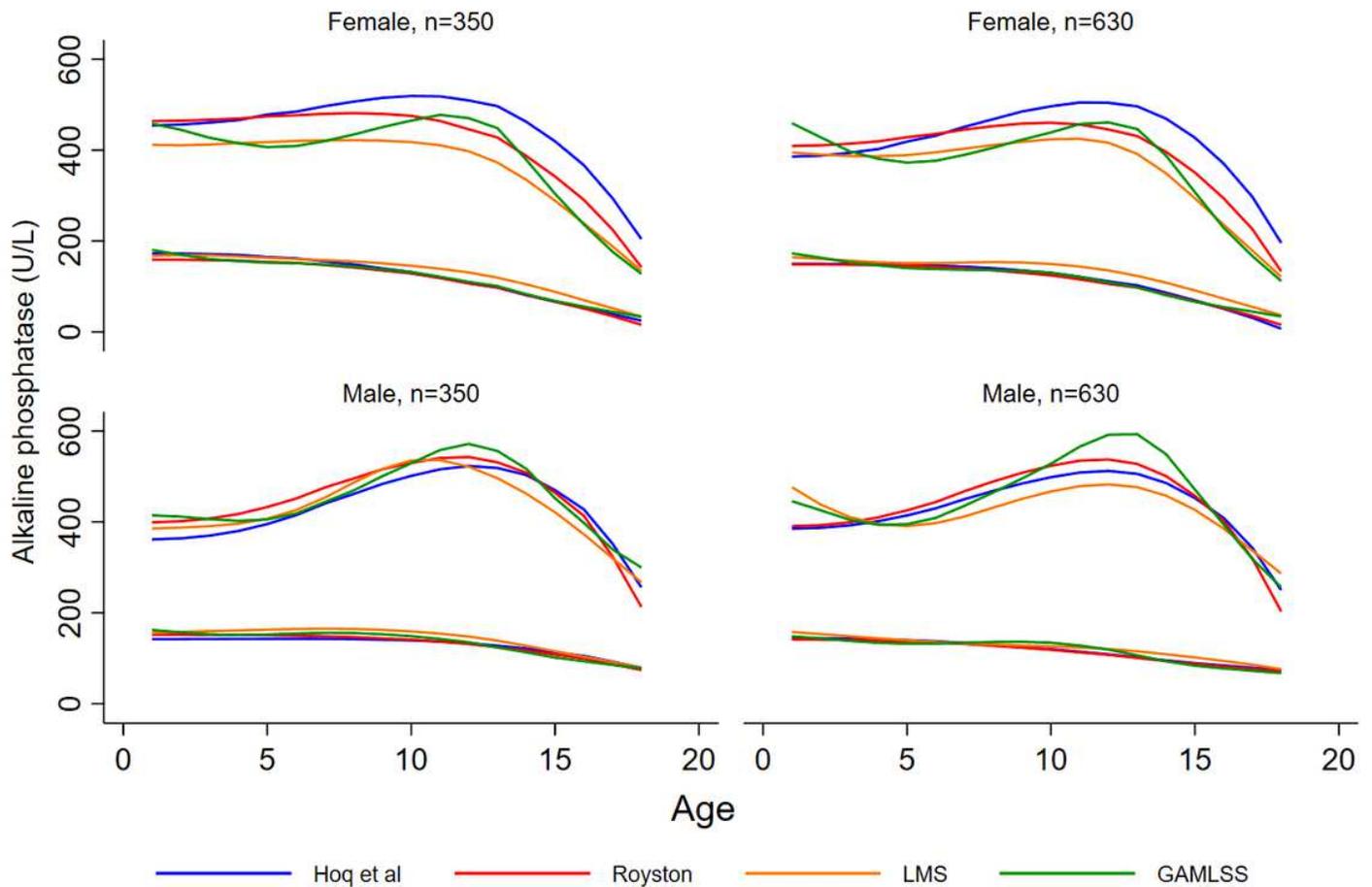


Figure 5

RIs for creatinine from the HAPPI Kids study applying four methods for two sample sizes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalDocument1StataandRCodesHoqetalcomparisonofstatisticalmethodforestimatingRIs.docx](#)
- [SupplementaryFigures1to8HoqetalcomparisonofstatisticalmethodsforestimatingRIs.docx](#)
- [SupplementaryTable1and2HoqetalcomparisonofstatisticalmethodsforestimatingRIs.docx](#)