

Anwar Borai*, Kiyoshi Ichihara, Abdulaziz Masaud, Waleed Tamimi, Suhad Bahijri, David Armbuster, Reo Kawano, Ziad Baarmah, Faris Joatar and Mohammed Almohammadi

Establishment of reference intervals for immunoassay analytes of adult population in Saudi Arabia

<https://doi.org/10.1515/cclm-2019-1049>

Received October 10, 2019; accepted January 17, 2020; previously published online March 11, 2020

Abstract

Background: This is a second part of report on the IFCC global multicenter study conducted in Saudi Arabia to derive reference intervals (RIs) for 20 immunoassay analytes including five tumor makers, five reproductive, seven other hormones and three vitamins.

Methods: A total of 826 apparently healthy individuals aged ≥ 18 years were recruited in three clinical laboratories located in western, central and eastern Saudi Arabia using the protocol specified for the global study. All serum specimens were measured using Abbott, Architect analyzers. Multiple regression analysis (MRA) was performed to explore sources of variation of each analyte: age, body mass index (BMI), physical exercise and smoking. The magnitude of variation of reference values (RVs) attributable to sex, age and region was calculated by ANOVA as a standard deviation ratio (SDR). RIs were derived by the parametric (P) method.

Results: MRA revealed that region, smoking and exercise were not relevant sources of variation for any analyte. Based on SDR and actual between-sex differences in upper limits (ULs), we chose to partition RIs by sex for

all analytes except for α -fetoprotein and parathyroid hormone (PTH). Age-specific RIs were required in females for ferritin, estradiol, progesterone, testosterone, follitropin, luteotropin and prolactin (PRL). With prominent BMI-related increase, RIs for insulin and C-peptide were derived after excluding individuals with BMI > 32 kg/m². Individuals taking vitamin D supplements were excluded in deriving RIs for vitamin D and PTH.

Conclusions: RIs of major immunoassay analytes specific for Saudi Arabians were established in careful consideration of various biological sources of variation.

Keywords: Architect; menopause; multicenter study; reproductive hormones; tumor markers; vitamin D.

Introduction

The reference interval (RI) is the most widely used tool for interpreting laboratory test results. This tool helps clinicians to differentiate between healthy and non-healthy individuals. RIs are typically derived from healthy subjects. The multicenter study conducted by the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL) was unique in the number of participating countries and number of included tests. In the first part of this study we attempted to derive RIs for 28 clinical biochemistry analytes, such as electrolytes, enzymes, glucose, lipids, iron, uric acid and proteins [1]. In this part we derived RIs for the most common analytes measured by immunoassays including hormones, tumor markers, vitamins, iron stores for the Saudi population according to harmonized protocol of the global study. We also investigated the association of those analytes with body mass index (BMI), sex, age, exercise and smoking status. In Saudi Arabia, there were studies targeting RIs for a few immunoassay analytes [2–4], but no study investigated RIs and influencing factors of their derivation for a diverse immunoassay analytes systematically.

The importance of our study resides in the following points:

1. It is part of the international multi-center project led by the IFCC, C-RIDL. The other collaborating countries,

*Corresponding author: Anwar Borai, PhD, King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Jeddah, Saudi Arabia, E-mail: boraiaa@ngha.med.sa

Kiyoshi Ichihara and Reo Kawano: Faculty of Health Sciences, Yamaguchi University Graduate School of Medicine, Ube, Japan. <https://orcid.org/0000-0002-6039-4537> (K. Ichihara); <https://orcid.org/0000-0001-8699-3768> (R. Kawano)

Abdulaziz Masaud, Waleed Tamimi, Ziad Baarmah, Faris Joatar and Mohammed Almohammadi: King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Jeddah, Saudi Arabia

Suhad Bahijri: Faculty of Medicine, Department of Clinical Biochemistry, King Abdulaziz University, Jeddah, Saudi Arabia

David Armbuster: Global Scientific Affairs, Abbott Diagnostics, Chicago, IL, USA

as of now, are the USA, Turkey, Japan, the UK, China, India, South Africa, Argentina, Russia, Pakistan, the Philippines, Nepal, Bangladesh, Kenya, Nigeria, Ghana, Egypt and Malaysia in the order of joining.

2. Twenty major immunoassay tests are targeted in the study.
3. Current analytical methods and the latest automated instruments were used.
4. The harmonized protocol developed in C-RIDL by consensus has been adopted for recruitment and data analyses
5. The Saudi population, similar in ethnic composition to the Arab gulf countries, is the only population involved in this global study, for which alcohol intake is prohibited. Therefore, this study provides a great opportunity to compare immunoassays reference values (RVs) with those from countries with widely different demographic profiles in term of typical partitioning factors such as BMI, alcohol intake and smoking.

As with part one of this study, it will be an opportunity to address the controversies over the selection between parametric (P) and non-parametric (NP)

methods for derivation of RIs [5, 6] and over what criteria to use in partitioning RVs by sex and age.

Materials and methods

The framework of the study was described in the first part of this report [1]. In brief, the number of apparently healthy Saudi subjects recruited in this study was 826 from across the Kingdom of Saudi Arabia. All subjects were recruited using the C-RIDL protocol with some modification to make it suitable for the culture of Saudi Arabia [7, 8]. Twenty-two of the subjects were excluded either due to unqualified status under the protocol or due to possession of overtly extreme values among test results. Subjects were distributed across the region as follow: western [Jeddah] (51% = 409), central [Riyadh] (19% = 152) and eastern [Hassa] (30% = 243). The ages of subjects were between 18 and 65 years (48.2% males, 52.8% females). Names, abbreviations, assay methods, and imprecision of the analytes are listed in Table 1.

The questionnaire items were adapted from the one used in the previous Asian project, but modified to meet Saudi local needs [7]. In each region, apparently healthy subjects were recruited from more than one city and from various professions as described previously. To eliminate the factor of different analytical methods, all samples were analyzed collectively in the central laboratory located in Jeddah as described in the global protocol.

Table 1: List of immunoassay tests with their principles of measurement and the calculated CV from mid-normal QC sera (within-day and between-day imprecision).

Abbr	Full test name	Architect method ^a	Within day CV	Between day CV	EFLM median CV ^b	
					Within-subject (CV) _i	CV _i /2
AFP	α -Fetoprotein	CMIA	4.0	3.2	26.7	13.4
CEA	Carcinoembryonic antigen	CMIA	2.7	5.0	17.9	9.0
CA125	Carbohydrate antigen 125	CMIA	1.3	4.6	9.1	4.6
PSA	Prostate specific antigen	CMIA	4.3	5.5	6.8	3.4
Ferritin	Ferritin	CMIA	3.6	3.3	NA	NA
VitB12	Vitamin B12	CMIA	2.2	10.3	NA	NA
Folate	Folate	CMIA	4.2	5.9	NA	NA
VitD	Vitamin D	CMIA	1.8	4.7	NA	NA
tHCY	Total homocysteine	CMIA	2.6	3.4	NA	NA
Insulin	Insulin	CMIA	2.2	3.4	27.5	13.8
CPep	C-peptide	CMIA	1.3	2.7	NA	NA
Cortisol	Cortisol	CMIA	4.9	5.4	24	12.0
Estradiol	Estradiol	CMIA	2.0	1.7	14.6	7.3
Progesterone	Progesterone	CMIA	7.4	7.8	NA	NA
Testo	Testosterone	CMIA	5.5	2.6	11.9	6.0
FSH	Follicle-stimulating hormone	CMIA	3.8	3.8	11.6	5.8
LH	Luteinizing hormone	CMIA	3.9	6.3	23.1	11.6
PRL	Prolactin	CMIA	3.6	3.2	19.9	10.0
TSH	Thyroid stimulating hormone	CMIA	2.0	4.7	15.9	8.0
PTH	Intact parathyroid hormone	CMIA	4.2	3.3	24.4	12.2

^aCMIA, Chemiluminescent microparticle immunoassay. ^bEFLM, European Federation of Clinical Chemistry and Laboratory Medicine (Biological Variation Database). CV, coefficient of variation; NA, not available.

Blood collection and handling

Participants were asked to avoid excessive eating/drinking the night before sampling. They were asked to be fasting for at least 10 h before sample collection. The amount of blood to be drawn was between 15 and 20 mL. The time of sampling was set at 7–10 AM. The blood was drawn after the participant had sat quietly for 30 min to avoid variations due to postural influence and physical stress. The waiting time was used for checking the questionnaire. For blood collection, four plain serum tubes containing clot activator were used. The tubes were left at room temperature before centrifugation, which was performed within 1 h. After separation of the serum, the specimens were then promptly divided into five to 10 aliquots of 1 mL each using well-sealed freezing containers (CryoTubes) and be immediately stored at -80°C . All the aliquots were shipped in a box filled with sufficient amount of dry ice to the central laboratory in the western region (Jeddah) for collective measurement. The study was approved by the King Abdullah International Medical Research Center, Research Ethics Committee, King Abdulaziz Medical City, Jeddah, Saudi Arabia (Study number RCJ0212-209).

Quality control

A quality control (QC) program was in use on a daily basis as usual. For the purpose of this study, a dedicated QC monitoring was also performed by use of multiple commutable specimens prepared in the central laboratory (Jeddah) as suggested by the C-RIDL common protocol and standard operation procedure 2 (SOP 2) [9]. Specifically, a mini-panel of sera from five healthy individuals (two males and three females) was prepared and measured over the period of collective measurements in order to assess between-day variations of test results.

RIs determined centrally in Jeddah laboratory were converted to those for each of the other participating laboratories (Riyadh and Hassa) through cross-checking results. Before sending samples to the central laboratory in Jeddah, 20 specimens were measured in each participating laboratory for the purpose of comparison. The linear structural relationship (the major axis regression) was used to convert RIs established by the centralized assay to the values of each participating laboratory. The participating laboratories in Jeddah, Riyadh and Hassa all used the same Abbott (Architect) analyzers.

Methods, instruments and reagents

In this second part of the Saudi multicenter RI study, we targeted 20 commonly tested analytes that were all measured using an Abbott, Architect i2000 analyzer. The analytes were five tumor markers (α -fetoprotein [AFP], carcinoembryonic antigen [CEA], carbohydrate antigen 125 [CA125], prostate specific antigen [PSA], ferritin), six reproductive hormones (estradiol, progesterone, testosterone, luteinizing hormone [LH], follicle-stimulating hormone [FSH], prolactin [PRL]), five other miscellaneous hormones (insulin, cortisol, C-peptide [CPep], thyroid stimulating hormone [TSH], parathyroid hormone [PTH]), three vitamins (vitamin B12 [VitB12], folate, vitamin D [VitD]) and total homocysteine [tHCY]. Assay reagents were

kindly donated by the manufacturer, Abbott, through Medi-Serve, Inc., (Saudi Arabia). Assay methods and their precisions are listed in Table 1. The auto-analyzer used for the centralized assays was an Abbott Architect i2000. The assays were performed at the King Abdulaziz Medical City Laboratory located in Jeddah, Saudi Arabia, using the manufacturer's reagents, calibrators and controls.

The results were retrieved and verified using stored data using i2000 Abbott software.

Statistical analysis

The tests results were evaluated by using the same statistical methods described in the previous studies [7–10] but methods are also described here.

Data validation: Data validation started by excluding subjects with overtly abnormal results (e.g. diabetic, hepatic, renal disease, active viral infection, etc.). This step represents the primary exclusion during recruitment and before blood sampling. Twenty-two subjects were excluded by identification of non-eligible criteria such as overt diabetes, early pregnancy, post-delivery <1 year, chronic diseases, donated blood within <4 weeks previously and the presence of abnormal hemoglobin variants.

Because we targeted heterogeneous combinations of analytes, the scheme for secondary exclusion of inappropriate subjects was customized analyte by analyte. For VitD and PTH, we excluded 123 individuals with VitD supplements, 24 individuals for ferritin because of iron supplement use. For TSH, we identified 10 individuals who were under thyroxine replacement therapy, but we ignored the fact because all were euthyroid. For reproductive hormones, we excluded 29 individuals who were using hormone replacement therapy (HRT) or using contraceptive pills.

Analyses of biological sources of variation: Multiple regression analysis (MRA) was performed separately for each sex with RVs for each analyte set as an objective variable. As explanatory variables, we constantly set age, BMI, level of regular exercise (0 [none] ~7 [every day]; the value halved for light exercise) and level of smoking (0 [none] ~2 [heavy smoker]). The association of a given explanatory variable with the objective one was expressed as standardized partial regression coefficient (r_p) corresponding to partial correlation coefficient with values between -1.0 and 1.0 . Because of analyzing hundreds of data, it is not appropriate to evaluate statistical significance of r_p by p-value. Therefore, we set its practical significance (effect size) to $|r_p| \geq 0.2$ as a midpoint effect size of Cohen's small (0.1) and medium (0.3) correlation coefficient [11] and its magnitude interpreted as "slight" for $0.20 \leq |r_p| < 0.30$, "moderate" for $0.30 \leq |r_p| < 0.50$ and "strong" for $0.50 \leq |r_p|$. Because the distributions of RVs of all analytes were skewed with long tailing to the higher side except for cortisol, prior to MRA (or ANOVA as described below), they were converted to a near Gaussian shape by either square-root or log transformation: the former for tHCY and folate, and the latter for all other analytes.

Partitioning criteria: The possible need for partitioning/subgrouping RVs by sex, region and age was primarily based on the actual magnitude of between subgroup differences expressed as a standard deviation (SD) ratio or SDR. In brief, the magnitude of between-region SD (SDreg), between-sex SD (SDsex), between-age SD (SDage),

and net-between individual SD (SD_{indiv}) were computed by the three-level nested ANOVA, and SDR for each factor was calculated as a ratio of each SD to the SD_{indiv}: i.e. SDR_{reg}, SDR_{sex} and SDR_{age}. The sampling sites were categorized into three regions represented by the main three cities (western; Jeddah, central; Riyadh and eastern; Hassa). Age was stratified into the following four groups: 18–29, 30–39, 40–49 and 50–65 years. The threshold level of SDR that requires partition by a given factor was set to 0.4 [11, 12]. In the calculation, each component of SDs for analytes that had been transformed were first calculated under the transformed scale, and then the SDs were reverse transformed to the original scale [12].

Because we occasionally encounter a situation where SDR does not represent actual between-subgroup difference at the lower or upper limits (LL, UL) of the RI, we adopted a secondary criterion called bias ratio (BR) at UL (or LL), BR_{UL} (or BR_{LL}), for judging the need to partition RVs. It was defined by the following formula as was used for a case of partitioning by sex.

$$BR_{UL} = \frac{UL_M - UL_F}{(UL_{MF} - LL_{MF}) / 3.92}$$

where UL_M , UL_F , and UL_{MF} denote ULs for male (M), female (F) and male + female (MF), respectively, and LL_{MF} denotes LL for MF. The numerator represents actual between-sex bias in UL, and the denominator represents SD comprising the RI, which corresponds to between-individual SD. Therefore, analogous with conventional specification for allowable analytical bias [13], we adopted 0.375 as a threshold for the minimal requirement for BR_{UL} (or BR_{LL}).

Derivation of RI: For analytes that showed close association with BMI by MRA, we examined the effect of excluding individuals with BMI ≥ 32 kg/m² by calculating BR_{UL} .

For calculation of RIs, the P method was primarily used, based on Gaussian transformation of RVs using the modified Box-Cox formula [14]. For comparative purpose, RIs were also calculated by a NP method. In both methods, confidence intervals (CIs) of lower and upper limits (LL and UL) of the RI were determined by the bootstrap method: i.e. the final dataset after the secondary exclusion steps was randomly resampled allowing replacement until the data size is the same as the source dataset, and RIs was computed from the resampled dataset, this resampling and recalculation of RIs was repeated 50 times, and CIs for LL and UL were predicted from the repeatedly calculated LLs and ULs of the RIs.

Results

Profile of the subjects

The demographic profile of the participants from each region was summarized in the first part of this study. The tabulation was made after deleting those with extreme values based on the criteria described in the Methods section. As a whole, there were more females than males (51.2% vs. 48.8%). The questionnaire shows 65% of males and 80% of females exercised <1 day per week.

More details about evaluation of between-region, gender, smoking status were described in the first part of this study.

QC monitoring

QC monitoring using routine QC specimens and a set of five sera (mini-panel) from healthy individuals was performed over the period of collective measurements. Between-day and within-day CVs computed from daily test results of mid-normal QC sera are as shown in Table 1. The critical level of CV was set to 1/2 of CV_i (within-individual CV), which was adopted from Biological Variation Database of the European Federation of Clinical Chemistry and Laboratory Medicine [15].

Sources of variation evaluated by MRA

MRA was performed separately for each sex to identify source factors associated with changes in RVs from among age, BMI, levels of regular exercise and smoking. As shown in Table 2, for age-related changes in males, a moderate increase with age was observed for AFP ($r_p = 0.395$) and FSH (0.392), a slight increase with age observed for VitD (0.220), and slight decrease with age for progesterone (−0.266), slight age-related increase was also noted for PSA (0.243).

On the other hand, females generally showed more pronounced age-related changes: a strong to moderate increase with age observed for FSH ($r_p = 0.638$) and LH (0.425), a moderate decrease with age for estradiol (−0.409) and testosterone (−0.373), a moderate increase was also noted for VitD (0.397), and a slight increase for AFP (0.265), CEA (0.225), ferritin (0.248), and VitB12 (0.263), and a slight decrease for PRL (−0.278) and progesterone (−0.237). These changes are illustrated in Supplementary Figure 1. Representative changes are listed in Figure 1.

For BMI-related changes, a slight to moderate increase with BMI was noted in males and females for insulin (0.289, 0.354) and CPep (0.368, 0.372), respectively. A moderate reduction due to the increase of BMI was observed for testosterone only in males (−0.344). In females, a slight increase proportionate to BMI was noted for PTH (0.251). These BMI-related changes are illustrated in Figure 2 for representative analytes. For the exercise and smoking-habit related changes, none of the analytes showed appreciable changes.

Table 2: Multiple regression analysis for sources of variation of RVs.

Male	n	R	Age	BMI	ExerLvl	SmkLvl	Female	n	R	Age	BMI	ExerLvl
AFP	380	0.394	0.395	-0.078	0.006	0.083	AFP	387	0.277	0.265	0.020	0.048
CEA	345	0.236	0.181	-0.136	0.023	0.115	CEA	326	0.213	0.225	-0.072	0.049
CA125	380	0.094	0.009	0.083	-0.017	-0.033	CA125	385	0.154	-0.135	-0.004	-0.074
PSA	378	0.262	0.243	-0.117	0.033	-0.049	PSA	389	0.064	-0.017	0.066	-0.015
Ferritin	380	0.269	0.156	0.168	-0.002	-0.095	Ferritin	388	0.266	0.248	0.038	-0.018
VitB12	372	0.157	-0.086	0.031	-0.058	-0.125	VitB12	377	0.245	0.263	-0.122	-0.004
Folate	379	0.292	0.233	0.081	-0.001	-0.120	Folate	383	0.205	0.160	-0.190	0.064
VitD	382	0.231	0.220	-0.079	0.045	-0.038	VitD	388	0.387	0.397	-0.126	0.126
tHcy	376	0.128	0.060	-0.068	0.014	0.098	tHcy	383	0.129	0.029	0.073	0.096
Insulin	373	0.320	0.044	0.289	0.025	0.129	Insulin	379	0.327	-0.157	0.354	0.010
CPep	382	0.454	0.160	0.368	0.023	0.194	CPep	385	0.450	0.139	0.372	-0.054
Cortisol	374	0.187	-0.073	-0.154	0.043	0.005	Cortisol	380	0.095	0.066	-0.096	-0.023
Estradiol	380	0.166	-0.117	0.127	-0.029	-0.044	Estradiol	385	0.410	-0.409	-0.002	-0.005
Progesterone	382	0.362	-0.266	-0.203	0.039	-0.002	Progesterone	389	0.267	-0.237	-0.030	0.089
Testosterone	372	0.423	-0.179	-0.344	0.013	0.081	Testosterone	387	0.390	-0.373	-0.035	-0.047
FSH	375	0.389	0.392	-0.009	0.036	0.009	FSH	388	0.620	0.638	-0.053	0.019
LH	375	0.079	0.020	-0.067	0.020	-0.036	LH	386	0.401	0.425	-0.104	-0.058
PRL	377	0.167	-0.117	-0.093	-0.009	-0.066	PRL	387	0.306	-0.278	-0.029	-0.099
TSH	339	0.183	0.037	0.048	-0.011	-0.169	TSH	357	0.125	0.043	0.095	0.041
PTH	379	0.195	0.149	0.089	0.006	-0.054	PTH	386	0.275	-0.089	0.251	-0.146

Listed values are standardized partial regression coefficients (r_p). $r_p = |0.2|$ was considered as a minimum effect size of practical significant. Graded red and blue background colors, respectively, indicate positive and negative correlation: light one, $0.2 \leq |r_p| < 0.30$; dark one $0.3 \leq |r_p|$. In females, SmkLvl was not analyzed because of low smoking rate of 1.9%. BMI, body mass index; ExerLvl, level of exercise level; SmkLvl, level of smoking-habit. Other abbreviations are as given in the text.

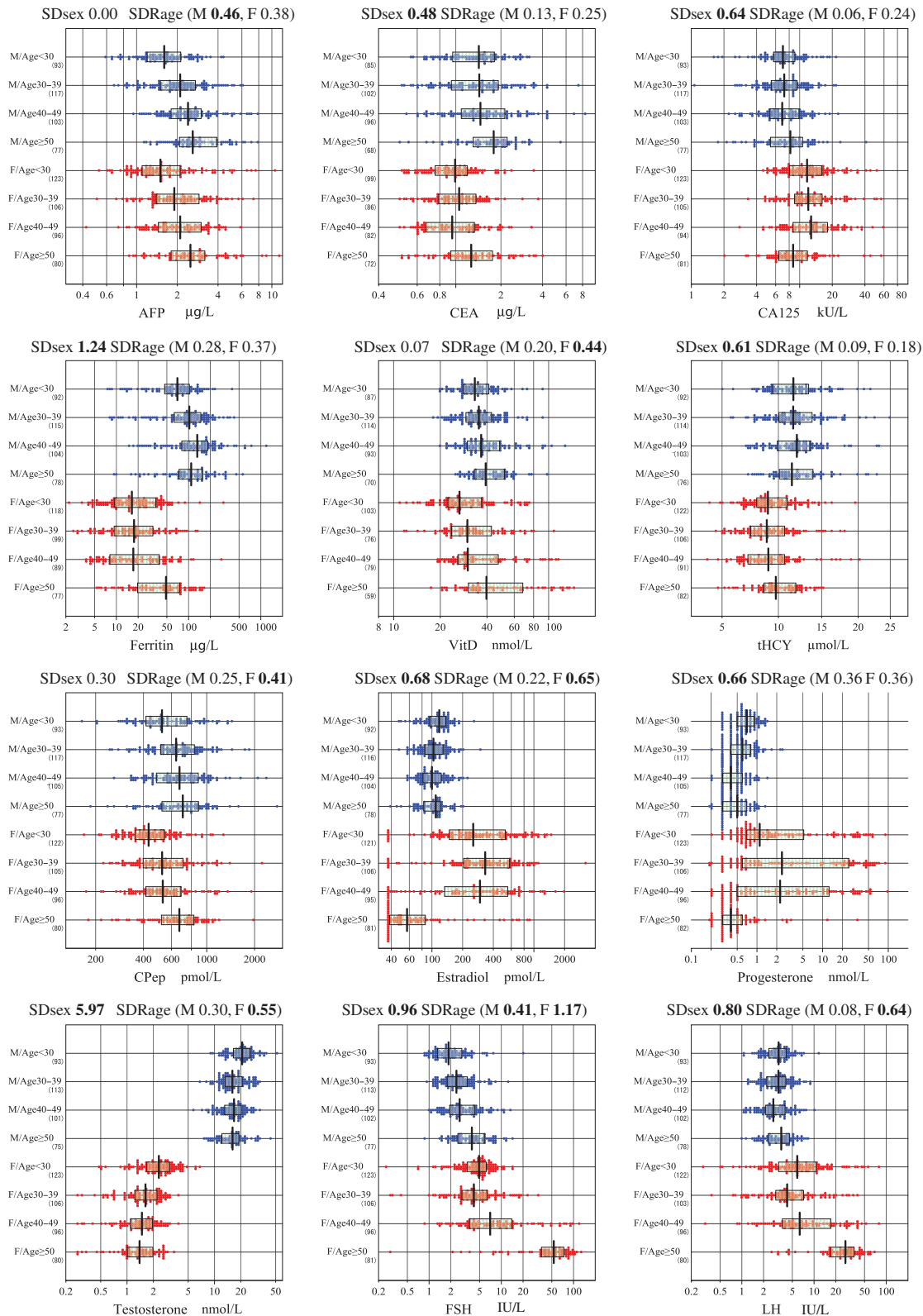


Figure 1: Sex and age-related changes in RVs of 12 major analytes.

RVs of 12 analytes with $\text{SDR} \geq 0.4$ for sex and/or age are shown subgrouped by sex and age (<30, 30–39, 40–49, 50+ years). The box in each scattergram represents central 50% range and the vertical bar in the middle represents median RVs. On top of each panel, the magnitudes of between-sex and between-age variations are shown as SDR_{sex} and SDR_{age} derived separately for males (M) and females (F). No secondary exclusion was performed in plotting data.

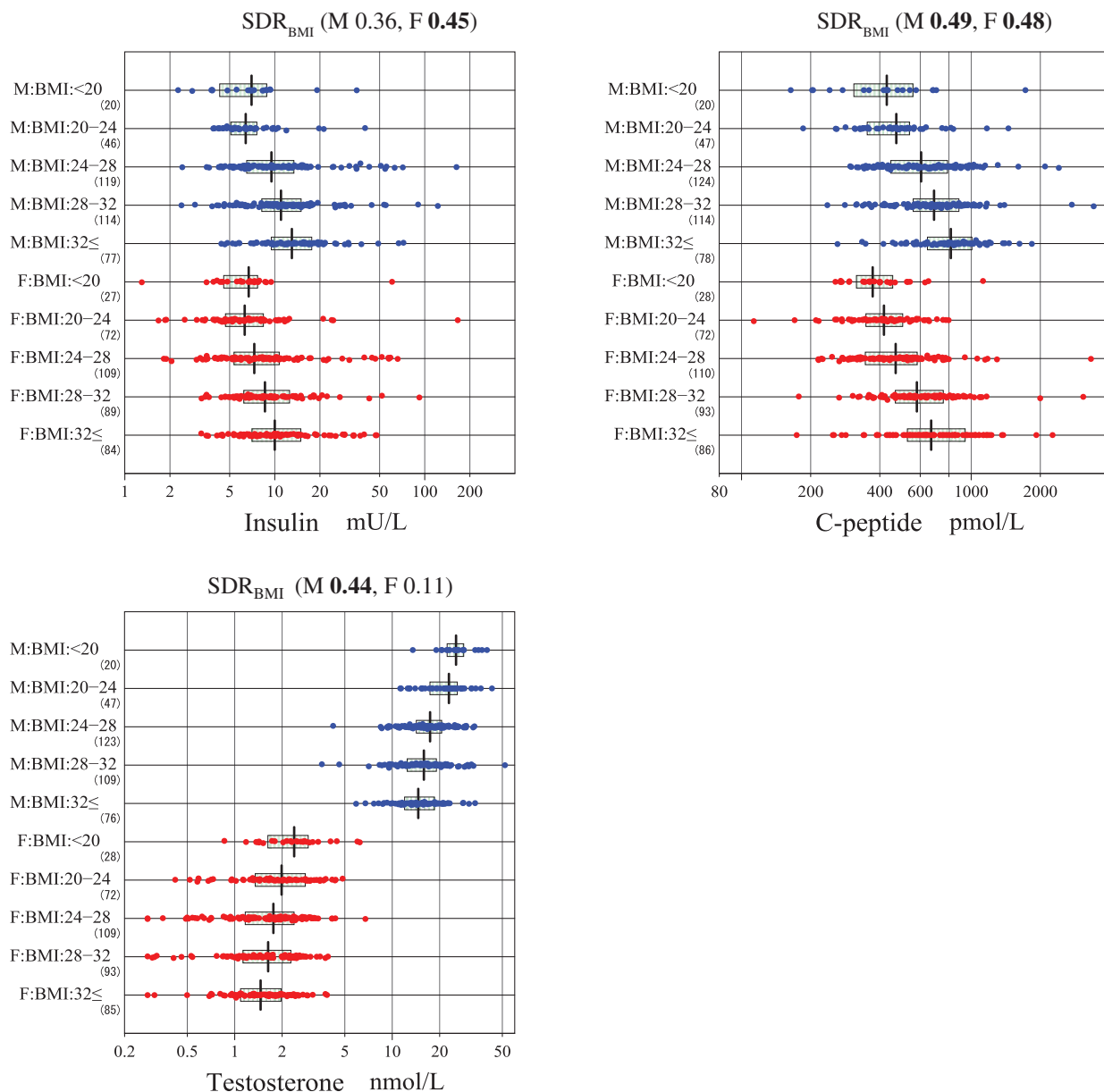


Figure 2: BMI-related changes in RVs of insulin, CPep and testosterone.

RVs of four analytes found associated with BMI by multiple regression analysis are shown subgrouped by sex and BMI (<20, 20–24, 24–28, 28–32, <32 kg/m²). The box in each scattergram represents central 50% range and the vertical bar in the middle represents a median point. The magnitude of between-subgroup variation is shown on top of each panel as SDR_{BMI} derived separately for males (M) and females (F).

SDR for sex and age-related changes

In judging the need for partitioning RVs by sex, region and age, we calculated the magnitude of between-subgroup variations of RVs as an SDR by use of ANOVAs: three-level nested ANOVA for combined analysis of the three factors, followed by a two-level nested ANOVA targeting region and age, performed separately for each gender. The results are summarized in Table 3. By adopting 0.4

as a significant effect size for SDR, between-sex differences expressed as SDR_{sex} was significantly high for 11 analytes (PSA, testosterone, ferritin, FSH, LH, estradiol, progesterone, tHcy, CA125, CEA and PLR, in that order of magnitude). The SDRs for region (SDR_{reg}) were below the threshold level for all analytes. SDRs for age (SDR_{age}) were significant for two analytes (AFP and FSH) in males, and in five analytes (FSH, estradiol, LH, testosterone, VitD, CPep) in females.

Table 3: Gender-, region- and age-related changes evaluated by the standard deviation ratio (SDR).

	SDRsex	SDRreg			SDRage			
		M + F	M	F	M + F	M	F	F [†]
AFP	0.000	0.000	0.000	0.000	0.419	0.462	0.384	
CEA	0.481	0.028	0.074	0.000	0.189	0.133	0.247	
CA125	0.644	0.000	0.000	0.000	0.183	0.057	0.240	
PSA	15.872	0.000	0.000	0.000	0.183	0.228	0.000	
Ferritin	1.243	0.000	0.000	0.000	0.329	0.277	0.368	
VitB12	0.000	0.000	0.026	0.000	0.236	0.035	0.304	
Folate	0.388	0.282	0.394	0.163	0.221	0.330	0.066	
VitD	0.069	0.030	0.155	0.000	0.378	0.204	0.444	
Homocysteine	0.612	0.206	0.281	0.072	0.144	0.093	0.182	
Insulin	0.144	0.230	0.157	0.279	0.212	0.194	0.228	
C-peptide	0.304	0.000	0.000	0.000	0.336	0.253	0.407	
Cortisol	0.000	0.196	0.297	0.079	0.119	0.127	0.114	
Estradiol [†]	0.682	0.000	0.000	0.000	0.638	0.216	0.654	0.789
Progesterone [†]	0.664	0.000	0.133	0.000	0.404	0.362	0.362	0.484
Testosterone [†]	5.974	0.210	0.000	0.257	0.484	0.301	0.548	0.394
FSH [†]	0.959	0.000	0.000	0.000	0.981	0.405	1.168	2.605
LH [†]	0.804	0.000	0.000	0.000	0.583	0.080	0.635	1.297
PRL [†]	0.420	0.186	0.353	0.000	0.232	0.109	0.283	0.379
TSH	0.074	0.188	0.221	0.170	0.120	0.119	0.121	
PTH	0.000	0.155	0.171	0.143	0.153	0.233	0.000	

SDRsex, SDRreg, and SDRage were first computed by three-level nested ANOVA, then gender-specific SDRreg and SDRage were computed by two-level nested ANOVA after partitioning RVs by sex. For computing SDRage for female sex hormones (†), RVs were partitioned in two ways: one by decade of age and the other by the status of MP (*1). SDR values that exceed 0.4 was shown in bold font and in three graded backgrounds color: $0.4 < \text{SDR} \leq 0.6$, $0.6 < \text{SDR} \leq 0.8$ and $0.8 < \text{SDR}$. SDR, standard deviation ratio; SDRsex, SDR for between-sex variations; SDRreg, SDR for between-region variations; SDRage, SDR for between-age subgroup variation.

Derivation of RIs

RIs were computed by both P and NP methods. Comparison of the RIs are shown in Supplementary Figure 2. It was obvious that ULs of RIs tended to be higher by the NP method than by the P method for the majority analytes. Ninety percent CIs of UL by the NP method were generally wider. On the other hand, the validity of the P method was confirmed by successful Gaussian transformation regardless of analytes as indicated by the linearity of probability paper plot (for the range of 10–90% cumulative frequency) and non-significant Kolmogorov-Smirnov (KS) test results shown in Supplementary Figure 3. The only exception was a bimodal distribution of progesterone RVs in female below 45 years of age, for which we adopted the RI by the NP method. We adopted RIs by the P method for all other analytes.

TSH was an exception to this scheme because we encountered an unnatural tailing of RVs toward the higher side, which we regarded as brought about by latent autoimmune thyroiditis. Therefore, we examined the effect of truncating TSH RVs above 8 mIU/L as they are obviously abnormal. As shown in Supplementary

Figure 4, we applied the probability paper plot analysis for each sex by setting cumulative frequency on Y-axis and TSH values on X-axis before and after the exclusion. The UL and LL of each RI was determined by a fitting least-square regression line, and reading out the cumulative frequencies of 2.5 and 97.5% as LL and UL. After this confirmation, we applied the P method to the truncated TSH RVs.

A list of RIs derived by various means are shown in Supplementary Table 1 for all analytes. The need for deriving RIs partitioned by sex, age or status of $\text{BMI} \leq 32 \text{ kg/m}^2$ were decided primarily on the basis of the magnitude of SDRsex and SDRage, and r_p for BMI. For deriving age specific RIs for sex hormones in females, we partitioned RVs by the status of menopause (MP), which was judged from age and information on menstruation in the questionnaire as well as a balance in values of estradiol, progesterone, FSH and LH. As a result, 90 belonged to the MP group and 317 to the pre-MP group. For the latter, 288 were valid for calculating the RIs after excluding females under HRT or taking contraceptive pills. On the other hand, for males and analytes other than sex-hormones, we chose to partition RVs arbitrarily at 45 years

Table 4: The list of RIs with 90% CI adopted in consideration of sex, age and BMI.

Item	Unit	Kit insert RI	BMI	Sex	Age	N	90% CI of LL		Reference interval			90% CI of UL	
									LL	Me	UL		
AFP	µg/L	0.9–8.8	All	MF	All	789	0.75	0.88	0.8	2.0	5.8	5.32	6.20
CEA	µg/L	0.0–5.0	All	M	All	350	0.56	0.62	0.59	1.45	3.93	3.40	4.46
			All	F	All	339	0.52	0.55	0.54	1.02	2.65	2.31	3.00
CA125	kU/L	0.0–35.0	All	M	All	386	2.2	3.2	2.7	7.4	16.1	14.5	17.7
			All	F	All	399	4.2	5.3	4.7	11.3	34.5	28.6	40.4
PSA	µg/L	<4.0	All	M	All	386	0.11	0.24	0.17	0.63	2.11	1.77	2.45
Ferritin	µg/L	101–536	All	M	All	384	1.7	9.5	6	101	296	254	337
			All	F	18–44	273	3.2	4.3	3.8	17	102	80	124
			All	F	45–65	131	0.0	3.6	1.6	32	163	135	191
VitB12	pmol/L	138–652	All	M	All	378	127	142	135	256	455	427	483
			All	F	All	391	123	135	129	248	577	522	631
Folate	nmol/L	7.0–46.4	All	M	All	388	5.7	7.1	6.4	15	30	29	32
			All	F	All	400	7.9	9.0	8.4	18.9	38	36	39
VitD	nmol/L	75–100	All	M	All	361	20	23	21	36	71	63	78
			All	F	All	312	15	20	17	33	102	85	119
tHCY	µmol/L	3.4–20.4	All	M	All	385	6.4	7.4	6.9	11.8	19.0	17.4	20.7
Insulin	mU/L	6.0–27.0	All	F	All	401	4.8	5.5	5.1	9.3	15.4	14.3	16.5
			All	M	All	378	3.2	4.4	3.8	9.8	35.0	25.5	44.5
CPep	pmol/L	364–1655	BMI < 32	F	All	286	2.2	4.1	3.1	7.4	26.2	10.5	42.0
			BMI < 32	M	All	300	220	284	252	607	1176	1094	1257
			BMI < 32	F	All	298	200	254	227	485	971	897	1046
Cortisol	nmol/L	101–536	All	MF	All	783	80	100	90	270	524	495	552
Estradiol	pmol/L	40.37–1615	All	M	All	388	52	60	56	107	178	169	186
			All	F	bfr MP	284	59	86	73	344	1301	1153	1449
			All	F	aft MP	88	29	35	32	59	293	144	441
Progesterone	nmol/L	<0.32–0.64	All	M	All	390	0.08	0.20	0.14	0.55	1.24	1.12	1.37
			All	F	bfr MP	288	0.18	0.35	0.30	1.73	59.3	54	68
			All	F	aft MP	87	0.18	0.21	0.20	0.41	1.4	0.4	2.5
Testosterone	nmol/L	M: 4.9–32.0 F: 0.38–1.97	All	M	18–44	268	5.2	9.6	7.4	17.8	32.3	30.3	34.4
			All	F	45–65	121	0.2	6.2	3.2	16.4	29.7	25.4	33.9
			All	F	bfr MP	286	0.4	0.6	0.5	1.9	4.2	3.8	4.6
FSH	IU/L	M: 1.0–12.0	All	F	aft MP	90	0.2	0.5	0.3	1.4	3.0	2.6	3.3
			All	M	18–44	265	0.9	1.1	1.0	2.2	6.8	5.8	7.9
			All	F	45–65	121	0.3	1.5	0.9	3.6	12.3	9.5	15.0
LH	IU/L	M: 0.57–12.07	All	F	bfr MP	283	1.24	1.92	1.58	5.1	15.3	12.6	18.1
			All	F	aft MP	90	4.58	18.2	11.4	57	119	107	131
			All	M	All	385	1.2	1.4	1.3	3.1	7.4	6.8	8.0
PRL	µg/L	M: 3.46–19.4 F: 5.18–26.53	All	F	bfr MP	281	0.8	1.9	1.3	5.8	35.2	27.6	42.7
			All	F	aft MP	90	3.4	12.7	8.1	28.1	60.9	50.9	70.8
			All	M	All	384	4.8	5.5	5.1	10.7	29.6	25.7	33.6
TSH*	mIU/L	0.35–4.94	All	F	bfr MP	283	6.0	8.2	7.1	16.7	47.3	40.2	54.4
			All	F	aft MP	87	3.6	5.4	4.5	10.4	30.5	15.4	45.7
			All	M	All	346	0.55	0.71	0.63	1.59	4.48	3.96	5.00
PTH	µg/L	15–68	All	F	All	359	0.45	0.65	0.55	1.72	5.77	5.04	6.51
			All	MF	All	674	23	28	25	56	120	112	128

All RIs were derived by the parametric method based on modified Box-Cox power transformation except the progesterone RI for females before menopause. Ninety percent CI of the limits of RIs were derived by the bootstrap method with 50-times iteration. For computing age-specific RI, RVs were in general partitioned at 45 years of age, but in females for sex hormones, RVs were partitioned by the status of MP as before or after MP (bfr MP, aft MP). The need for partitioning RVs by sex and age were decided primarily in reference to SDRsex and SDRage. The need for BMI restriction was decided by its effect on bias ratio (BR). The detailed on these decisions are shown in Supplementary Table 1. For TSH, subjects with TSH > 8 mIU/L were excluded assuming presence of latent autoimmune thyroiditis.

of age to ensure 100 or more subjects in the higher age group.

From the extensive list of RIs in the Table, we chose to adopt RIs partitioned by sex for CEA, CA125, ferritin, tHCY, estradiol, progesterone, testosterone, FSH and LH, based on their high SDRsex. As an exception, sex-specific RIs were also adopted for analytes with $\text{SDRsex} < 0.4$ when between-sex difference in ULs was high: i.e. $\text{BR}_{\text{UL}} \geq 0.375$, as commented in the column named “SDR & decision” of Supplementary Table 1. These include VitB12, folate, VitD, insulin, CPep and TSH. For AFP and cortisol, between-sex differences at UL were greater than the threshold of BR_{UL} , but the actual difference seen in Supplementary Figure 1 are obviously minor and thus we chose not to partition RVs for the two analytes.

For adopting age-specific RIs in males, we took into consideration both SDRage and between age-subgroup differences in the UL, BR_{UL} , in a similar manner as above. As a result, we found age partitioned RIs necessary for FSH and testosterone in males and for ferritin, estradiol, progesterone, testosterone, FSH, LH and PRL in females.

Due to the high association (r_p) between BMI and test results of insulin, CPep and testosterone (as shown in Table 2) and due to the high prevalence of increased BMI in Saudi individuals, we examined the effect of five graded BMI restrictions on RIs: BMI all, ≤ 32 , ≤ 30 , ≤ 28 and ≤ 26 kg/m² as shown in Supplementary Table 2. We found that, with stepwise increase in BMI restriction, data sizes reduced progressively, but the changes in RIs were not as prominent or mixed. Therefore, as a compromise for avoiding reduction in data size, we chose the threshold of BMI = 32 kg/m², which we adopted in deriving RIs for chemistry analytes [1]. We also examined the effect of applying the latent abnormal values exclusion method, which we applied in the study of chemistry RIs. However, the effect was a little less than that of BMI restriction (data not shown). Consequently, based on the BR of UL with/without BMI restriction shown in Supplementary Table 2, we adopted the BMI restricted RIs of insulin for females and CPep for both sexes.

The list of final RIs we adopted accordingly, from Supplementary Table 1, are shown in Table 4.

Discussion

Due to the heterogeneity of the included tests in this study, we encountered various difficulties in finding appropriate schemes for deriving RIs optimized for each analyte. For example, for insulin, CPep, testosterone and PTH, we

noted an appreciable association of their RVs with BMI and examined the effect of excluding individuals with $\text{BMI} \geq 32$ kg/m² ($n = 165$ or 20.5%). An appreciable lowering of UL was noted for insulin and CPep but not for testosterone and PTH.

One of the difficulties in deriving RIs is the situation in which RVs of all analytes, except for cortisol, were prominently skewed with long tailing to the high concentration side. In such a case, SDR, which represents the variations of subgroup means from the grand mean, tends to give a lower value, but actual between-subgroup differences at ULs, measured as BR_{UL} , tended to reveal wide between-UL difference. This phenomenon was true to between-sex differences in RVs of VitB12, folate, VitD, insulin, CPep, PRL and TSH, for which SDRsex was below the threshold of 0.4, but, by contrast, BR_{UL} well exceeded the threshold of 0.375. This finding led us to evaluate BR_{UL} in making the final decision for the need of partitioning RIs when the distribution of RVs is highly skewed.

In relation to the finding of highly skewed distributions of RVs in nearly all immunoassay analytes, it is of note that the P method was unanimously better than the NP method because the latter method is known to be highly sensitive to the presence of extreme values in the periphery of skewed RV distribution [16].

Regarding the regionality in test results for any of the 20 analytes examined within the three different regions of Saudi Arabia, based on SDRreg, we found no major regional differences in RVs for any analyte. However, a slight tendency of between-region differences was observed for folate, tHCY, cortisol, and PRL in males and for insulin and testosterone in females. These may be attributable to the fact that age of participants was lower in Riyadh, and BMI distribution of participants from Hassa was higher with some differences in lifestyles among the three regions of Saudi Arabia.

Results obtained from different countries involved in the global study compared to the Saudis were investigated in the first part of this study [17, 18]. One of the most significant results was the median BMI value for Saudis which was the highest compared to other participating countries for both males and females. Unfortunately the number of included immunoassays varies among the involved countries (Japan, China, India, Turkey, Russia, the United Kingdom, South Africa and the USA). In the interim report of the global study, it was possible to align RVs based on test results of a serum panel measured in common. Comparison of the panel test results for 17 immunoassays (five tumor markers, six reproductive hormones, four other hormones and two vitamins) were done across five to nine countries depending on the analyte. Although variable

degrees of between-laboratory bias existed in the majority of analytes, narrow-scatter around the regression line made it possible to align values across the countries. In this study, we re-aligned the test results of selected analytes from other countries to those of Saudi Arabia based on all-pairwise comparisons of the panel test results. From the analysis, we found RVs for insulin is peculiar to Saudi with the values highest among the countries in both genders (Supplementary Figure 5). Increased levels of insulin can be explained by the high prevalence of obesity and diabetes in Saudi population. Another interesting finding was the decreased level of ferritin in Saudi females. This is compatible with a previous report of low serum ferritin in Saudi females [19].

In Table 4, RIs are compared to the manufacturer RIs (kit inserts). Several immunoassays ULs and LLs were appreciably different than the kit insert RIs. For example, ULs for insulin and PRL in males are appreciably higher compared to the kit insert RIs. For PTH, the derived LL and UL are almost 60% and 72% higher than the kit insert LL and UL, respectively. The high PTH can be explained as a compensation for VitD deficiency in adult Saudi population as explained below.

Ferritin, folate, VitB12 and CPep immunoassays had lower ULs when compared to their kit insert limits. In Table 4, the lowered UL compared to kit inserts values for ferritin, VitB12 and folate strongly support our conclusion regarding the low iron level in Saudis compared to other populations in the first report [1] and the findings agree with a previous study [19]. Unfortunately, little information about samples collection, methods and procedures are given in the kit inserts. Therefore, most of the above observations cannot be discussed in detail. For example, the high ULs for PRL in Saudi males and females are increased for about 53% and 81%, respectively, when compared to the kit insert ULs.

In general, compared to the kit insert RIs, our derived RIs are more reliable because they are based on a well-defined harmonized protocol that specifies sampling procedures, measurements and statistical analyses considering sources of variation of each analyte.

The decreased VitD LL values for males and females (22 and 18 nmol/L, respectively) compared to the LL value suggested by the expert panel (75 nmol/L) as indicated by the kit insert [20], is evidence for the high prevalence of VitD deficiency in the adult Saudi population. This finding is supported by the previous published studies [3, 21, 22]. It is known that the recommended minimum level of VitD is 50 nmol/L, as advised by Food and Nutrition Board at the Institute of Medicine of the National Academies [23]. Our data shows that about 79.3% of males and 70.4% of

females in adult Saudi population who participated in this study are deficient in VitD (Supplementary Figure 6A). In addition to this, we found that the median level of PTH differed appreciably from the cut-off limit of VitD (<50 or ≥ 50 nmol/L) regardless of sex [$p < 0.001$ by the Mann-Whitney U test for both sexes] (Supplementary Figure 6B).

One limitation of this study was the relatively small number of subjects recruited to partition RVs of female fertility hormones according to the status of MP. For TSH, we did not test for antithyroid autoantibodies for the definitive exclusion of individuals with autoimmune thyroiditis. Therefore, we had to resort to the empirical graphical method to reduce the influence of individuals with latent thyroid dysfunction.

Conclusions

This study was conducted as a part of the international collaborative project for the derivation of RIs. It is the largest and first study for the Saudi population to establish RIs for the major immunoassay tests of high clinical demand by use of the internationally harmonized protocol. Partitioning of RIs by sex, age and BMI status was necessary for many analytes. As stated in our first report, the outcome from this study should be applicable to other neighboring countries in the Middle East, taking into account the common culture, religion, language, life style, foods and the prevalence of related diseases such as increased BMI and diabetes.

Acknowledgments: We are grateful to King Abdullah International Medical Research Center-Western Region (KAIMRC-WR) for their support and for the collaborating laboratory staff at King Abdulaziz Medical Cities in Jeddah, Riyadh, Hassa and Dammam for their valuable contributions to the success of the study. Finally, we are grateful for the collaboration and the kind support from the Saudi Society for Clinical Chemistry (SSCC).

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This study was planned by the C-RIDL of the IFCC and supported financially by the Japan Society for the Promotion of Science (JSPS) [No. 24256003: 2012–2014]. All reagents required for testing were generously offered by Abbott (Medi-Serve). The sampling equipment and all required facilities were provided by Ministry of Saudi National Guard.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Borai A, Ichihara K, Al Masaoud A, Tamimi W, Bahijri S, Armbuster D, et al. Establishment of reference intervals of clinical chemistry analytes for the adult population in Saudi Arabia: a study conducted as a part of the IFCC global study on reference values. *Clin Chem Lab Med* 2016;54:843–55.
2. Alqahatani M, Tamimi W, Aldaker M, Alenzi F, Tamim H, Alsadhan A. Young adult reference ranges for thyroid function tests on the Centaur immunoassay analyser. *Br J Biomed Sci* 2006;63:163–5.
3. Alsuwadia AO, Farag YM, Al Sayyari AA, Mousa DH, Alhejaili FF, Al-Harbi AS, et al. Prevalence of vitamin D deficiency in Saudi adults. *Saudi Med J* 2013;34:814–8.
4. Rabah DM, Farhat KH, Al-Atawi MA, Arafa MA. Age-specific reference ranges of prostate-specific antigen among Saudi men as a representation of the Arab population. *Med Princ Pract* 2019;28:242–46.
5. Ichihara K. Statistical considerations for harmonization of the global multicenter study on reference values. *Clin Chim Acta* 2014;432:108–18.
6. Ichihara K, Boyd JC, Intervals ICoR, Decision L. An appraisal of statistical procedures used in derivation of reference intervals. *Clin Chem Lab Med* 2010;48:1537–51.
7. Ichihara K, Ceriotti F, Tam TH, Sueyoshi S, Poon PM, Thong ML, et al. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. *Clin Chem Lab Med* 2013;51:1429–42.
8. Ichihara K, Ceriotti F, Kazuo M, Huang YY, Shimizu Y, Suzuki H, et al. The Asian project for collaborative derivation of reference intervals: (2) results of non-standardized analytes and transfer of reference intervals to the participating laboratories on the basis of cross-comparison of test results. *Clin Chem Lab Med* 2013;51:1443–57.
9. Ozarda Y, Ichihara K, Barth JH, Klee G, Committee on Reference I, Decision Limits IFCC, et al. Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values. *Clin Chem Lab Med* 2013;51:1027–40.
10. Ichihara K. Statistical considerations for harmonization of the global multicenter study on reference values. *Clin Chim Acta* 2014;432:108–18.
11. Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. *Clin Chem and Lab Med* 2010;48:1537–51.
12. Ichihara K, Itoh Y, Lam CW, Poon PM, Kim JH, Kyono H, et al. Sources of variation of commonly measured serum analytes in 6 Asian cities and consideration of common reference intervals. *Clin Chem* 2008;54:356–65.
13. Fraser CG. *Biological variation: from principles to practice*. Washington, DC: AACC Press, 2001:1–145.
14. Ichihara K, Kawai T. Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM470) and NCCLS proposed guideline (C28-P;1992): trial to select reference individuals by results of screening tests and application of maximal likelihood method. *J Clin Lab Anal* 1996;10:110–7.
15. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Biological Variation Database, 2018–2019. <https://biologicalvariation.eu/>.
16. Klee GG, Ichihara K, Ozarda Y, Baumann NA, Straseski J, Bryant SC, et al. Reference Intervals: comparison of calculation methods and evaluation of procedures for merging reference measurements from two US Medical Centers. *Am J Clin Pathol* 2018;150:545–54.
17. Ichihara K, Ozarda Y, Barth JH, Klee G, Qiu L, Erasmus R, et al. A global multicenter study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. *Clin Chim Acta* 2017;467:70–82.
18. Ichihara K, Ozarda Y, Barth JH, Klee G, Shimizu Y, Xia L, et al. A global multicenter study on reference values: 2. Exploration of sources of variation across the countries. *Clin Chim Acta* 2017;467:83–97.
19. Al-Buhairan AM, Oluboyede OA. Determination of serum iron, total iron-binding capacity and serum ferritin in healthy Saudi adults. *Ann Saudi Med* 2001;21:100–3.
20. Souberbielle JC, Body JJ, Lappe JM, Plebani M, Shoenfeld Y, Wang TJ, et al. Vitamin D and musculoskeletal health, cardiovascular disease, autoimmunity and cancer: recommendations for clinical practice. *Autoimmun Rev* 2010;9:709–15.
21. Ardawi MS, Sibiany AM, Bakhsh TM, Qari MH, Maimani AA. High prevalence of vitamin D deficiency among healthy Saudi Arabian men: relationship to bone mineral density, parathyroid hormone, bone turnover markers, and lifestyle factors. *Osteoporos Int* 2012;23:675–86.
22. Al-Raddadi R, Bahijri S, Borai A, AlRaddadi Z. Prevalence of lifestyle practices that might affect bone health in relation to vitamin D status among female Saudi adolescents. *Nutrition* 2018;45:108–13.
23. Institute of Medicine, Food and Nutrition Board. *Dietary reference intakes for calcium and vitamin D*. Washington, DC: National Academy Press, 2010.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2019-1049>).