From Population to Subject-Specific Reference Intervals

Murih Pusparum[1,2][0000−0001−9560−0612], Gökhan
Ertaylan[2][0000−0001−5602−6435], and Olivier Thas[1,3,4][0000−0001−6442−4089]

[1] Hasselt University, Hasselt 3500, Belgium
[2] Flemish Institute for Technological Research (VITO), Mol 2400, Belgium
[3] Department of Data Analysis and Mathematical Modelling, Ghent University,
Ghent 9000, Belgium
[4] National Institute for Applied Statistics Research Australia (NIASRA),
Wollongong 2500, NSW, Australia

Abstract. In clinical practice, normal values or reference intervals are the main point of reference for interpreting a wide array of measurements, including biochemical laboratory tests, anthropometrical measurements, physiological or physical ability tests. They are historically defined to separate a healthy population from unhealthy and therefore serve a diagnostic purpose. Numerous cross-sectional studies use various classical parametric and nonparametric approaches to calculate reference intervals. Based on a large cross-sectional study (N = 60,799), we compute reference intervals for subpopulations (e.g. males and females) which illustrate that subpopulations may have their own specific and more narrow reference intervals. We further argue that each healthy subject may actually have its own reference interval (subject-specific reference intervals or SSRIs). However, for estimating such SSRIs longitudinal data are required, for which the traditional reference interval estimating methods cannot be used. In this study, a linear quantile mixed model (LQMM) is proposed for estimating SSRIs from longitudinal data. The SSRIs can help clinicians to give a more accurate diagnosis as they provide an interval for each individual patient. We conclude that it is worthwhile to develop a dedicated methodology to bring the idea of subject-specific reference intervals to the preventive healthcare landscape.

Keywords: Clinical statistics· Clinical biochemistry· Reference intervals · Longitudinal data · Quantile mixed models.

1 Introduction

In the era of personalized medicine, we are surrounded by sensors and methodologies to capture and store data from a single individual at an unprecedented scale. These data later can be associated with Linked Data technologies [1]. Furthermore, artificial intelligence, and more specifically various machine learning (ML) techniques, are making their way to daily practice. We are presented with

algorithms that can outperform pathologists [2] for diagnosing and staging of diseases [2, 3]. However, one disadvantage of AI and ML approaches is the lack of transparency in the decision making process. Although there is progress to make these algorithms transparent, these efforts are in their infancy. Hence, statistical methods and reasoning are essential to bridge the gap towards their implementation in clinical practice. Specifically, in the preventive healthcare landscape, there is a clear lack of methodological development to render these technologies viable in practice.

In the field of medicine, interpreting clinical laboratory results can be done in several ways. The most common way is by comparing them to a standard value or range that has been calculated from a reference population of healthy individuals. Such intervals are known as normal values or reference intervals, but for further purposes we will refer to them as *population reference intervals* (PRI). In clinical practise, a patient will be considered healthy when the laboratory results show values within this PRI. In the preventive healthcare landscape, however, a PRI only gives little advantage since it is only designed for diagnostic purposes. Moreover, it is assumed to be constant over time and space.

To obtain the PRI, a cross-sectional prospective or retrospective study is typically considered. In this type of studies, the data of a particular physiological or clinical parameters will be collected from a large number of healthy subjects. The participants must be as similar as possible with the target population in which the PRIs will be used. For example, a study for estimating PRIs of BNP (brain natriuretic peptide) using only university students will be inappropriate as this BNP test is normally run for elderly people [4].

The classical definition of a PRI is the central 95% of the reference population of the parameter of interest. This central 95% is located between the 2.5 and 97.5 percentiles of the reference population. Various methods for estimating the PRIs have been proposed. Parametric methods start from the assumption that the distribution of the parameter of interest can be described by a particular distribution (usually Gaussian). The percentiles can then be directly computed from this distribution when its parameters (mean and variance for the Gaussian distribution) are estimated from a dataset. In general, when the distributional assumption holds, the parametric methods will be better in the sense they provide more precise estimates of the PRI than the nonparametric methods for the same sample size. However, without any distributional assumption, the nonparametric methods are more suitable as they can still produce unbiased estimators of the PRI, whereas the parametric methods may give biased results when the wrong distribution is used. With nonparametric methods, a minimum number of 120 participants is proposed for calculating the intervals [4]. Statistically speaking, larger sample sizes will result in better estimates in terms of bias and precision.

These classical estimation methods typically require a cross-sectional dataset, containing a single measurement of the parameter for each subject in the study. The application of these methods can be seen in many studies using a large number of cross-sectional samples for estimating PRIs for common clinical markers [5, 6, 8, 7, 9]. Longitudinal studies, on the other hand, are characterised by multi-

ple (repeated) measurements of the parameter for each subject in the study. Several studies involving longitudinal dataset for calculating PRIs have been performed [10, 11]. However, instead of using the classical methods, a simple random effects model and a semi-parametric method were used in these two studies, respectively, to produce pointwise PRIs i.e. PRIs that only have a valid probabilistic interpretation for each time point separately.

In this paper, we demonstrate the use of classical methods for reference interval calculation on a large cross-sectional study. We also apply the methods to subpopulations (e.g. males and females, for illustration purpose) and we will argue that reference intervals can be made more specific (i.e. more informative) when applied to such subpopulations. By extending this reasoning to every individual, we end up with reference intervals for each subject (subject-specific reference intervals, SSRI). The assumption under SSRI is that each parameter measured from any subject has a biological variation that is specific to this individual and has potential upper and lower boundaries that can be inferred from data, allowing a better interpretation of this parameter range taking into account subject specific variation. For the estimation of these SSRI, data on single subject level are required, and hence data from longitudinal studies are needed. We will propose to estimate subject-specific reference intervals with linear quantile mixed models (LQMM). Data and methods are described in Section 2, while the results are discussed in Section 3. A conclusion and some suggestions for future research will be given in Section 4.

2   Material and Methods

2.1   Data Description

There are two different types of datasets that are used in this paper. The first dataset comes from a cross-sectional study conducted in 2012-2016 and consisting of 60,799 participants from the Balearic Island, Spain, with ages ranging from 19 to 70 years [12]. This dataset will be referred to as the *Balearic data*. The measurements fall into three categories: a personal and health habits category (e.g. gender, age and smoking status), an anthropometric or a physiological measurements category (e.g. BMI and body fat percentage), and a clinical category (e.g. HDL and LDL cholesterol level). To reduce the scope of the research, only 5 parameters from each of the physiological and the clinical category were considered. Physiological parameters consist of a body shape index (ABSI), body mass index (BMI), waist circumference, systolic blood pressure, and diastolic blood pressure. The clinical parameters include total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, and glucose level.

The second dataset comes from our in-house ongoing longitudinal cohort study of 30 individuals with monthly physiological and clinical measurements over a period of 9 months and with age ranging from 45 to 60 years at the time of recruitment. This dataset will be referred to as the *IAM Frontier data* [13]. The same 10 physiological and clinical measurements as for the Balearic data were assessed. Characteristics of the two datasets are presented in Table 1. Due

to privacy reasons and confidentiality, the order of the individuals presented in each graph in Section 3 is randomised.

Table 1. Data characteristics of the Balearic and the IAM Frontier datasets.

	Balearic data	I AM Frontier data
Type	Cross-sectional	Longitudinal
Number of subjects	60,799	30
Time points	Single (1 time point)	Multiple (9 time points)

Category	Parameter
Physiological	A body shape index (ABSI)
	Body mass index (BMI)
	Waist circumference
	Systolic blood pressure (Systolic BP)
	Diastolic blood pressure (Diastolic BP)
Clinical	Total cholesterol
	HDL cholesterol
	LDL cholesterol
	Triglycerides
	Glucose

2.2   Classical Parametric and Nonparametric Reference Intervals

Let n denote the total number of sample observations. For the Balearic data, the RIs were estimated using a classical parametric method and two nonparametric methods. The parametric method estimates the 2.5 and 97.5 percentiles as

$$\bar{x} \pm z_{0.975} s_x \tag{1}$$

where $\bar{x}$ and s_x indicate the sample mean and the sample standard deviation, and $z_{0.975}$ is the 97.5 percentile of a standard normal distribution [14].

With the nonparametric methods the bounds of the reference interval are computed as the sample 2.5 percentile and the sample 97.5 percentile. These are estimated from the order statistics, which is the ordered set of sample observations. In particular, for a sample if n observations, the order statistics can be denoted by $y_{[1]} \leq y_{[2]} \leq \cdots \leq y_{[n]}$. We consider two nonparametric methods. The first estimates the 2.5 and the 97.5 percentile as the $0.025(n + 1)$-th and $0.975(n + 1)$-th order statistics. The second method estimates these percentiles as the $[(0.025 \times n) + 0.5]$-th and $[(0.975 \times n) + 0.5]$-th order statistics. If any of these numbers is not an integer then it is rounded to the nearest value, for example a value of 12.3 is rounded to 12 and 12.6 is rounded to 13. For rounding

off a .5 decimal, it follows the 'round-to-even' rule, therefore 12.5 equals 12 and 13.5 equals 14. These two nonparametric methods will be referred to as NP1 and NP2. The bootstrap or resampling technique was also applied in combination with these methods [15]. We will call the PRIs obtained from these five approaches the *classical reference intervals* (CRIs) and the summary is presented in Table 2.

For some clinical parameters, one-sided reference intervals are needed. For example, for LDL cholesterol only an upper bound is used in clinical practice. In such cases, the PRIs still refer to 95% of the reference population, but now the lower bound is fixed at the minimal value of 0, and the upper bound is given by the 95 percentile of the distribution. The methods described in the previous paragraphs can still be used, but with the 97.5 percentile replaced with the 95 percentile.

Table 2. Summary of classical parametric and nonparametric methods

Method	Formula	Characteristics
Parametric	$\bar{x} \pm z_{0.975}s_x$	Rely on distributional assumptions
Nonparametric		Without distributional assumptions,
Nonparametric 1 (NP1)	[0.025(n+1), 0.975(n+1)]	based on order statistics
Bootstrapped NP1	[0.025(n+1), 0.975(n+1)] in combination with bootstrapped samples	
Nonparametric 2 (NP2)	[0.025n + 0.5, 0.975n + 0.5)]	
Bootstrapped NP1	[0.025n + 0.5, 0.975n + 0.5)] in combination with bootstrapped samples	

2.3   Linear quantile mixed models Longitudinal Data

For the IAM Frontier data, linear quantile mixed models (LQMM) were fitted to obtain the RIs estimates. Linear quantile regression models [16] are a class of statistical models that express a particular quantile or percentile (e.g. quantile $\tau \in (0,1)$) of the outcome distribution as a linear function of one or more regressors. In our setting, we do not have regressors, but we do have repeated measurements on multiple subjects. This can be formulated as a simple special case of a linear quantile mixed model (LQMM), which extend the class of linear quantile regression models by the inclusion of random effects. In particular, we propose a LQMM which only includes one fixed-intercept and one random-intercept model the between-subject variability of the reference intervals. With Y the outcome variable (i.e. clinical parameter of interest) of subject $i = 1, \ldots, n$, with random effect u_i and with $Q(\tau \mid u_i)$ the subject-specific quantile function of outcome Y evaluated in the $100 \times \tau$ percentile, the model can be written as

$$Q(\tau \mid u_i) = \beta_0^{(\tau)} + u_i, \tag{2}$$

in which $\beta_0^{(\tau)}$ represents the fixed intercept. The model is completed by specifying the distribution of the random effects; in this paper this is restricted to the zero-mean normal distribution with variance Ψ_u^2. Note that the intercept parameter

$\beta_0^{(\tau)}$ has the interpretation of the $100 \times \tau$ percentile bound of the PRI, whereas $\beta_0^{(\tau)} + u_i$ has the interpretation of the SSRI for subject i. We need this model with $\tau = 0.025$ (lower bound) and with $\tau = 0.975$ (upper bound).

This class of models were first described in a study in 2007 and the authors gave details on how the model parameters can be estimated from longitudinal data [17]. They also proposed a method for predicting the subject-specific random effects u_i. Their models and methods were further generalised and improved in [18]. The methods are implemented in the *lqmm* package [19, 20] of the statistical software R [21].

An important characteristic of the LQMM and its parameter estimation procedure, is that it can give subject-specific RIs with only few repeated measurements for each subject. This is a typical feature of random effects models: the random effects distribution allows for information-sharing between subjects.

3   Results and Discussions

### 3.1   Population Reference Intervals for the Balearic Data

For each parameter a boxplot was produced with reference lines corresponding to the lower and upper bounds of PRIs that have been previously published [22–26]. Figure 1 shows boxplots for two physiological and two clinical parameters, split by gender. The boxplots for the other parameters can be found in Appendix. The graphs illustrate that for some parameters there may be difference between males and females. To the contrary, the published PRIs used in clinical practice often do not have gender-specific intervals. The example of this case can be seen in systolic blood pressure, body mass index, diastolic blood pressure, and triglycerides level (Figure 5 in Appendix). This suggests that, for these parameters, it may be better to work with PRIs for subpopulations.

The numerical results for the published PRIs for all ten parameters are shown in Table 3. Figure 1 also shows a fairly long tail in the distributions of almost all parameters. his is an indication of a skewed distribution and hence the parametric methods based on the normal assumption may not be appropriate here. The nonparametric methods may thus be advised.

Figure 2 shows the published PRI and PRIs computed by applying the parametric and nonparametric methods (CRIs) to the Balearic dataset. The PRIs for all parameters can be found in Table 4. From Figure 2, it can be seen that the CRIs computed by the five parametric and nonparametric methods give wider intervals than the published PRIs. Only for waist circumference the published PRI is very close to the CRIs. Among the CRIs, the nonparametric methods generally give similar intervals as the parametric method. However, for some of the parameters such as HDL, BMI and glucose level (see Figure 6 in Appendix), the intervals calculated by the parametric method are quite different as compared to the nonparametric. In these parameters, we observed deviations from the Normal distribution. Since the parametric method relies on distributional assumptions (usually Gaussian), a departure from this assumption may result in different estimates of intervals of the nonparametric methods.
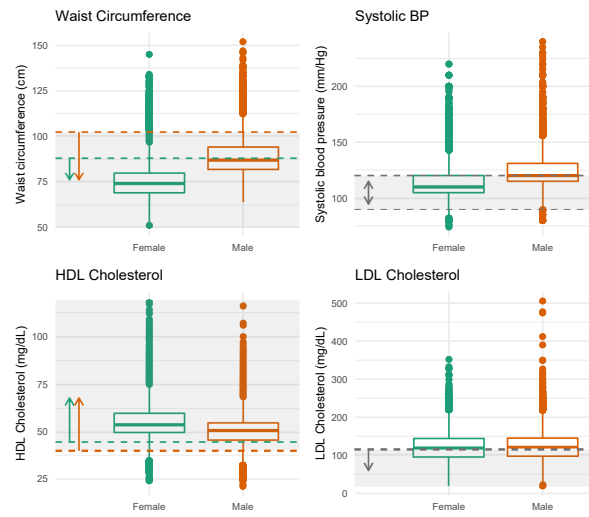
Fig. 1. Boxplots for 4 parameters in the physiological (top) and clinical (bottom) categories. The grey transparent area and the dashed lines correspond to the published PRIs while the arrows indicate their directions. The red and green dashed lines represent the lower/upper bounds of the published PRIs for males and females, respectively, and the grey dashed lines represent the published PRI for males and females together.
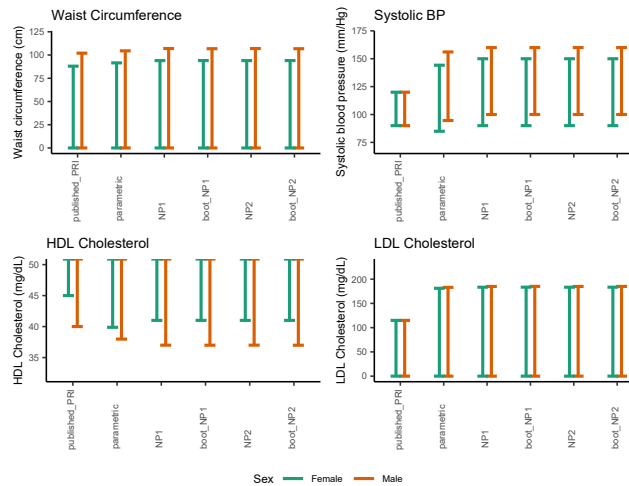
Fig. 2. For 4 parameters in the physiological (top) and clinical (bottom) categories, gender-specific PRIs are shown, estimated with various methods. Red and green lines represent the RIs for males and females, respectively. For waist circumference and LDL cholesterol, only the upper bounds were computed, and for HDL cholesterol only the lower bounds. For systolic blood pressure (BP) both lower and upper bounds were computed. For all calculations the Balearic dataset was used.

Table 3. Published PRIs for all ten parameters. Only for waist circumference and HDL cholesterol level, gender-specific PRIs are reported.

Parameter	Unit	Lower limit	Upper limit
Physiological category			
Waist circumference [1]	cm	0	F < 88, M < 102*
A body shape index [2]	Z-score	$-\infty$	0.229
BMI [3]	kg/m^2	18.5	24.9
Systolic BP [1]	mm/Hg	90	120
Diastolic BP [1]	mm/Hg	60	80
Clinical category [4]			
Total cholesterol	mg/dL	0	190
HDL cholesterol	mg/dL	F $\geq$ 45, M $\geq$ 40*	∞
LDL cholesterol	mg/dL	0	115
Triglycerides	mg/dL	0	150
Glucose	mg/dL	70	110

*F for females and M for males. [1] National Heart, Lung, and Blood Institute (NHLBI), [2] Krakauer & Krakauer (2012), [3] World Health Organization (WHO), [4] Laposata (2019).

Table 4. Reference intervals calculated by various methods for all parameters in the Balearic dataset.

| Parameter	Sex*	Published PRI	Parametric RI	NP1	Boot. NP1	NP2	Boot. NP2
Physiological category							
Waist circumference	M	(0, 102)	(0, 104.51)	(0, 107)	(0, 106.81)	(0, 107)	(0, 106.79)
	F	(0, 88)	(0, 91.55)	(0, 94)	(0, 94.04)	(0, 94)	(0, 94.04)
A body shape index	M	$(-\infty, 0.229)$	$(-\infty, 1.34)$	$(-\infty, 1.29)$	$(-\infty, 1.29)$	$(-\infty, 1.29)$	$(-\infty, 1.29)$
	F	$(-\infty, 0.229)$	$(-\infty, 0.77)$	$(-\infty, 0.99)$	$(-\infty, 0.99)$	$(-\infty, 0.99)$	$(-\infty, 0.99)$
BMI	M	(18.5, 24.9)	(18.59, 35.17)	(20.02, 36.82)	(20.01, 36.82)	(20.02, 36.82)	(20.02, 36.83)
	F	(18.5, 24.9)	(15.40, 34.65)	(18.36, 37.29)	(18.36, 37.31)	(18.36, 37.29)	(18.36, 37.30)
Systolic BP	M	(90, 120)	(94.61, 156.08)	(100, 160)	(100, 160)	(100, 160)	(100, 160)
	F	(90, 120)	(84.97, 144.21)	(90, 150)	(90, 150)	(90, 150)	(90, 150)
Diastolic BP	M	(60, 80)	(54.75, 97.24)	(60, 100)	(60, 100)	(60, 100)	(60, 100)
	F	(60, 80)	(50.02, 90.83)	(53, 93)	(52.65, 92.63)	(53, 92)	(52.72, 92.61)
Clinical category							
Total cholesterol	M	(0, 190)	(0, 260.53)	(0, 263)	(0, 263.00)	(0, 263)	(0, 263.02)
	F	(0, 190)	(0, 252.82)	(0, 256)	(0, 256.22)	(0, 256)	(0, 256.25)
HDL cholesterol	M	$(40, \infty)$	$(37.98, \infty)$	$(37, \infty)$	$(36.99, \infty)$	$(37, \infty)$	$(36.99, \infty)$
	F	$(45, \infty)$	$(39.88, \infty)$	$(41, \infty)$	$(41.00, \infty)$	$(41, \infty)$	$(41.00, \infty)$
LDL cholesterol	M	(0, 115)	(0, 183.05)	(0, 185)	(0, 185.09)	(0, 185)	(0, 185.08)
	F	(0, 115)	(0, 181.16)	(0, 183)	(0, 183.47)	(0, 183)	(0, 183.43)
Triglycerides	M	(0, 150)	(0, 271.30)	(0, 275)	(0, 275.03)	(0, 275)	(0, 275.01)
	F	(0, 150)	(0, 163.46)	(0, 167)	(0, 166.55)	(0, 167)	(0, 166.49)
Glucose	M	(70, 110)	(49.06, 132.18)	(65, 135)	(65.37, 135.40)	(65, 135)	(65.40, 135.44)
	F	(70, 110)	(55.66, 114.75)	(64, 113)	(63.99, 113.14)	(64, 113)	(63.99, 113.11)

* F for females and M for males.

When separately computing the PRIs for the males and females, we see that the bounds may be quite different. This is a first argument in favour of refining the PRIs towards smaller sub-populations. For example, for systolic blood pressure there are no gender-specific reference intervals published, but when estimated from the Balearic data we observe a clear difference between males and females. Similar findings were also observed in the other parameters, which are displayed in Figure 6 in Appendix.

### 3.2   Subject-specific Reference Intervals for the IAM Frontier Data

The IAM Frontier dataset contains data of 30 individuals that were measured at nine time-points. The individual profiles are shown in Figure 3. They show for all subjects how the measurements evolve over time. The plot indicates differences between the two genders: males generally have larger waist circumference, higher systolic BP and higher LDL cholesterol than females, but they have lower HDL cholesterol. Females have higher HDL than males at least until the age of 50 [27, 28] and the difference on the sex hormones between males and females can explain this phenomenon [29]. The plot also suggests a large between-subject variability and small within-subject variability, which is a common characteristic of repeated measurements. This phenomenon can be quantified by the intra-class correlation (ICC). Figure 3 also shows the ICC for each parameter. A large ICC is an indication that the within-subject variance is small as compared to the between-subject variance, or, equivalently, that the correlation between observations of the same individual is large. Large ICCs are observed for waist circumference and HDL cholesterol levels. Systolic blood pressure, on the other hand, has an ICC of only 65%.
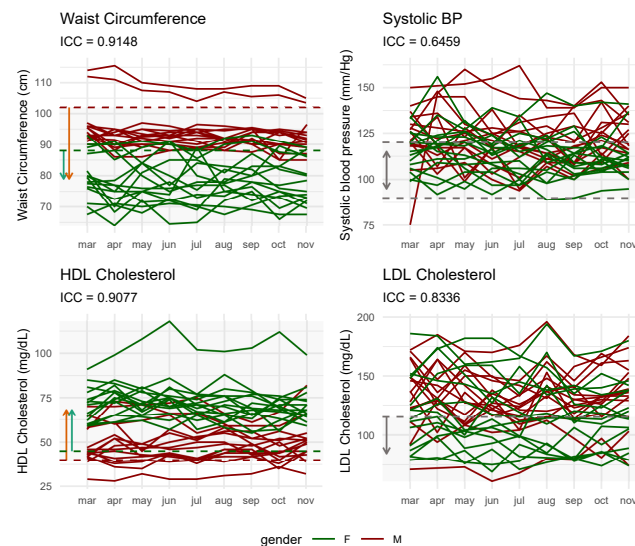
Fig. 3. Subject-specific profiles for all individuals in the IAM Frontier dataset. The grey transparent area and the dashed lines correspond to the published PRIs. The red and green dashed lines represent the lower/upper bounds of the published PRIs for males and females, respectively, and the grey dashed lines represent the published PRI for males and females together (no distinction between genders). The arrows indicate the directions of the intervals.

We argue that for parameters with a large ICC, a subject-specific RI (SSRI) would be preferred over a population RI (PRI). The former can be calculated from with quantile mixed models (LQMM). The results of this approach are displayed in Figure 4. The graph also shows the PRIs that were computed with

the classical nonparametic method, using all observations. Since these classical methods are not valid with longitudinal data, these PRIs are only shown for illustration purposes. The results for the other parameters can be consulted in Appendix. Figure 4 shows that the SSRIs vary between subjects. For the two-sided intervals, the SSRIs are generally smaller than the PRIs computed from the same data. For the one-sided intervals, we see that the SSRI bounds vary about the PRI bound; this variation follows the subject-specific observations. Our results suggests that SSRIs may be more informative than PRIs.
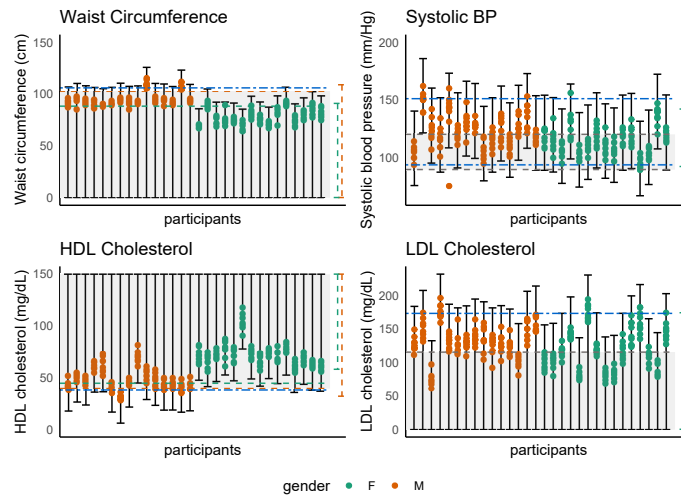
Fig. 4. SSRI for all subjects, estimated using LQMM with the IAM Frontier dataset. Red and green points refer to males and females observations. The grey area together with the red and green dashed lines indicate the published PRI for males and females, respectively, and the grey dashed lines indicate the published PRI for males and females together (no distinction between genders). The blue and the vertical red and green dashed lines indicate the estimated PRI and PRIs for males and females computed from the same data (the order of the individuals is randomised in each graph).

3.3   Discussion and Conclusion

We have applied conventional methods for estimating reference intervals for many parameters in the Balearic dataset, which comes from a cross-sectional study with 60,799 participants. Since such reference intervals are computed from a large cross-sectional sample from a reference population of healthy individuals, they are referred to as population reference intervals (PRI). **Our analyses demonstrated that parametric and nonparametric methods do not always give the same results, from which we conclude that it is better to rely on the nonparametric methods for they do not rely on distributional assumptions**. By computing reference intervals for subgroups of participants (e.g. males and females), we demonstrated that reference intervals for subpopulations may be different. This pleas for not using a single PRI for all subjects, but rather work with PRI for subpopulations.

In this paper, we considered reference intervals for individuals, referred to as Subject-Specific Reference Intervals (SSRI). Our motivation came from *the perspective of personalised medicine, which starts from the supposition that each person is unique, and from the observation in longitudinal data that often the within-subject variability of a clinical parameter over time is small as compared to the between-subject variability.* However, since longitudinal data often do not include a very large number of observations for each individual, the conventional nonparametric methods for RI calculation cannot be used for individual subjects. We have proposed to use linear quantile mixed models (LQMM) for the calculation of the SSRIs. This method makes use of the assumption that the upper (and lower) SSRI bounds vary between subjects as a normal distribution, allowing for the calculation of SSRIs even with only 9 observations per subject. We have applied the method to several parameters in the longitudinal IAM Frontier dataset. **The results show, as expected, that there is variability between the SSRI, which is an indication for the need of subject-specific intervals**. The results also show that for some parameters the lengths of the SSRI are smaller than those of the PRI. If such intervals were used in clinical practice then a deviation from the healthy status may be sooner detected. Similarly, for one-sided intervals, the bounds of the SSRI vary about the PRI, following the distribution of the repeated measurements of the individual.

Despite our first positive findings of the use of LQMM for the calculation of SSRI, more research is needed. The LQMM relies on the distribution assumption that quantiles vary between subjects according to a normal distribution. This assumption need to be assessed, and the consequences of deviations from this assumption need to be evaluated. Moreover, the theory behind the LQMM is asymptotic in nature, which does not guarantee that the SSRIs are unbiased when only limited numbers of time-points are available. Future research could focus on a thorough evaluation of the LQMM for SSRI calculcation and on further improving the methods so as to give reliable SSRIs even if model assumptions are not satisfied.

We believe that when SSRIs are widely used in clinical practice, they will allow for more precise diagnoses and hence they will be beneficial both for the patients and clinicians. We anticipate that in the future, the collaboration with artificial intelligence (AI) and machine learning (ML) algorithms could produce SSRIs for subjects for which even no longitudinal data is available. The well understood statistical methods produced from this research can perhaps eventually overcome the lack of algorithm transparency that is often criticised in the AI and ML approaches.

Appendix
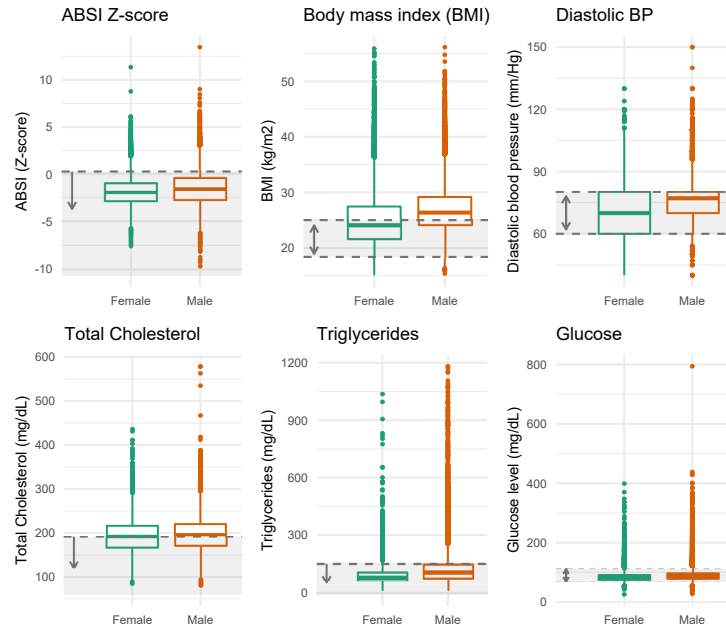
Fig. 5. Boxplots for 4 parameters in the physiological (top) and clinical (bottom) categories. The grey transparent area and the dashed lines represent the lower and upper bounds of the published PRIs for males and females together (no distinction between genders). The arrows indicate the directions of the intervals.
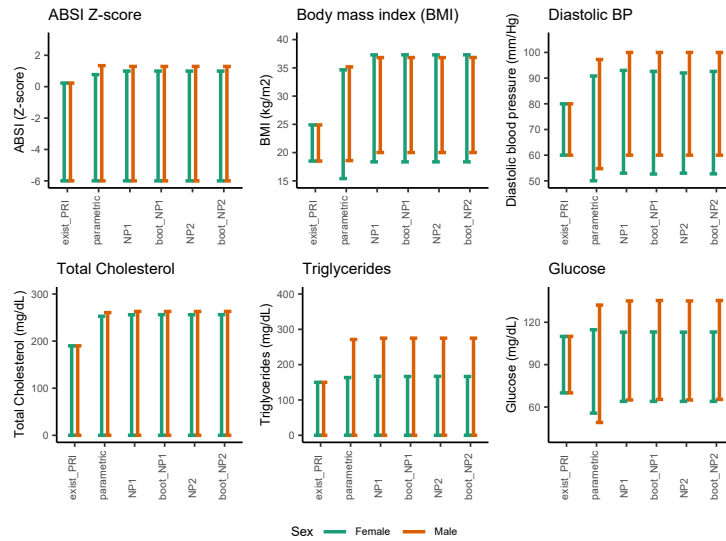
Fig. 6. For 4 parameters in the physiological (top) and clinical (bottom) categories, gender-specific PRIs are shown, estimated with various methods. Red and green lines represent the RIs for males and females, respectively. For ABSI Z-score, total cholesterol, and triglycerides level, only the upper bounds were computed. For diastolic blood pressure (BP) and glucose level both lower and upper bounds were computed. For all calculations the Balearic dataset was used.
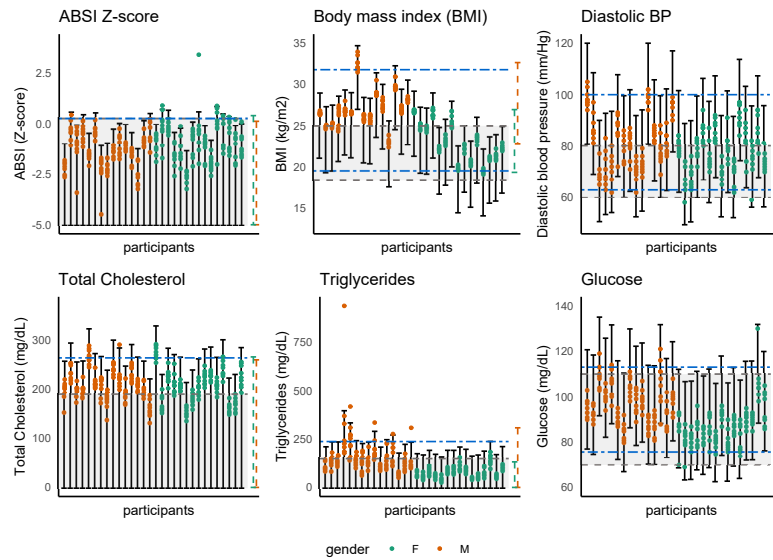
Fig. 7. Subject-specific reference intervals (SSRI) for all subjects, estimated using LQMM with the IAM Frontier dataset. Red and green dotted points refer to males and females observations. The grey transparent area together with the dashed lines indicate the published PRI while the blue and the vertical red and green dashed lines indicate the estimated PRI and PRI for males and females computed from the same data (the order of the individuals is randomised in each graph).

References

1. Zaveri, A., Ertaylan, G.: Linked data for life sciences. Algorithms **10**(4), 126 (2017)
2. Ščupáková, K., Balluff, B., Tressler, C., Adelaja, T., Heeren, R., Glunde, K., Ertaylan, G.: Cellular resolution in clinical MALDI mass spectrometry imaging: the latest advancements and current challenges. Clinical Chemistry and Laboratory Medicine (CCLM) (2019). https://doi.org/10.1515/cclm-2019-0858
3. de Kok, T.M., O'Donnovan, S., Lenz, M., Wimmenauer, F., Lukas, A., Neeven, J., Stumm, T., Krishna, A., Lopatta, D., Smirnov, Evgueni, Ertaylan, Gökhan, van Riel, N., Jennen, D., Cavill, R., Driessens, K., Peeters, R.: Deep learning methods to translate gene expression changes induced in vitro in rat hepatocytes to human in vivo. Toxicology Letters **314**, S170–S170(2019)
4. Graham, J., Barker, A.: Reference intervals. The Clinical Biochemist Reviews **29**(i), 93–97 (2008)
5. Rustad, P., Felding, P., Franzson, L., Kairisto, V., Lahti, A., Martensson, A., Petersen, P.H., Simonsson, P., Steensland, H., Auldall, A.: The Nordic reference interval project 2000, recommended reference intervals for 25 common biochemical properties. Scandinavian Journal of Clinical Laboratory Investigation **64**, 271–284 (2004)
6. Katayev, A., Balciza, C., Seccombe, D.W.,: Establishing reference intervals for clinical laboratory test results, is there a better way? American Journal of Clinical Pathology **133**(2), 180–186 (2010)
7. Ichihara, K., Yomamoto, Y., Hotta, T., Hosogaya, S., Miyachi, H., Itoh, Y., Ishibashi, M., Kang, D.: Collaborative derivation of reference intervals for major clinical laboratory tests in Japan. Annals of Clinical Biochemistry **53**(3), 347–356 (2016)
8. 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 **54**(6), 358–413 (2017)

9. Cheneke, W., Demie, T., Getachew, A., Bekele, S., Eskinder, K., Sahlemariam, Z., Adisu, W., Asres, Y., Yemane, T., Tesfaye, G., Tesfaw, G., Kebede, E., Mekonnen, Z., Gedefaw, L.: Reference interval for clinical chemistry test parameters from apparently healthy individuals in Southwest Ethiopia. Ethiopian Journal of Laboratory Medicine **5**(5), 62–69 (2018)
10. Royston, P.: Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. Statistics in Medicine **14**, 1417–1436 (1995)
11. Vogel, M., Kirsten, T., Kratzsch, J., Engel, C., Kiess, W.: A combined approach to generate laboratory reference intervals using unbalanced longitudinal data. Journal of Pediatric Endocrinology and Metabolism **30**(7), 767–773 (2017)
12. Romero-Saldaña, M., Tauler, P., Vaquero-Abellán, M., López-González, A., Fuentes-Jiménez, F., Aguiló, A., Álvarez-Fernández, C., Molina-Recio, G., Bennasar-Veny, M.: Validation of a non-invasive method for the early detection of metabolic syndrome: a diagnostic accuracy test in a working population. BMJ Open **8**(10), 1–11 (2018)
13. I AM Frontier study - VITO, Belgium, http://https://iammyhealth.eu/en/i-am-frontier. Last accessed 8 Jan 2020
14. Solberg, H.E.: Approved recommendation (1987) on the theory of reference values. Part 5: Statistical treatment of collected reference value. Determination of reference limit. Clinica Chimica Acta **170** S13–S32 (1987)
15. Linnet, K.: Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. Clinical Chemistry **46**(6), 867–869 (2000)
16. Koenker, R. Quantile Regression. Cambridge University Press (2005).
17. Geraci, M., Bottai, M.: Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics **8**(1), 140–154 (2007)
18. Geraci, M., Bottai, M.: Linear quantile mixed models. Statistics and Computing **24**, 461–479 (2014)
19. Geraci, M.: Linear quantile mixed models: The lqmm package for laplace quantile regression. Journal of Statistical Software **57**(13), 1–29 (2014)
20. Geraci, M.: **lqmm**: Linear Quantile Mixed Models. R package version 1.5, http://CRAN.R-project.org/package=lqmm. Last accessed 22 Jan 2020
21. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/. Last accessed 22 Jan 2020
22. World Health Organization: Mean Body Mass Index (BMI), https://www.who.int/gho/ncd/risk-factors/bmi-text/en/. Last accessed 22 Jan 2020
23. Krakauer, N.Y., Krakauer, J.C.: A new body shape index predicts mortality hazard independently of body mass index. Plos One **7**(7), e39504 (2012)
24. National Heart, Lung, and Blood Institute (NHLBI): Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. NIH Publication, Maryland USA (1998)
25. National Heart, Lung, and Blood Institute (NHLBI): Low Blood Pressure, https://www.nhlbi.nih.gov/health-topics/low-blood-pressure. Last accessed 22 Jan 2020
26. Laposata, M.: Laposata's Laboratory Medicine: Diagnosis of Disease in the Clinical Laboratory. 3rd edition. McGraw-Hill Education, Ohio USA (2019).
27. Kim, H.K., Park, H.A., Cho, Y.G., Kang, J.H., Kim, K.W., Kang, J.H., Kim, N., Chung, W., Kim, C.H., Whang, D.H., Park, J.K.: Gender difference in the level of HDL cholesterol in Korean adults. Korean Journal of Family Medicine **32**(3), 173–181 (2011)
28. Davis, C.E., Williams, D.H., Oganov, R.G., Tao, S.C., Rywik, S.L., Stein, Y., Little, J.A.: Sex difference in high density lipoprotein cholesterol in six countries. American Journal of Epidemiology **143**(11), 1100–1106 (1996)
29. Rossouw, J.E.: Hormones, genetic factors, and gender differences in cardiovascular disease. Cardiovascular Research **53**(3), 550–557 (2002)