Cureus
Part of SPRINGER NATURE

Open Access Original Article

Establishing Biological Reference Intervals From 2 Million Routine Hospital Laboratory Results of 59,141 Patients Using an Indirect Method: Evidence From Adult and Pediatric Moroccan Populations

Review began 10/14/2026
Review ended 10/24/2026
Published 10/31/2026

© **Copyright** 2026
Krimi et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118895

Kholoud Krimi [1, 2], Dounia El Moujtahide [1, 2], Abdelmounaim Kerkri [3], Mohamed Amine Madani [4], Elhoucine Sebbar [1, 2], Mohammed Choukri [1, 2]

1. Central Laboratory, Mohammed VI University Hospital, Oujda, MAR 2. Faculty of Medicine and Pharmacy, Mohammed First University, Oujda, MAR 3. Laboratory of Engineering Sciences and Intelligent Systems, National School of Applied Sciences, Mohammed First University, Oujda, MAR 4. Laboratory of Engineering Sciences and Intelligent Systems, National School of Applied Sciences, Mohammed First University, Oujda, MAR

Corresponding author: Kholoud Krimi, kholoudkrimi2020@gmail.com

Abstract

Objective: Establishing reference intervals (RIs) using direct sampling is costly and particularly challenging in pediatric populations. Moreover, RIs derived from Western populations may not be fully applicable to the Moroccan population. This study aimed to establish indirect RIs for 26 biochemical analytes in adults and children, with stratification by sex and age.

Methods: We retrospectively analyzed 2,563,513 laboratory results between January 1, 2022, and August 7, 2024. After cleaning, 1,972,082 results from 59,141 patients were retained, and only the earliest result per patient and analyte was used for estimation. Pathological values were handled statistically by modelling the main distribution peak: RIs were estimated using Box-Cox transformation and peak modelling. Each RI covered the central 95% of the modeled non-pathological distribution, and its limits were given with 90% bootstrap confidence intervals (CIs).

Results: RIs were successfully established for nearly all age and sex groups; several implausible pediatric limits were withheld. Several expected biological patterns were observed, including higher creatinine, uric acid, and gamma-glutamyl transferase (GGT) concentrations in men; higher alkaline phosphatase and phosphorus levels in younger children; increasing creatinine concentrations with age; and higher GGT and bilirubin levels in infants. Overall agreement with existing RIs was good for transaminases, alkaline phosphatase, urea, and creatinine. In contrast, shifts were observed in sodium, calcium, albumin, glucose, glycated hemoglobin (HbA1c), uric acid, GGT and lactate dehydrogenase (LDH) limits, which may be related to the hospital based population included in the study.

Conclusion: Indirect estimation using large-scale routine laboratory data provided plausible and cost-effective RIs for both adults and children in the Moroccan population. These findings may complement direct RI studies and provide a useful basis for local laboratory practice, although further validation in representative healthy populations remains necessary.

Categories: Hematology, Internal Medicine, Infectious Disease
Keywords: box-cox transformation, box–cox transformation, clinical biochemistry, clinical chemistry, indirect method, morocco, paediatrics, paediatrics., real-world data, reference intervals

Introduction

The reference interval (RI) is a fundamental tool that enables clinicians and laboratory professionals to determine whether a laboratory test result falls within an expected range. By international convention, it corresponds to the central interval containing 95% of the values in a healthy reference population, bounded by the 2.5th and 97.5th percentiles [1]. The quality of clinical interpretation therefore depends directly on the appropriateness of the RI used: an unsuitable interval may lead to misclassification, unnecessary additional investigations, or, conversely, missed diagnoses.

The so-called “direct” method, recommended by the Clinical and Laboratory Standards Institute (CLSI) and International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) guidelines, is considered the reference approach. It involves selecting a sample of healthy individuals according to strict criteria, collecting specimens, and calculating the corresponding percentiles [2]. However, this approach faces major challenges: it is costly and time consuming, and recruiting a sufficient number of healthy volunteers is difficult, particularly at the extremes of age (infants, young children, and older adults), where additional

ethical concerns arise regarding sample collection from healthy individuals [3]. Consequently, many laboratories, particularly in Morocco and more broadly in Africa, continue to use RIs provided by reagent manufacturers or derived from reference textbooks, which are largely based on predominantly Caucasian populations [4]. Reference values may vary according to ethnicity, diet, socioeconomic status, or altitude, meaning that their direct transferability to the Moroccan population cannot be assumed [4].

To overcome these difficulties, which programs such as Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) have addressed through direct sampling of healthy children [5], indirect methods have attracted renewed interest [3]. Initially proposed by Hoffmann in 1963 [6] and subsequently formalized by Bhattacharya [7], their principle is straightforward: results accumulated in the laboratory information system (LIS) represent a mixture of values from healthy individuals, who constitute the majority for many analytes, and from diseased individuals. Using appropriate statistical tools, it is possible to isolate the distribution of healthy individuals and derive reference percentiles without additional sampling [7,8]. Modern approaches, including truncated maximum-likelihood methods and refined algorithms, have substantially improved the reliability of this process [9,10], to the extent that indirect methods are now recognized as a legitimate alternative or complement to direct approaches for RI determination [3].

Indirect methods are particularly relevant in pediatrics, where the need is especially pronounced: concentrations of many analytes vary considerably with growth, requiring age-specific RIs, as demonstrated by the North American CALIPER program [11]. Several groups have therefore established pediatric RIs using hospital-based databases [12], while emphasizing the need to verify their validity [13].

In this context, we analyzed a large biochemical database from a Moroccan university hospital, covering nearly three years :. The primary objective was to establish indirect RIs for 26 routine biochemical analytes, by sex in adults and by age group in children. The secondary objectives were: (1) to compare adult intervals with conventional values, and (2) to compare pediatric patterns with CALIPER data, in order to assess the relevance and limitations of this approach in our population.

Materials And Methods

Study design and data source

This retrospective, observational, single-center study used data extracted from the information system of the Central Laboratory of Mohammed VI University Hospital, Oujda, Morocco, from January 1, 2022, to August 7, 2024. The database initially included 2,563,513 results, 430 analytes, and 52 requesting departments. Available variables were anonymized patient identifier, age, sex, test name, numerical result, requesting department, and date. No identifiable personal data were used.

Selection of analytes and data cleaning

We selected 26 routinely measured biochemistry analytes with high testing volumes, covering electrolytes, renal and liver function, proteins, glucose and glycated hemoglobin (HbA1c), and lipid profile. HbA1c and lipid tests were included because they are among the most frequently requested analyses, despite being primarily ordered in patients with suspected diabetes or dyslipidaemia. Therefore, the derived limits reflect the tested population and should not be interpreted as clinical decision limits. Analytes predominantly associated with specific pathological conditions, such as C-reactive protein, procalcitonin, troponin, and tumour markers, were excluded because indirect methods perform poorly for these tests [3]. Non-numerical results were removed, values were restricted to broad physiological ranges, and records with indeterminate sex or an age above 105 years were excluded. Individual pathological results were retained and addressed statistically by modelling the main distribution peak. To ensure statistical independence, only the earliest result for each patient and analyte was included. After data cleaning, the final dataset comprised 1,972,082 results from 59,141 patients.

Subgroup definition

RIs were estimated in adults ⩾18 years, overall and by sex, and in children divided into five age groups: 0-<1, 1-<3, 3-<6, 6-<12, and 12-<18 years. These partitions were defined a priori on physiological grounds. At least 120 values were required per subgroup, the CLSI minimum for direct estimation [2]. Because indirect methods require larger samples, estimates from subgroups with fewer than 1,000 values should be considered less reliable [3].

Statistical method for indirect estimation

For each analyte, non-diseased individuals were assumed to form the main distribution peak, while diseased individuals occupied the tails. Inspired by Bhattacharya and Arzideh [7,9], values were normalized using Box-Cox transformation [14], with λ estimated on the 5th-75th percentiles and constrained to 0-2. On the transformed scale, the non-pathological peak center was identified by Gaussian kernel density estimation, and σ was derived from peak curvature by fitting a parabola to the log-density near the mode. Reference limits were calculated as center $\pm 1.96\sigma$ and then back-transformed. 90% confidence intervals (CIs) were obtained by bootstrap resampling [2]. Adult RIs were compared with harmonized conventional values [15],

while pediatric patterns were compared with CALIPER data [11]. Analyses were performed using Python 3.12.

Ethical considerations

The study used anonymized routine data only, without intervention or additional sampling, and followed the Declaration of Helsinki.

Results

Population characteristics

In total, 1,972,082 results from 59,141 patients were analyzed. The sex distribution was balanced (27,684 men (46.8%) and 31,457 women (53.2%)). The median age was 47 years (interquartile range (IQR) 24-64 years). The study population included 47,856 (80.9%) adults and 11,285 (19.1%) children under 18 years, including 2,263 (3.8%) infants under one year of age, enabling coverage of the full age spectrum, including the neonatal period (Table *1*).

Characteristics	Value
Study period	January 1, 2022-August 7, 2024
Results analysed (after cleaning)	1,972,082 (76.9%)
Patients (1 sample per patient)	59,141 (100%)
Men	27,684 (46.8%)
Women	31,457 (53.2%)
Median age (IQR)	47 years (24-64)
Adults (≥ 18 years)	47,856 (80.9%)
Children (< 18 years)	11,285 (19.1%)
Of whom infants (< 1 year)	2,263 (3.8%)
Analytes studied	26 (routine biochemistry)

TABLE 1: Characteristics of the study population

IQR: Interquartile range

Principle of the method

Figure *1* illustrates the method using creatinine and alanine aminotransferase (ALT). On the original scale, both distributions are right-skewed, with a narrow peak of non-diseased individuals and a pathological tail that would falsely increase the upper reference limit if direct percentiles were used. After Box-Cox transformation, the non-pathological peak becomes nearly Gaussian, allowing estimation of its center and dispersion, calculation of the 2.5th and 97.5th percentiles, and back-transformation to the original scale. By focusing on the peak rather than the tails, the method remains stable despite pathological values (Figure *1*).

Figure 3. Principle of the indirect estimation (examples: creatinine and ALT)

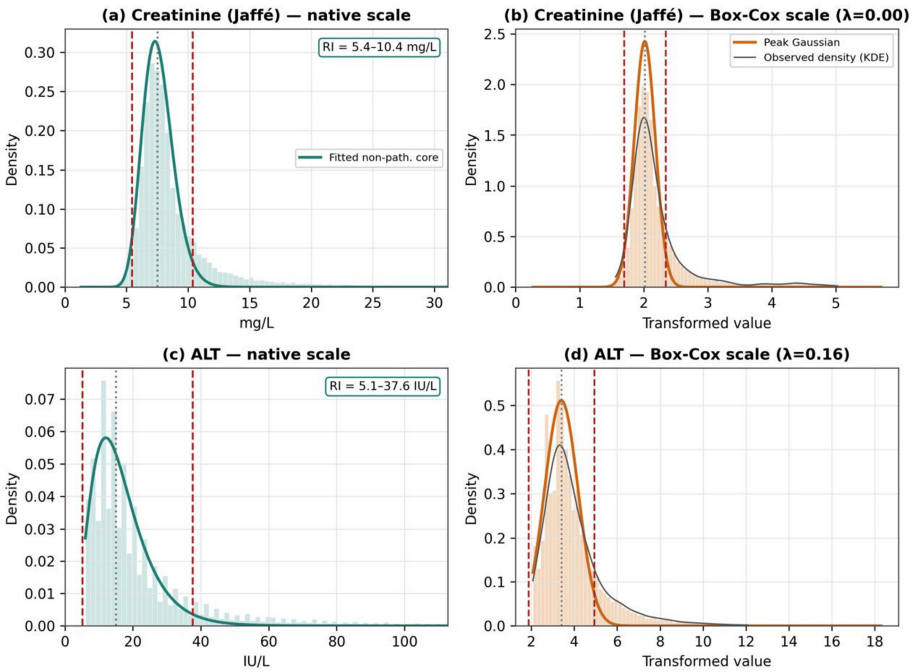


FIGURE 1: Principle of the indirect method illustrated with creatinine (Jaffé) and ALT in adults: histogram on the original scale (a, c) and after Box-Cox transformation (b, d), with the Gaussian fitted to the main peak and the estimated lower and upper reference limits (vertical lines)

ALT: Alanine aminotransferase

Adult RIs and sex-related differences

Adult sex-stratified RIs showed expected differences, with higher creatinine, uric acid, and gamma-glutamyl transferase (GGT) limits in men than in women, while electrolytes showed no meaningful sex differences. Pooled adult RIs were coherent for electrolytes, renal, liver, metabolic, and lipid markers. Two creatinine assays showed distinct RIs, with lower limits for the enzymatic method than the Jaffé method, illustrating the impact of analytical technique and supporting method-specific RIs (Table *2*).

Analyte	Unit	Sex	n	Median	Lower limit (90% CI)	Upper limit (90% CI)
Sodium	mmol/L	M	18,470 (38.6%)	138	131 (130–132)	144 (144–145)
		F	18,545 (38.8%)	138	132 (131–132)	144 (143–144)
Potassium	mmol/L	M	17,942 (37.5%)	4.2	3.1 (3.0–3.2)	5.0 (5.0–5.2)
		F	18,162 (38.0%)	4.1	3.1 (3.1–3.2)	4.9 (4.8–5.0)
Chloride	mmol/L	M	18,222 (38.1%)	105	97.0 (96.0–98.0)	112 (111–113)
		F	18,302 (38.2%)	106	98.8 (97.7–99.2)	113 (112–113)
Bicarbonate reserve (HCO_3^-)	mmol/L	M	14,614 (30.5%)	22.7	15.8 (14.6–16.4)	29.3 (28.4–30.3)
		F	14,511 (30.3%)	22.4	15.6 (14.7–16.5)	28.4 (27.1–29.1)
Total calcium	mg/L	M	18,340 (38.3%)	89.9	76.8 (75.1–77.5)	101 (100–103)
		F	19,314 (40.4%)	90.1	78.1 (76.8–78.9)	101 (100–102)
Phosphate	mg/L	M	4,476 (9.4%)	33.2	18.5 (17.0–18.9)	48.5 (47.9–50.2)
		F	5,248 (11.0%)	34.5	18.8 (16.5–19.3)	46.1 (45.8–47.4)

Magnesium	mg/L	M	1,710 (3.6%)	20.2	14.6 (14.1–15.0)	24.6 (24.3–24.9)
		F	1,874 (3.9%)	19.8	14.6 (14.0–15.0)	23.9 (23.6–24.3)
Urea	g/L	M	19,271 (40.3%)	0.30	0.15 (0.14–0.15)	0.54 (0.53–0.59)
		F	20,706 (43.3%)	0.26	0.11 (0.10–0.12)	0.47 (0.45–0.49)
Creatinine (Jaffé method)	mg/L	M	19,248 (40.2%)	8.4	5.9 (5.7–6.0)	11.8 (11.7–12.5)
		F	22,155 (46.3%)	7.2	5.1 (4.7–5.1)	9.4 (9.5–10.2)
Creatinine (enzymatic method)	mg/L	M	2,802 (5.9%)	7.4	4.4 (4.2–4.5)	11.7 (11.7–12.4)
		F	2,311 (4.8%)	5.8	3.5 (3.3–3.6)	9.6 (9.3–10.0)
Uric acid	mg/L	M	4,856 (10.1%)	54.5	19.5 (17.2–22.5)	87.6 (84.3–91.1)
		F	6,242 (13.0%)	46.3	19.9 (18.0–21.7)	77.4 (74.3–81.8)
Glucose	g/L	M	17,338 (36.2%)	0.95	0.61 (0.58–0.63)	1.5 (1.4–1.6)
		F	19,668 (41.1%)	0.90	0.62 (0.60–0.63)	1.3 (1.3–1.3)
HbA1c	%	M	4,819 (10.1%)	5.7	4.7 (4.6–4.8)	6.8 (6.7–7.0)
		F	7,791 (16.3%)	5.6	4.7 (4.6–4.8)	6.7 (6.6–6.9)
Total protein	g/L	M	18,616 (38.9%)	71.3	56.6 (54.9–58.7)	83.4 (8.2–84.5)
		F	18,687 (39.0%)	72.5	58.7 (57.9–60.8)	84.0 (82.7–84.8)
Albumin	g/L	M	13,845 (28.9%)	40.3	26.2 (24.7–30.0)	50.7 (49.3–51.2)
		F	13,735 (28.7%)	40.9	31.9 (29.6–32.5)	48.2 (47.6–49.2)
ALT	UI/L	M	11,138 (23.3%)	16.4	6.0 (6.0–6.0)	45.9 (45.2–58.1)
		F	13,788 (28.8%)	14.2	6.0 (6.0–6.0)	31.5 (29.5–33.8)
AST	UI/L	M	11,124 (23.2%)	20.6	9.0 (8.5–9.7)	4.4 (44.7–53.6)
		F	13,752 (28.7%)	18.2	9.8 (9.3–10.1)	33.2 (32.3–36.0)
GGT	UI/L	M	9,131 (19.1%)	23.7	7.9 (6.9–8.4)	71.4 (67.8–89.6)
		F	10,140 (21.2%)	19.9	7.2 (6.2–7.5)	55.4 (54.6–71.3)
Alkaline phosphatase	UI/L	M	8,943 (18.7%)	72.7	38.9 (36.7–40.2)	125 (123–135)
		F	9,918 (20.7%)	73.6	34.1 (32.4–36.8)	137 (129–145)
LDH	UI/L	M	7,574 (15.8%)	273	123 (116–130)	608 (554–683)
		F	7,847 (16.4%)	259	122 (113–131)	551 (491–620)
Total biluribin	mg/L	M	8,613 (18.0%)	6.2	1.7 (1.6–1.9)	14.6 (13.9–15.4)
		F	9,797 (20.5%)	5.3	1.8 (1.7–2.0)	11.9 (11.6–12.9)
Direct bilirubin	mg/L	M	8,537 (17.8%)	2.5	1.0 (1.0–1.0)	5.8 (5.7–6.3)
		F	9,680 (20.2%)	2.1	1.0 (1.0–1.0)	4.9 (4.9–5.5)
Total cholesterol	g/L	M	6,326 (13.2%)	1.6	0.63 (0.57–0.70)	2.3 (2.3–2.4)
		F	9,406 (19.7%)	1.8	0.75 (0.67–0.90)	2.4 (2.4–2.5)
HDL cholesterol	g/L	M	6,174 (12.9%)	0.37	0.11 (0.081–0.15)	0.53 (0.51–0.55)
		F	8,713 (18.2%)	0.43	0.15 (0.062–0.19)	0.60 (0.60–0.63)
LDL cholesterol	g/L	M	6,101 (12.7%)	0.95	0.23 (0.18–0.31)	1.6 (1.5–1.6)
		F	8,619 (18.0%)	1.1	0.27 (0.24–0.38)	1.7 (1.6–1.7)
Triglycerides	g/L	M	6,799 (14.2%)	0.97	0.34 (0.30–0.36)	2.0 (2.0–2.3)
		F	9,925 (20.7%)	1.0	0.32 (0.30–0.37)	2.2 (2.1–2.4)

TABLE 2: Indirect adult (≥18 years) RIs, by sex. Limits correspond to the 2.5th and 97.5th percentiles; the 90% CI obtained by bootstrap resampling is shown in parentheses.

ALT and direct bilirubin lower limits correspond to the assay quantification limit. HbA1c and lipid limits describe the tested population and are not decision limits.

CI: Confidence interval; HbA1c: Glycated hemoglobin; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase; LDH: Lactate dehydrogenase; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; RI: Reference interval; M: Men; F: Women; n (%): Number of results per sex and its percentage of the 47,856 adults included in the study

Pediatric RIs and age-related variation

Pediatric RIs reproduced expected age-related patterns. Alkaline phosphatase and phosphate were high in young children, then declined towards adult values, with an adolescent rise linked to pubertal growth. Creatinine increased steadily with age, reflecting muscle mass. Infants showed markedly higher GGT and total bilirubin, consistent with hepatic immaturity and neonatal jaundice. Phosphate and aspartate aminotransferase (AST) decreased with age, while total protein increased, reflecting protein synthesis maturation. These profiles were close to CALIPER data for alkaline phosphatase, phosphate, creatinine, GGT, and bilirubin. In contrast, some limits were implausible, suggesting contamination by diseased children or estimator instability: ALT and AST upper limits at 3-<6 years (56.1 and 92.6 U/L), Jaffé creatinine upper limit at 6-<12 years (12.5 mg/L), enzymatic creatinine (1.0-17.4 mg/L) and glucose lower limit (0.21 g/L) in infants, triglyceride upper limits of 3.5-3.7 g/L before three years, and high-density lipoprotein (HDL) lower limits of 0.050 g/L. These limits were withheld from Table *3*. However, infant RIs should be interpreted cautiously, as they reflect rapidly changing neonatal physiology, frequent abnormalities, jaundice, and hemolysed specimens, making them hospital-based approximate ranges rather than normative values for healthy neonates (Table *3*, Figure *2*).

Analyte	Unit	0–<1 year	1–<3 years	3–<6 years	6–<12 years	12–<18 years
Sodium	mmol/L	128–146	130–142	131–143	132–142	132–143
Potassium	mmol/L	3.3–5.8	3.3–5.1	3.2–4.8	3.2–4.7	3.2–4.7
Chloride	mmol/L	98.6–115	99.4–113	98.1–113	98.8–112	99.0–112
Bicarbonate reserve (HCO_3^-)	mmol/L	15.1–25.0	16.1–24.2	15.8–24.4	16.1–25.6	16.1–26.4
Total calcium	mg/L	67.6–111	78.0–106	79.3–104	80.9–103	80.1–103
Phosphate	mg/L	14.9–75.2	31.4–62.4	26.0–62.2	27.4–59.4	16.4–59.2
Magnesium	mg/L	16.1–25.9	16.6–24.6	15.0–24.7	15.3–23.8	15.7–23.4
Urea	g/L	0.043–0.60	0.056–0.35	0.097–0.34	0.10–0.37	0.100–0.36
Creatinine (Jaffé method)	mg/L	2.9–6.2	2.3–7.4	3.1–7.7	3.1–12.5	3.8–11.3
Creatinine (enzymatic method)	mg/L	1.0–17.4	—	—	1.0–11.2	1.2–10.8
Uric acid	mg/L	10.6–87.5	11.4–57.7	10.0–50.7	14.6–57.1	13.3–62.8
Glucose	g/L	0.21–1.5	0.48–1.2	0.55–1.2	0.60–1.2	0.61–1.2
HbA1c	%	—	—	—	4.6–6.2	4.6–6.0
Total protein	g/L	37.1–73.7	57.7–76.7	57.7–78.7	61.5–80.3	60.5–83.4
Albumin	g/L	18.4–46.3	34.6–48.3	35.5–48.6	36.4–48.7	36.8–49.5
ALT	UI/L	7.1–52.9	6.0–31.2	6.0–56.1	6.0–24.5	6.0–24.0
AST	UI/L	15.1–111	14.8–56.9	10.5–92.6	12.0–42.8	9.8–35.6
GGT	UI/L	4.8–429	5.9–26.5	5.6–23.7	7.0–25.6	7.2–29.2
Alkaline phosphatase	UI/L	51.3–400	40.1–296	36.7–279	45.0–304	31.4–560
LDH	UI/L	202–2095	198–664	204–497	158–498	128–509
Total bilirubin	mg/L	31.5–226	0.80–7.5	1.1–8.2	1.2–9.2	1.3–11.5
Direct bilirubin	mg/L	1.0–12.9	1.0–4.6	1.0–2.5	1.0–7.0	1.0–5.4
Total cholesterol	g/L	—	—	—	0.72–1.9	0.85–2.0
HDL cholesterol	g/L	—	—	—	0.050–0.64	0.050–0.61
LDL cholesterol	g/L	—	—	—	0.077–1.3	0.39–1.4
Triglycerides	g/L	0.47–3.5	0.41–3.7	0.33–3.0	0.29–2.2	0.28–2.3

TABLE 3: Indirect pediatric RIs by age group (2.5th–97.5th percentile limits). A dash indicates an insufficient sample size (<120). Units are the same as in Table 2.

HbA1c: Glycated hemoglobin; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase; LDH: Lactate dehydrogenase; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; RI: Reference interval

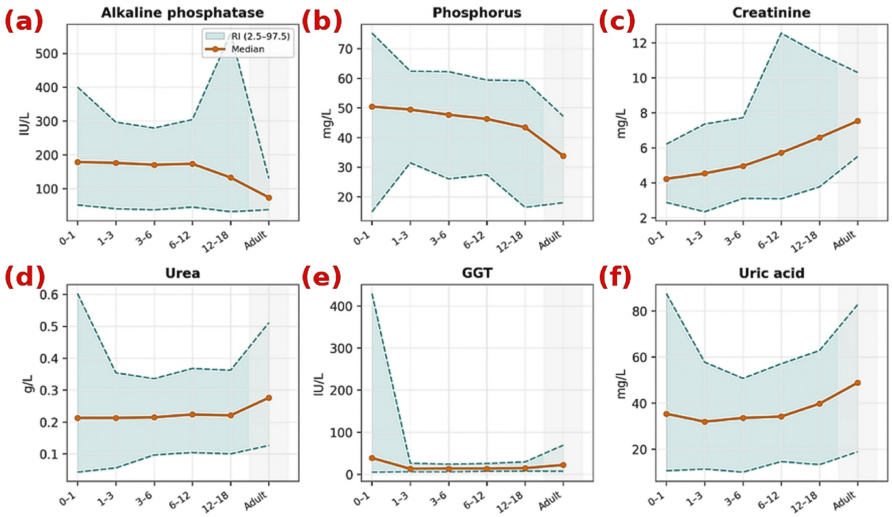


FIGURE 2: Age-related variation in RIs from childhood to adulthood

(a) Alkaline phosphatase, (b) Phosphorus, (c) Creatinine, (d) Urea, (e) GGT, and (f) uric acid. The orange line represents the median value for each age group. The teal dashed lines represent the lower and upper reference limits (2.5th and 97.5th percentiles), while the shaded teal area represents the corresponding RI. Age groups are shown from infancy to adulthood.

RI: Reference interval; GGT: Gamma-glutamyl transferase

Comparison with conventional reference values

The comparison between adult indirect RIs and conventional values was generally satisfactory, with good agreement for ALT, AST, alkaline phosphatase, urea, and creatinine. However, some limits deviated in relation to the hospital-based population. Lower limits for sodium, calcium, total protein, albumin, and bicarbonate were reduced, reflecting frequent hyponatraemia, hypocalcaemia, hypoalbuminaemia, and metabolic acidosis in patients from intensive care or oncology. Upper limits were higher for glucose and HbA1c, consistent with frequent diabetes and unknown fasting status, and for GGT and chloride; uric acid was wider at both ends. Conversely, lactate dehydrogenase (LDH) remained overestimated (Table *4*, Figure *3*). These discrepancies are presented in the Discussion section.

Analyte	Unit	Indirect RI (adults)	Conventional value
Sodium	mmol/L	131–144	136–145
Potassium	mmol/L	3.0–5.0	3.5–5.1
Chloride	mmol/L	98.1–112	98–107
Bicarbonate reserve (HCO_3^-)	mmol/L	15.5–28.2	22–29
Total calcium	mg/L	76.7–101	88–103
Phosphate	mg/L	17.9–47.1	25–45
Urea	g/L	0.13–0.51	0.15–0.45
Creatinine (Jaffé method)	mg/L	5.5–10.3	6–12
Uric acid	mg/L	18.9–82.7	30–70
Glucose	g/L	0.62–1.3	0.74–1.06
HbA1c	%	4.7–6.7	4–6
Total protein	g/L	57.8–84.0	64–83
Albumin	g/L	30.2–49.1	35–52
ALT	UI/L	6.0–38.1	10–40
AST	UI/L	9.4–38.5	10–40
GGT	UI/L	7.0–68.9	10–55
Alkaline phosphatase	UI/L	37.5–130	40–130
LDH	UI/L	123–562	135–225

TABLE 4: Comparison of indirect adult RIs with harmonized conventional values from the literature

Source: [16]

For lipids, the usual values are decision thresholds rather than RIs in the strict sense.

HbA1c: Glycated hemoglobin; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase; LDH: Lactate dehydrogenase; RI: Reference interval

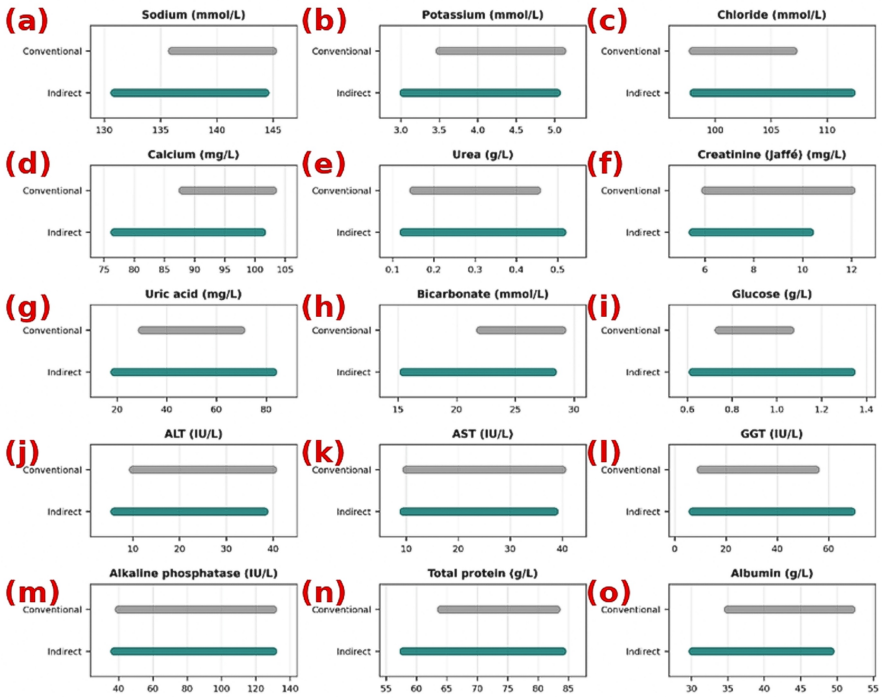


FIGURE 3: Comparison of indirectly estimated RIs with conventional reference values in adults

(a) Sodium, (b) Potassium, (c) Chloride, (d) Calcium, (e) Urea, (f) Creatinine (Jaffé), (g) Uric acid, (h) Bicarbonate, (i) Glucose, (j) ALT, (k) AST, (l) GGT, (m) Alkaline phosphatase, (n) Total protein, and (o) Albumin. Teal horizontal lines represent the RIs estimated indirectly from the university hospital laboratory data, whereas gray horizontal lines represent the conventional RIs used for comparison. The endpoints of each horizontal line indicate the lower and upper reference limits.

RI: Reference interval; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase

Discussion

This study demonstrates that it is possible to establish, from a large Moroccan hospital database and without any additional sampling, sex- and age-specific indirect RIs for most routine biochemistry analytes. Three main findings emerge: (i) the method faithfully reproduces the expected physiological variations, supporting its internal validity; (ii) it provides directly usable intervals for analytes minimally influenced by disease, but biased limits for those strongly affected by pathology; and (iii) for several analytes, it highlights discrepancies compared with imported values, thereby supporting the development of RIs specific to the Moroccan population. To our knowledge, this is among the first applications of this approach to a Moroccan population encompassing both adults and children.

Main findings and physiological coherence

The internal validity of the method is supported by its faithful reproduction of well-established physiological phenomena. Sex-related differences, including higher creatinine, uric acid, and GGT levels in males, are consistent with differences in muscle mass and metabolic characteristics. The feasibility of deriving pediatric RIs indirectly from routine laboratory data has also been demonstrated in previous studies, including the work of Zierk et al. [17]. Importantly, the pediatric patterns observed in our study are consistent with those reported by the CALIPER program [11], including elevated alkaline phosphatase and phosphate levels in young children due to intense bone activity, a progressive increase in creatinine with increasing muscle mass, and characteristic neonatal GGT and bilirubin values related to hepatic immaturity. This agreement, obtained independently from routine laboratory data, provides additional support for the relevance of the indirect approach to establishing pediatric RIs, particularly in populations in which direct studies are difficult to perform.

The two creatinine assays differed only in their lower reference limits. As they were performed in different patient populations, this difference may reflect both the populations tested and the analytical methods used. Therefore, these findings support the use of method-specific RIs but do not establish the sensitivity of the approach.

Strengths of the study

The principal strength of this work lies in the size of the dataset: 1,972,082 results from 59,141 patients. This confers a high degree of statistical precision on the estimates (narrow CIs) and allows for fine stratification by age and sex, including in infants. The method is also cost-effective, rapid, and reproducible, and it reflects the local population actually investigated by the laboratory. Finally, deduplication (one sample per patient) and the prior exclusion of analytes dominated by pathological conditions strengthen the validity of the estimates.

Where the Method Works… and Where It Reaches Its Limits

Our results clearly delineate the scope of validity of the approach, and the bias mechanism is itself informative. For an indirect method to be reliable on one side of a distribution, non-diseased individuals must predominate in the vicinity of the mode on that side. This condition is met at both ends for analytes that are only minimally affected by common hospital pathologies transaminases, GGT, alkaline phosphatase, urea, creatinine, and lipids for which the indirectly derived RIs, close to conventional values, appear readily usable. By contrast, when an abnormality is highly prevalent on one side, it encroaches on the modal region and shifts the corresponding limit. This explains the lowering of the lower limits for sodium, calcium, albumin, and bicarbonate reserve an immediate consequence of the high frequency of hyponatremia, hypocalcemia, hypoalbuminemia, and metabolic acidosis in an intensive care and oncology population as well as the upward widening of the glucose distribution (frequent hyperglycemia) and of LDH (a ubiquitous analyte, sensitive to hemolysis). Similarly, the neonatal class, characterized by physiological jaundice, yields an elevated upper limit for bilirubin. These observations are consistent with the cautions raised by Shaw et al. [18], who showed that pediatric RIs derived from hospital data may differ from those obtained in healthy children, and they confirm that indirect estimation cannot replace the direct method for analytes most sensitive to selection effects. In practice, interpretation of an indirectly derived limit should always take into account the prevalence, in the source population, of the pathology that alters the analyte in question.

Comparison with the literature

Our observations fit within and converge with an active line of research. In pediatrics, Zierk et al. indirectly derived age-continuous RIs [17] whose profiles reproduce those reported by the CALIPER program [11], consistent with our findings for alkaline phosphatase, phosphate, and creatinine. As early as 2006, a large Turkish study estimated indirect RIs for 40 analytes using more than 400,000 records and demonstrated good overall agreement with values obtained from healthy individuals [16]. From a methodological perspective, modern approaches, including truncated maximum likelihood [9] and, in particular, the refineR algorithm, which achieved the lowest estimation error in comparative evaluations [10], have strengthened the reliability of indirect RI estimation. This approach is now supported by international recommendations from the IFCC working group [3]. Finally, the need for population-specific RIs has also been highlighted in Morocco, where a study established hematological reference values specific to the population of northern Morocco because imported values were considered inappropriate [4]. Our work extends this rationale to biochemistry and pediatrics, areas in which this need remains largely unmet.

Clinical implications

From a practical standpoint, indirect estimation provides the laboratory with a low cost surveillance tool that can be readily updated. First, it makes it possible to identify analytes for which imported RIs clearly deviate from local values, thereby helping to prioritize candidates for a direct study or for a verification procedure, particularly because conducting a full direct study for each analyte is rarely feasible. Second, in pediatrics, it provides quantitative benchmarks where local reference standards are often lacking, which may reduce the risk of interpretive errors in children. Finally, because indirect estimation relies on patients actually tested by the laboratory and on the laboratory's own analyzers, it inherently incorporates institution-specific analytical and population characteristics. In accordance with recommendations [19], however, these intervals should be implemented in routine practice only after a local verification step, for example using a small sample of presumably healthy individuals, so that they complement rather than replace the direct method. Harmonized adult and pediatric RIs developed for chemistry analytes also provide a useful external framework for comparison with locally derived intervals [15].

Study limitations

Several limitations should be emphasized: (i) the dataset is single-center and hospital-based: the population does not reflect the general healthy population but rather the more severely ill patients referred to the university hospital, which biases analytes that are sensitive to selection effects, (ii) the lack of clinical information (diagnosis, fasting status, treatments) precluded any targeted exclusion of diseased individuals, (iii) age was available only in completed years, preventing a fine stratification of the neonatal period, which is nevertheless characterized by rapid changes, (iv) as with any indirect method, the approach assumes that non-diseased individuals dominate the distribution an assumption that is violated for certain analytes (sodium, calcium, albumin, LDH, neonatal bilirubin), for which the derived limits should be regarded as indicative, (v) partitions were set a priori without objective partitioning criteria [20,21], and children were not partitioned by sex, (vi) the requesting department was not used to exclude high-risk wards, (vii) possible

method or calibration drift over 31 months was not assessed, and the algorithm was not validated by simulation, (viii) assay quantification limits truncate some lower limits (ALT, direct bilirubin), and (ix) finally, the results were not prospectively verified in a sample of healthy individuals, a step that is essential prior to any clinical use [19].

Perspectives

This work opens several avenues for further research: confirmation of the intervals through a direct verification study in healthy individuals; comparison with the refineR algorithm; extension to other analyte families (hormones, specialized biomarkers); and, ideally, multicenter data pooling in order to propose harmonized RIs for the Moroccan population.

Conclusions

Based on nearly 2 million routine test results, indirect estimation made it possible to establish plausible, sex- and age-specific RIs for 26 common clinical chemistry analytes in Moroccan adults and children. The method faithfully reproduces physiological variations related to sex (higher creatinine, uric acid, and GGT in males) and growth (higher alkaline phosphatase and phosphate in young children, an age-related increase in creatinine, and characteristic neonatal values), supporting its reliability. It appears particularly valuable in pediatrics, where the direct method is the most difficult to implement and where local standards are most often lacking.

Its limitations for analytes most affected by hospital-based selection sodium, calcium, albumin, bicarbonate reserve, and LDH underscore that it should be used judiciously, taking into account the pathological profile of the source population, and that it represents a cost-effective complement to the direct method rather than a substitute; local verification remains essential prior to any clinical use. Overall, this work demonstrates the feasibility and relevance of the indirect approach in the Moroccan context and supports the longer-term, ideally multicenter, development of harmonized RIs specific to the Moroccan population, with the potential to improve the accuracy of biological interpretation and, ultimately, the quality of care.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Kholoud Krimi, Dounia El Moujtahide, Abdelmounaim Kerkri, Mohamed Amine Madani, Elhoucine Sebbar, Mohammed Choukri

Acquisition, analysis, or interpretation of data: Kholoud Krimi, Dounia El Moujtahide, Abdelmounaim Kerkri, Mohamed Amine Madani, Elhoucine Sebbar, Mohammed Choukri

Drafting of the manuscript: Kholoud Krimi

Critical review of the manuscript for important intellectual content: Kholoud Krimi, Dounia El Moujtahide, Abdelmounaim Kerkri, Mohamed Amine Madani, Elhoucine Sebbar, Mohammed Choukri

Supervision: Kholoud Krimi, Dounia El Moujtahide, Abdelmounaim Kerkri, Mohamed Amine Madani, Elhoucine Sebbar, Mohammed Choukri

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Ozarda Y: Reference intervals: current status, recent developments and future considerations . Biochem Med (Zagreb). 2016, 26:5-16. 10.11613/BM.2016.001
2. Clinical and Laboratory Standards Institute: EP28-A3C: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline, Third Edition. Clinical and Laboratory Standards Institute, Wayne, USA; 2010. https://clsi.org/shop/standards/ep28/.
3. Jones GR, Haeckel R, Loh TP, et al.: Indirect methods for reference interval determination - review and

recommendations. Clin Chem Lab Med. 2018, 57:20-9. 10.1515/cclm-2018-0073

4. Bakrim S, Motiaa Y, Benajiba M, Ouarour A, Masrar A: Establishment of the hematology reference intervals in a healthy population of adults in the Northwest of Morocco (Tangier-Tetouan region). Pan Afr Med J. 2018, 29:169. 10.11604/pamj.2018.29.169.13042
5. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N: The Canadian laboratory initiative on pediatric reference intervals: a CALIPER white paper. Crit Rev Clin Lab Sci. 2017, 54:358-413. 10.1080/10408363.2017.1379945
6. Hoffmann RG: Statistics in the practice of medicine . JAMA. 1963, 185:864-73. 10.1001/jama.1963.03060110068020
7. Bhattacharya CG: A simple method of resolution of a distribution into Gaussian components . Biometrics. 1967, 23:115-35. 10.2307/2528285
8. Haeckel R, Wosniok W, Arzideh F: A plea for intra-laboratory reference limits. Part 1. General considerations and concepts for determination. Clin Chem Lab Med. 2007, 45:1033-42. 10.1515/CCLM.2007.249
9. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, Haeckel R: A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-laboratory databases demonstrated by catalytic activity concentrations of enzymes. Clin Chem Lab Med. 2007, 45:1043-57. 10.1515/CCLM.2007.250
10. Ammer T, Schützenmeister A, Prokosch HU, Rauh M, Rank CM, Zierk J: refineR: a novel algorithm for reference interval estimation from real-world data. Sci Rep. 2021, 11:16023. 10.1038/s41598-021-95301-2
11. Colantonio DA, Kyriakopoulou L, Chan MK, et al.: Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem. 2012, 58:854-68. 10.1373/clinchem.2011.177741
12. Katayev A, Balciza C, Seccombe DW: Establishing reference intervals for clinical laboratory test results: is there a better way?. Am J Clin Pathol. 2010, 133:180-6. 10.1309/AJCPN5BMTSF1CDYP
13. Concordet D, Geffré A, Braun JP, Trumel C: A new approach for the determination of reference intervals from hospital-based data. Clin Chim Acta. 2009, 405:43-8. 10.1016/j.cca.2009.03.057
14. Box GE, Cox DR: An analysis of transformations. J R Stat Soc Series B Stat Methodol. 1964, 26:211-43. 10.1111/j.2517-6161.1964.tb00553.x
15. Tate JR, Sikaris KA, Jones GR, et al.: Harmonising adult and paediatric reference intervals in Australia and New Zealand: an evidence-based approach for establishing a first panel of chemistry analytes. Clin Biochem Rev. 2014, 35:213-35.
16. Ilcol YO, Aslan D: Use of total patient data for indirect estimation of reference intervals for 40 clinical chemical analytes in Turkey. Clin Chem Lab Med. 2006, 44:867-76. 10.1515/CCLM.2006.139
17. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, Rauh M: Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clin Chem. 2015, 61:964-73. 10.1373/clinchem.2015.239731
18. Shaw JL, Cohen A, Konforte D, Binesh-Marvasti T, Colantonio DA, Adeli K: Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clin Biochem. 2014, 47:166-72. 10.1016/j.clinbiochem.2013.11.008
19. Ozarda Y, Higgins V, Adeli K: Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations. Clin Chem Lab Med. 2018, 57:30-7. 10.1515/cclm-2018-0059
20. Harris EK, Boyd JC: On dividing reference data into subgroups to produce separate reference ranges . Clin Chem. 1990, 36:265-70. 10.1093/clinchem/36.2.265
21. Lahti A, Hyltoft Petersen P, Boyd JC, Fraser CG, Jørgensen N: Objective criteria for partitioning Gaussian-distributed reference values into subgroups. Clin Chem. 2002, 48:338-52. 10.1093/clinchem/48.2.338