Wendy P.J. den Elzen[a,*], Nannette Brouwer[a], Marc H. Thelen, Saskia Le Cessie, Inez-Anne Haagen and Christa M. Cobbaert

NUMBER: standardized reference intervals in the Netherlands using a 'big data' approach

https://doi.org/10.1515/cclm-2018-0462
Received April 30, 2018; accepted May 21, 2018

Abstract

Background: External quality assessment (EQA) programs for general chemistry tests have evolved from between laboratory comparison programs to trueness verification surveys. In the Netherlands, the implementation of such programs has reduced inter-laboratory variation for electrolytes, substrates and enzymes. This allows for national and metrological traceable reference intervals, but these are still lacking. We have initiated a national endeavor named NUMBER (Nederlandse UniforMe Beslisgrenzen En Referentie-intervallen) to set up a sustainable system for the determination of standardized reference intervals in the Netherlands.
Methods: We used an evidence-based 'big-data' approach to deduce reference intervals using millions of test results from patients visiting general practitioners from clinical laboratory databases. We selected 21 medical tests which are either traceable to SI or have Joint Committee for Traceability in Laboratory Medicine (JCTLM)-listed reference materials and/or reference methods. Per laboratory, per test, outliers were excluded, data were transformed to a normal distribution (if necessary), and means and standard deviations (SDs) were calculated. Then, average means and SDs per test were calculated to generate pooled (mean ± 2 SD) reference intervals. Results were discussed in expert meetings.
Results: Sixteen carefully selected clinical laboratories across the country provided anonymous test results (n = 7,574,327). During three expert meetings, participants found consensus about calculated reference intervals for 18 tests and necessary partitioning in subcategories, based on sex, age, matrix and/or method. For two tests further evaluation of the reference interval and the study population were considered necessary. For glucose, the working group advised to adopt the clinical decision limit.
Conclusions: Using a 'big-data' approach we were able to determine traceable reference intervals for 18 general chemistry tests. Nationwide implementation of these established reference intervals has the potential to improve unequivocal interpretation of test results, thereby reducing patient harm.

Keywords: big data approach; reference intervals; standardization.

[a]Wendy P.J. den Elzen and Nannette Brouwer contributed equally to this work.
***Corresponding author: Wendy P.J. den Elzen,** Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Postal Zone E2-P, P.O. Box 9600, 2300 RC Leiden, The Netherlands, Phone: +31 71 526 2278,
E-mail: wpjdenelzen@lumc.nl
Nannette Brouwer: Diagnost-IQ, Expert Centre for Clinical Chemistry, Purmerend, The Netherlands
Marc H. Thelen: Amphia Ziekenhuis, Laboratory for Clinical Chemistry and Haematology, Breda, The Netherlands; and Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek, Nijmegen, The Netherlands
Saskia Le Cessie: Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; and Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
Inez-Anne Haagen: Department of Hematology and Clinical Chemistry Laboratories, OLVG, Amsterdam, The Netherlands
Christa M. Cobbaert: Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, The Netherlands

Introduction

Interchangeability of laboratory test results across laboratories and in time is a major topic in laboratory medicine. The implementation of the Dutch External Quality Assessment (EQA) Program 'SKML Combi New Style' in 2005, using commutable and targeted sera, has proven to be very effective in reducing median inter-laboratory coefficients of variation for electrolytes, substrates and enzymes in the Netherlands [1]. However, despite the standardization of medical tests, national reference intervals and decision limits are still lacking for many of these tests in the Netherlands. Reference intervals are still established per laboratory using variable not evidence-based approaches: e.g. based on information from manufacturers' product inserts, from literature or from healthy controls analyzed by the laboratory itself. As a result, reference intervals

differ according to manufacturer, method principle and/or method generation and are often based on more strict pre-analytical conditions compared to those applied in daily clinical practice, leading to higher between-laboratory variation in reference intervals than the analytical variation in measurement results [2–4]. This hinders the national use of reference intervals in clinical guidelines, prevents adequate interpretation of laboratory test results and leads to incorrect and unequal treatment of patients [3, 4]. Interchangeability of laboratory test names, units and reference intervals in the Netherlands has become an absolute necessity due to recent developments, such as the introduction of electronical patient records in which laboratory results from different laboratories are combined, and the re-organization of the Dutch healthcare system in which patients with multiple health conditions are treated in more than one hospital or are treated by different physicians in primary, secondary and tertiary care.

The conventional settlement of reference intervals with the direct method, e.g. collecting and analyzing material from healthy control donors, is a time consuming and costly process. Following the CLSI protocol EP28-A3c, at least 120 samples should be tested, for each subcategory when applicable [5], leading to sometimes >1000 measurements per analyte when age and sex subcategories are necessary. The Committee on Reference Intervals and Decision Limits (C-RIDL) from the International Federation for Clinical Chemistry and Laboratory Medicine has published a detailed protocol on conducting multicenter reference interval studies, including requirements for recruitment of healthy volunteers (using health questionnaires), sample collection, cross-checking, ethics and data-analysis [6, 7]. These recommended methods for sample collection, centrifuging and storage differ from everyday practice [8], leading to reference intervals that will not always be applicable to routine practice because it will lead to excess flagging [9]. Given the workload, burden and costs of the direct method of establishing reference intervals, alternative methods are explored for establishing reference intervals, as described by the IFCC Committee on Reference Intervals and Decision Limits [10].

Under the umbrella of SKML/Calibration 2.000 [11], we have initiated a national endeavor named NUMBER (Nederlandse UniforMe Beslisgrenzen En Referentie-intervallen), in order to accomplish standardized reference intervals in the Netherlands. In this project, we adopt an indirect 'big data' approach to determine reference intervals [12], based on the evidence-based approach that was applied in Australia and New Zealand (https://www.aacb.asn.au/aboutus/harmonization-committee) [4, 13, 14], using millions of test results that are readily available in existing clinical laboratory databases from patients visiting general practitioners. The 'big data' approach is particularly usable in the Netherlands, because a unique, category 1 EQA system with value assigned, commutable EQA-materials, covering the clinically relevant concentration range, is in place. The Dutch EQA organization SKML collaborates with JCTLM-listed reference laboratories which perform value assignments [15] using primary reference methods and reference materials [16]. Furthermore, a multi-sample evaluation scoring system (MUSE) has been introduced [17], based on the Six Sigma metric combined with the bias and imprecision criteria from the Milan conference models [18], allowing monitoring of trueness and precision over time, based on predefined scientific criteria. This approach is also a method which allows periodic review of the established reference interval, as required by ISO 15189 (Chapter 5.5.2). We initially focused on SI-traceable general chemistry tests and enzyme tests defined by their reference measurement procedure, in which trueness can be verified 2-weekly by the Dutch EQA scheme 'SKML Combi General Chemistry' [15] using its category 1 trueness verification program.

Materials and methods

A schematic overview of the project outline is shown in Figure 1.

Study design

Clinical laboratories across the Netherlands were asked to provide anonymised laboratory results of patients visiting general practitioners. In the Netherlands, the general practitioner is gatekeeper to hospital and specialist care. All Dutch residents are registered in one general practice. Health insurance is mandatory, which covers a standard benefit package including primary care delivered by general practitioners.

We covered the main available manufacturers, matrices and geographic regions in the Netherlands, to enhance generalizability of the results.

Inclusion criteria for participating laboratories were:

- Adherence to the 'Venipuncture' guideline of the Dutch Society for Clinical Chemistry and Laboratory Medicine [19]
- Availability of test results ordered by general practitioners
- Participation in the Dutch EQA programs 'SKML Combi General Chemistry', with a sigma score ≥2 ('multi sample evaluation [MUSE]' score ≥1 in EQA reports) [17]
- Being able to supply anonymous test results for a period of 12 months (1 July 2015–1 July 2016) from the laboratory information system (Excel 2007 or above, .xlsx file)
- Provision of additional information on general pre-analytical conditions, test characteristics and performance of the test in the EQA during the period of data mining.

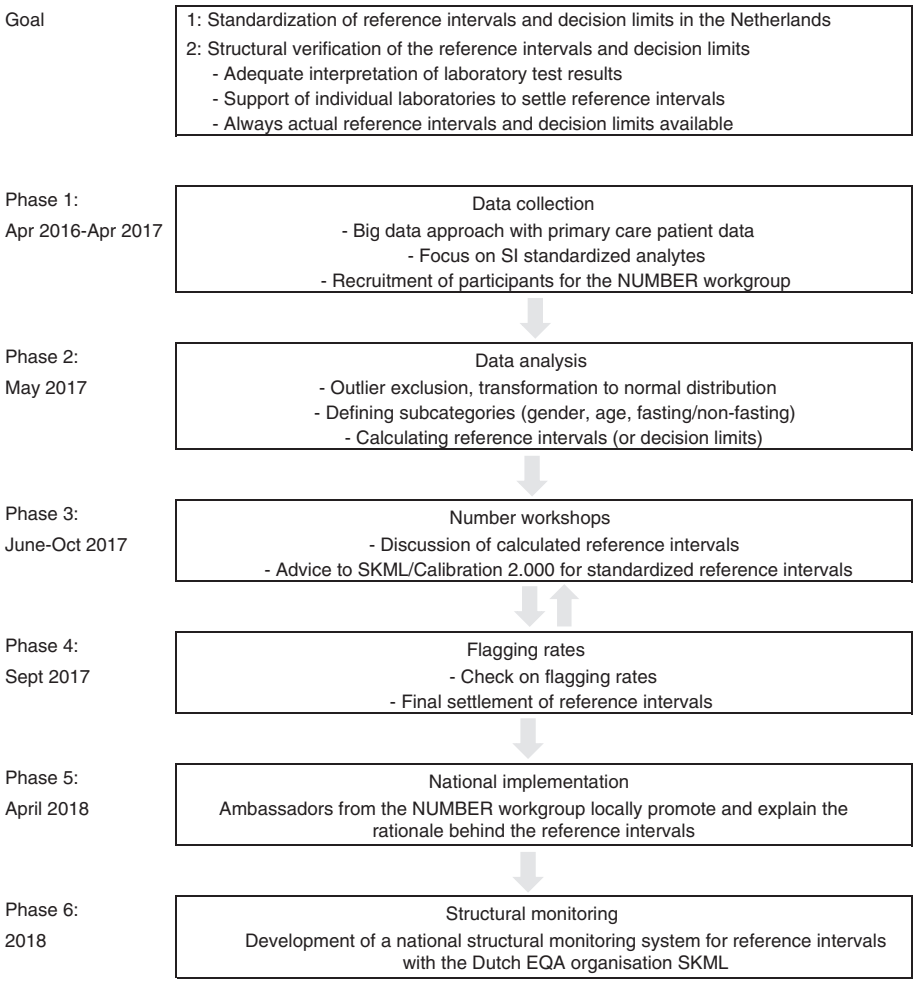

Figure 1: Project outline.

Data were centrally collected at the data center of the department of Clinical Chemistry and Laboratory Medicine at the Leiden University Medical Center.

Confidentiality of data and mutual performance of the participants was guaranteed and secured by a non-disclosure agreement, signed by the distributing and receiving parties. In addition, all participating laboratories gave permission to gain insight in the SKML EQA scores of the laboratories for the periods of data collection. As we received anonymous data from participating laboratories, the Medical Research Involving Human Subjects Act (WMO) did not apply to this project and we were exempted from obtaining ethical approval. However, the Medical Ethics Committee of the Leiden University Medical Center reviewed the study protocol and declared to have 'no objection' to the execution of this project in the Leiden University Medical Center (G16.056).

Data quality

We selected SI-traceable general chemistry tests from the Dutch EQA scheme 'SKML Combi General Chemistry (blood)' (www.skml.nl/rondzendingen) for which the reference measurement systems are listed in the JCTLM database (www.jctlm.org) and for which SKML has a trueness verification program in place [15]. Included analytes were serum/plasma alanine aminotransferase (ALT), albumin, alkaline phosphatase (ALP), amylase, (anorganic) phosphate, aspartate amino transaminase (AST), (total) bilirubin, calcium, chloride, creatinine, creatinine kinase (CK), γ-glutamyltransferase (GGT), glucose, iron, potassium, lactate dehydrogenase (LD), magnesium, sodium, total protein, urea and uric acid. The quality of the data was assured by excluding data from poor performers, defined as a 'multi sample evaluation (MUSE)' score of zero in EQA reports, indicating a total allowable error (Tea) sigma value below 2 [17].

Pre-analytical considerations

All samples were taken using a routine phlebotomy order of draw (citrate, serum, heparin, EDTA, sodium fluoride) as advised by the EFLM pre-analytical workgroup [20]. In the Netherlands, phlebotomy services are often located in or near general practices or hospitals. The Netherlands (41,528 km^2) has 134 general and university hospitals [21]. The average distance to a general practitioner is 0.9 km; six out of 10 inhabitants live within a radius of 5 km from a

hospital [22, 23], resulting in relatively short transportation times of blood samples to the laboratory. For the participating laboratories, maximum transportation time (i.e. maximum time to centrifugation) ranged from 10 min to 6 h. Fifty-six percent of the laboratories performed regular temperature monitoring in transportation cases. Centrifugation conditions ranged from 1800*g* to 3000*g*, 5–10 min (20 °C – room temperature).

Analytical considerations

From each laboratory only data for which their SKML participation resulted in a passing performance score were included (MUSE score ≥1, sigma ≥2) [17]. Only specific methods that measure well-defined *measurands* were considered for calculating reference intervals. For example, results from laboratories that used the Jaffe method for measuring creatinine, which exhibits considerable analytical variability (in terms of bias, imprecision and lack of specificity) [24], were not included in the statistical analyses. In addition, given the significant differences between bromcresol green (BCG) and bromcresol purple (BCP) dye-binding methods for albumin [25], we performed separate analyses for laboratories using the BCG and BCP method.

Matrix considerations

As potassium concentrations are known to be 0.1 to 0.7 mmol/L higher in serum than in heparin plasma as a result of platelet rupture due to clotting, analyses for potassium were stratified for serum and plasma [26, 27]. Moreover, separate analyses in serum and heparin plasma were performed for total protein, because serum normally contains ±4% less protein than plasma, mostly because soluble fibrinogen is converted to fibrin during the clotting process [28].

Clinical considerations

Laboratory results were excluded when phlebotomy was performed at home, as this is often only requested for severely ill patients. Iron measurements were only included when, at the same phlebotomy date and time, hemoglobin levels were available and results were >8.1 mmol/L for men or >7.5 mmol/L for women [29]. In addition, to reduce the risk of including glucose results from patients with diabetes visiting the laboratory for routine check-ups, glucose results were only included in the analyses when HbA_{1c} was not ordered at the same phlebotomy date and time.

Statistical analyses

The data analyses were performed in three steps.

First, reference intervals were calculated per test per laboratory. Per laboratory, per test, outliers (i.e. results that were supposed not to belong to the reference population) were discarded using the Tukey method [30]. In short, the lower and upper cut-offs for outliers were defined as $Q1-(1.5\times IQR)$ and $Q3+(1.5\times IQR)$, respectively, where Q1 is the lower sample quartile, Q3 is the upper sample quartile and IQR = Q3–Q1. When an outlier was detected, results from related tests were discarded as well. For this purpose, the following groups of related tests were defined:

- Electrolytes: calcium, chloride, potassium, sodium
- Bone: calcium, magnesium, phosphate
- Liver: alkaline phosphatase, GGT, ALT, AST, (total) bilirubin
- Kidney: creatinine, urea
- Proteins: albumin, total protein

Histograms were visually inspected and formal tests were performed (Z score for skewness and kurtosis between –1.96 and 1.96 and p value Shapiro-Wilk test >0.05) to determine the presence of a normal Gaussian distribution. Given the large numbers of test results, the formal tests of normality were very sensitive to a deviation from normality. Therefore, decisions on normality were based on the visual inspection of the histograms. If a normal distribution was absent, the data were log-transformed. Again, the Tukey method was applied to discard outliers, histograms were visually inspected and formal tests were performed to check if the assumption of an approximately normal Gaussian distribution was met. Supplementary Material Table 1 shows the total number of available test results (combined for all participating laboratories) per test and an overview of the proportion of excluded test results (range 2%–15%).

Albumin, (anorganic) phosphate, calcium, chloride, potassium, magnesium, sodium, total protein followed a Gaussian distribution. For the other tests, we obtained a Gaussian distribution after log transformation.

Per laboratory, per test, the mean and standard deviation were calculated, for the total dataset, and stratified in the following predefined subgroups (based on consensus):

- Sex: Male/female
- Age:
  - Newborns/infants: <28 days of age (WHO definition), 28 days to <1 year
  - 1–5 years, 6–12 years, 12–18 years, 19–50 years, 51–65 years, 66–80 years, 80+ years.

Results were included in the pooled analyses when a minimum of five test results per laboratory were available.

Second, per test, average means and average SDs of all participating laboratories were calculated to generate pooled (mean ± 2 SD) reference intervals. The means and SDs of the individual participating laboratories were all considered of equal value, and were considered equally accurate reflections of the underlying true distributions of the test results for patients visiting general practitioners in the catchment areas of the participating laboratories. Therefore, no weights based on sample size were assigned. The pooled reference intervals were considered for implementation when they were based on at least 120 measurements in total.

Third, the calculated pooled reference intervals were verified in five representative and independent datasets from participating laboratories (additional data extraction period 1 December 2016 – 1 June 2017) covering all manufacturers and matrices. The percentage of measurements below the lower reference limit and above the upper reference limit were calculated per test per laboratory, further referred to as "flagging rates". Theoretically, in case of suitable standardized reference intervals, 2.5% flagging rates below, respectively, above the defined lower reference limit and upper reference limit are expected.

Data analyses were performed using IBM SPSS Statistics 23.

Expert meetings

Expert meetings were organized with all participating laboratories. The expert meetings took place on February 9th 2017, May 30th 2017 and September 5th 2017. During the first workshop, the general concepts of (inter)national standardization of post-analysis for clinical chemistry tests were discussed, consensus was obtained about the methods and data analysis plan, and the first results were discussed during breakout sessions. The aims of the second workshop were (1) to check for the availability of reference materials and reference methods in the JCTLM database and (2) discuss the need for partitioning based on manufacturer, matrix, sex and age based on a literature search and the results of the data analyses. During the third workshop, all results (per test, per laboratory, per subgroup) and flagging rates were discussed during breakout sessions and consensus-based decisions were made about the applicability of the calculated standardized reference intervals. Minutes were recorded of all meetings.

Results

Sixteen clinical laboratories across the country provided anonymous test results from patients visiting general practitioners over a period of 1 year, leading to a total of 7,574,327 analyzed results. The 16 laboratories represented the complete geographical area of the Netherlands, the four main clinical chemistry analyzer manufacturers and both heparin plasma and serum matrices (Figure 2). Thirteen laboratories used lithium heparin collection tubes, and three used serum tubes, either with or without clotting activator (Becton Dickinson or Greiner). The majority of laboratories (85%) used tubes with a gel for separation. Besides providing the test results for the data analyses and additional information on pre-analytical conditions and test characteristics, the laboratory specialists in clinical chemistry, residents in clinical chemistry or senior technical staff of the participating laboratories took seat in the NUMBER expert meetings.

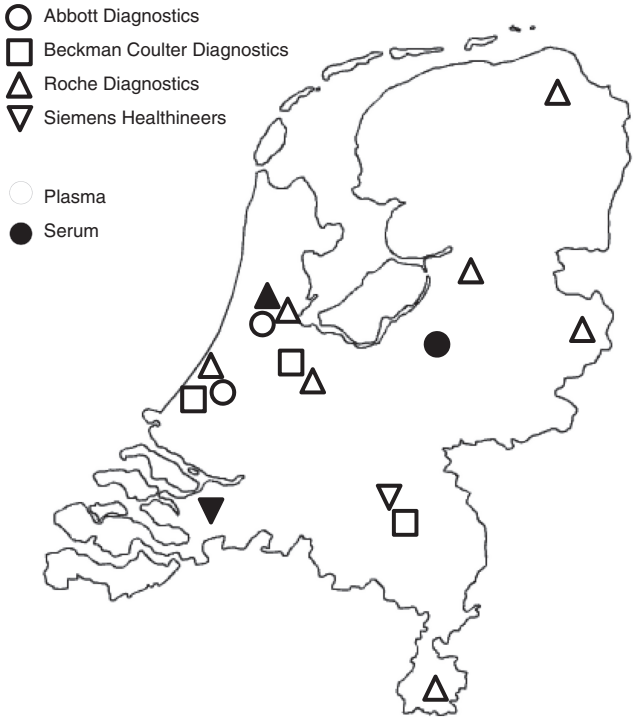

Figure 2: Map of the Netherlands indicating the location, platform and matrix of the participating laboratories.

During the three expert meetings organized in 2017, consensus was obtained about (a) the standardized reference intervals for ALT, albumin, ALP, amylase, (anorganic) phosphate, AST, (total) bilirubin, calcium, chloride, GGT, iron, potassium, creatinine, LD, magnesium, sodium, total protein and urea, and (b) the need for partitioning in sex, age, matrix or method categories for each analyte. An overview of the standardized reference intervals is presented in Table 1.

Flagging rates were calculated in five representative independent datasets to verify the plausibility of the calculated reference intervals using the 'big data' approach. Flagging rates were all within the plus or minus 5% (±2%) range (Figure 3), which was considered acceptable during the expert meetings (consensus decision). The higher flagging rates below the lower limit of the reference interval for iron were expected, as no hemoglobin data were available in the validation data sets.

No consensus has been reached yet for implementation of the calculated reference intervals for creatinine kinase and uric acid, as these reference intervals were substantially higher than currently applied in the participating laboratories and higher than to be expected in healthy individuals. For glucose (fasting and non-fasting), it was decided in expert meeting 1, in compliance with the Milan hierarchy [18], to recommend the use of outcome-based clinical decision limits rather than a reference interval, such as provided by the WHO and American Diabetes Association [31, 32]. As a result of this decision, we did not calculate standardized reference intervals for glucose.

We observed substantial differences in the calculated standardized reference intervals using data from laboratories using the BCG and the BCP method for albumin, confirming the need to introduce separate reference intervals for BCG and BCP (Figure 4). In addition, we observed larger variation in currently applied reference intervals for potassium (based on manufacturers' product inserts, literature, or healthy controls analyzed by the laboratory itself) in the participating laboratories than in the calculated reference intervals using the 'big data' approach with separate

Table 1: Reference intervals as calculated in the NUMBER project for selected medical tests.

Test	Unit	Gender	Age, years	n	Proposed reference interval in:					
					Serum + plasma		Plasma		Serum	
					Low	High	Low	High	Low	High
ALT	U/L	M	1–5	1376	9	35				
			6–12	3796	10	35				
			13–18	8857	9	42				
			19–50	98,326	13	71				
			51–65	103,797	13	64				
			66–80	88,210	11	54				
			81+	20,462	9	42				
		F	1–5	1112	10	33				
			6–12	4206	9	33				
			13–18	18,502	8	35				
			19–50	157,705	9	44				
			51–65	121,128	11	52				
			66–80	103,883	10	46				
			81+	34,073	8	37				
Albumin – BCP	g/L	M	1–5	191	33	44				
			6–12	506	34	46				
			13–18	1096	35	48				
			19–50	9931	35	47				
			51–65	10,205	33	45				
			66–80	10,459	31	44				
			81+	3228	30	42				
		F	1–5	139	34	46				
			6–12	559	36	45				
			13–18	2243	34	46				
			19–50	19,539	32	45				
			51–65	14,217	33	44				
			66–80	13,126	32	43				
			81+	5998	31	42				
Albumin – BCG	g/L	M	1–5	24	nd	nd				
			6–12	111	nd	nd				
			13–18	236	40	52				
			19–50	4636	39	51				
			51–65	4468	37	49				
			66–80	9907	36	48				
			81+	1833	36	46				
		F	1–5	26	nd	nd				
			6–12	121	39	50				
			13–18	785	40	51				
			19–50	8132	38	49				
			51–65	6735	38	49				
			66–80	6835	37	48				
			81+	3646	36	47				
Amylase	U/L	M + F	1–5	32	nd	nd				
			6–12	257	30	128				
			13–18	1029	28	120				
			19–50	16,260	28	119				
			51–65	14,315	27	134				
			66–80	10,738	27	141				
			81+	2864	27	139				
ALP	U/L	M	1–5	50	nd	nd				
			6–12	91	nd	nd				
			13–18	1426	63	190				
			19–50	29,543	45	128				
			51–65	29,777	46	126				
			66–80	24,695	44	134				
			81+	5698	45	145				

Table 1 (continued)

Test	Unit	Gender	Age, years	n	Proposed reference interval in:					
					Serum + plasma		Plasma		Serum	
					Low	High	Low	High	Low	High
		F	1–5	40	nd	nd				
			6–12	189	107	209				
			13–18	3745	44	149				
			19–50	44,377	38	123				
			51–65	36,555	48	142				
			66–80	29,028	47	138				
			81+	8849	45	146				
(Anorganic) phosphate	mmol/L	M	1–5	16	nd	nd				
			6–12	64	nd	nd				
			13–18	215	0.88	1.53				
			19+	12,671	0.62	1.32				
		F	1–5	14	nd	nd				
			6–12	87	nd	nd				
			13–18	801	0.82	1.52				
			19+	19,688	0.73	1.40				
AST	U/L	M	1–5	290	24	53				
			6–12	937	19	42				
			13–18	2710	14	39				
			19+	117,620	14	43				
		F	1–5	233	23	49				
			6–12	1009	15	46				
			13–18	5660	12	33				
			19+	150,029	13	38				
(Total) bilirubin	µmol/L	M	1–5	279	2	11				
			6–12	835	3	16				
			13–18	2005	4	29				
			19+	62,721	4	24				
		F	1–5	200	2	10				
			6–12	880	3	16				
			13–18	3979	3	22				
			19+	91,411	3	19				
Calcium	mmol/L	M + F	1–5	104	nd	nd				
			6–12	408	2.29	2.56				
			13–18	1564	2.23	2.57				
			19+	75,237	2.18	2.55				
Chloride	mmol/L	M + F	13+	8963	97	108				
Creatinine	µmol/L	M	1–5	81	nd	nd				
			6–12	3505	38	63				
			13–18	9509	48	101				
			19–50	140,432	61	113				
			51–65	246,639	61	120				
			66–80	253,514	62	134				
			81+	56,456	65	149				
		F	1–5	86	nd	nd				
			6–12	3988	38	65				
			13–18	20,251	46	83				
			19–50	214,100	48	91				
			51–65	261,732	48	99				
			66–80	289,028	48	113				
			81+	102,806	49	132				
GGT	U/L	M	1–5	822	6	17				
			6–12	2592	7	23				
			13–18	6546	7	36				
			19–50	79,249	9	102				
			51–65	78,949	11	117				
			66–80	63,516	10	110				

Table 1 (continued)

Test	Unit	Gender	Age, years	n	Proposed reference interval in:					
					Serum + plasma		Plasma		Serum	
					Low	High	Low	High	Low	High
			81+	15,431	8	105				
		F	1–5	703	6	19				
			6–12	2763	7	22				
			13–18	13,326	6	35				
			19–50	121,881	6	62				
			51–65	90,963	7	90				
			66–80	75,431	8	91				
			81+	26,037	7	89				
Iron	µmol/L	M	1–5	40	nd	nd				
			6–12	325	7	30				
			13–18	761	9	35				
			19–50	4312	8	39				
			51–65	3883	8	35				
			66–80	3694	7	31				
			81+	951	8	29				
		F	1–5	113	nd	nd				
			6–12	593	7	31				
			13–18	2219	6	35				
			19–50	14,512	7	35				
			51–65	6188	8	30				
			66–80	4138	8	30				
			81+	1279	7	28				
LD	U/L	M	1–5	155	197	329				
			6–12	435	168	307				
			13–18	1032	127	273				
		F	1–5	110	nd	nd				
			6–12	485	148	313				
			13–18	1851	116	234				
		M + F	19–50	32,031	118	251				
			51–65	28,328	128	274				
			66–80	30,239	130	282				
			81+	13,894	136	301				
Magnesium	mmol/L	M + F	6+	38,315	0.71	0.98				
Potassium	mmol/L	M + F	6+	1,012,436 (plasma), 408,045 (serum)			3.4	4.9	3.8	5.2
Sodium	mmol/L	M + F	1+	923,427	136	145				
Total protein	g/L	M + F	1+	14,805 (plasma), 55,457 (serum)			63	81	61	79
Urea	µmol/L	M	1–5	8	nd	nd				
			6–12	287	2.8	7.1				
			13–18	822	2.7	7.6				
			19–50	13,389	2.9	8.6				
			51–65	19,934	3.2	9.8				
			66–80	22,991	3.6	11.8				
			81+	7113	4.3	14.2				
		F	1–5	6	nd	Nd				
			6–12	334	2.4	7.3				
			13–18	1820	2.2	6.9				
			19–50	19,180	2.5	7.0				
			51–65	21,143	2.9	9.4				
			66–80	24,835	3.1	11.8				
			81+	10,883	3.6	13.9				

nd – not determined due to n < 120.

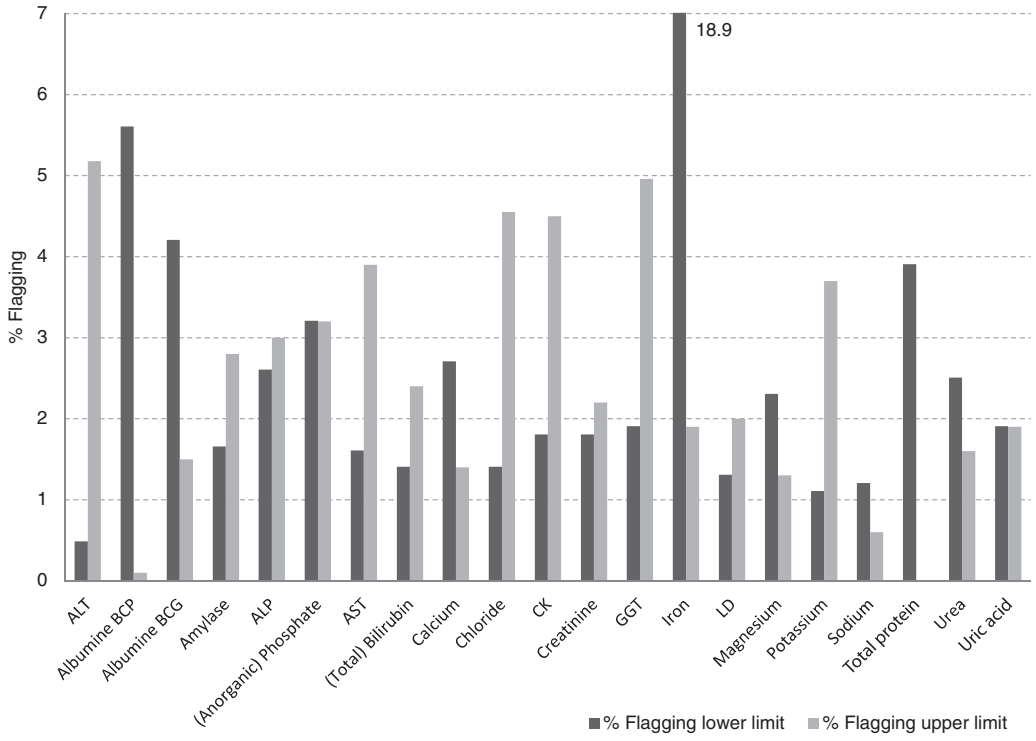

Figure 3: Percentage flagging rates when applying standardized reference intervals in an independent dataset.

reference intervals for plasma and serum matrices (Figure 5). The observed differences in reference intervals for laboratories using heparin plasma and serum for potassium and total protein were in accordance with previous studies [26–28]. For plasma/serum ALT, ALP, AST, GGT and creatinine, considerable differences were found between the different age groups (Figure 6).

Discussion

Building on the successes of the Dutch National Endeavor Calibration 2000 (now called Calibration 2.000) [11, 33] and inspired by the evidence-based approach that was applied in Australia and New Zealand [4], we set up a project and road map for sustainable determination of actual reference intervals, eventually stratified by gender and age, using a 'big-data' approach and medical test results from patients visiting their general practitioners. As metrological traceability of test results produced by medical laboratories in the Netherlands is guaranteed through the category 1 EQA scheme with value-assigned trueness verifiers, standardized reference intervals should be feasible. The fully commutable, targeted EQA materials that have been introduced in the Netherlands since 1998, in combination with the MUSE scoring system [17], have significantly reduced inter-laboratory variation for electrolytes, substrates and enzymes in the Netherlands [1], and has led to significant, further harmonization of test results [34]. Unfortunately, the standardization of tests has not yet led to harmonization of the total test process, because of substantial variation in both the pre-analytical phase and post-analytical phase. Reference intervals that are currently applied by Dutch medical laboratories originate from variable sources such as literature, manufacturer inserts, data mining from the laboratory information system, or regional agreements, leading to higher variation in reference intervals than the variation of analytical results from the different clinical chemistry platforms [2–4]. This situation is suboptimal and may lead to misinterpretation of the generated results by treating physicians, additional unnecessary diagnostic evaluations, over- or under-treatment or inaccurate estimations about prognosis [3, 4]. The use of standardized reference intervals for medical tests for which the traceability chain is in place, is also an essential prerequisite for standardizing the total test process [35]. Standardized reference intervals will facilitate the uniform interpretation of test results in the case of referral of the patient to a different hospital or from primary to secondary care and vice versa. The upcoming digitalization in the medical world, and the development of national or personal

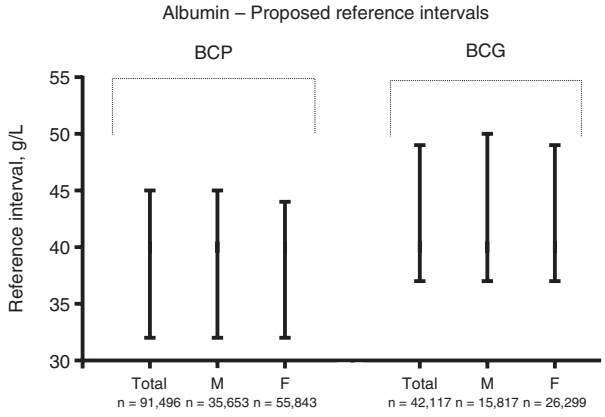

Figure 4: Reference intervals (calculated as mean ± 2 SD) for serum/plasma albumin stratified by method (BCP vs BCG).

health records is another driver for standardization of the total test process.

Because the establishment of reference intervals according to CLSI protocol EP28-A3c [5] and C-RIDL [7] is a time consuming, unsustainable and a costly expedition, we searched for an efficient and future proof approach for periodic (re-)establishment of standardized reference intervals in the Netherlands. We therefore adopted the evidence-based approach from Australia and New Zealand using existing data from patients visiting general practitioners to establish the reference intervals using a 'big data' approach. The calculated standardized reference intervals in our project are comparable to the reference intervals found in other important reference interval studies such as CALIPER (direct method) [36], ARIA (indirect method) [4] and NORIP (direct method) [8] (Supplementary Material Table 2), supporting the chosen approach. The reference intervals that we propose can be implemented in the near future by all Dutch laboratories if they (a) adhere to the 'Venipuncture' guideline of the Dutch Society for Clinical Chemistry and Laboratory Medicine [19], (b) have comparable pre-analytical work-up conditions (such as collection tubes, transportation, temperature monitoring and centrifugation conditions) as applied by the participating laboratories in this project, (c) are using selective tests with clearly defined *measurands* that are traceable to JCTLM-listed reference methods and/or reference materials, and (d) have a MUSE performance score of at least one (sigma 2 or higher) [17] in the SKML EQA scheme (Supplementary Material Figure 1).

Interestingly, our study shows significant age and/or lifestyle effects for ALT, ALP, AST, GGT and creatinine. The proposed standardized reference intervals for these tests are substantially higher than the (IFCC) reference intervals that are commonly applied for these tests in the Netherlands [37–42]. The results with respect to creatinine build on evidence from earlier studies and are likely to reflect an aging effect. It is well known that in old age renal function decreases as a result of loss of glomeruli and decline in renal blood flow [43, 44]. This is illustrated by results from the Leiden 85-plus Study, a population-based prospective follow-up study of 85 year olds in Leiden The Netherlands, in which median creatinine clearance (estimated by the Cockcroft-Gault formula at the time) at age 85 was 45 mL/min [44]. However, the age-dependent increase found in the reference intervals for ALT, AST and GGT might be due to natural aging of the organs and/or due to lifestyle of the Dutch population, where obesity, metabolic syndrome, lack of exercise and alcohol consumption are prevalent [45, 46]. The elevated reference intervals are not found in the oldest age category (81+), suggesting that aging is not solely responsible for the increase in reference intervals for these enzymes. Data are increasingly becoming available that question the extrapolation of 'common' medical knowledge into the highest age groups. The effects of some classical (laboratory) determinants of disease and mortality in middle age, such as hypothyroidism [47], hypercholesterolemia

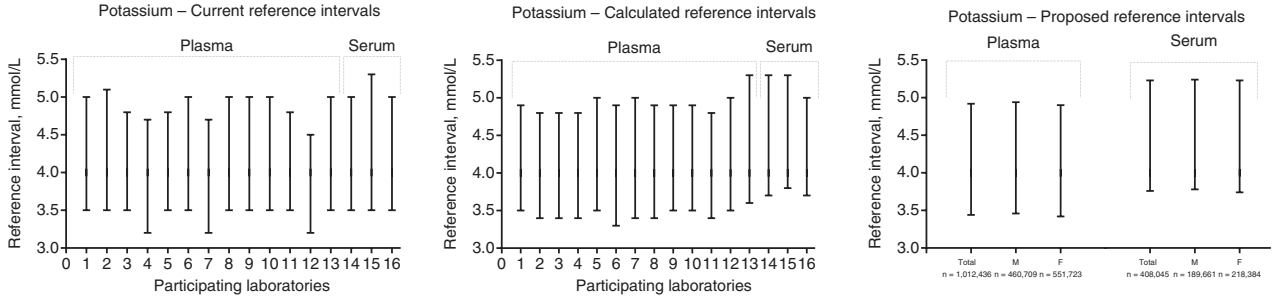

Figure 5: Current reference intervals for potassium in participating laboratories (left panel), calculated reference intervals for potassium in participating laboratories (middle panel) and proposed standardized reference intervals (right panel) for potassium in serum and heparin plasma.

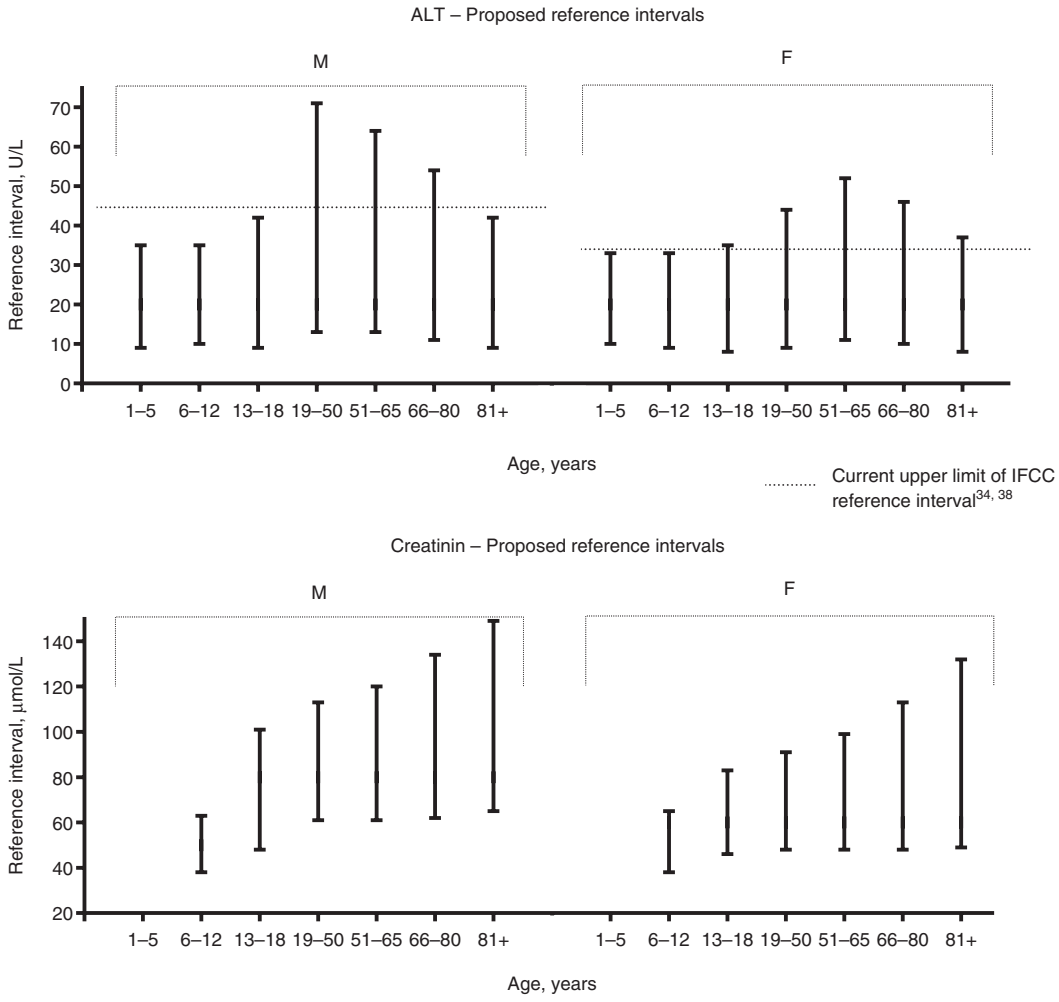

Figure 6: Age and sex effect on the reference intervals for ALT and creatinine.

[48] and elevated ALT [49–51], have been shown to disappear or even reverse in the oldest old, indicating complex relationships between common laboratory values and clinical outcomes in the elderly. This is an important topic of further research. The increased upper limits of the liver enzymes compared to the IFCC reference intervals that are commonly applied in the Netherlands [37, 39–42, 52] was discussed extensively during the expert meetings. Even though a lifestyle component may be involved, our results are supported by the flagging rate analyses. This is important, as frequent flagging (marked on the laboratory rapport) may distract attention from true pathological results [9].

Our study has several strengths. First, compared to the direct method of establishing reference intervals, the applied 'big data' approach in NUMBER is cost-efficient in the sense that it avoids collection and analyzing material from healthy control donors. Secondly, the chosen indirect method and statistical methodology has the potential to be a sustainable method that can be repeated at set time intervals to re-evaluate the applied reference intervals in the future, to identify effects of population changes due to epidemiological transitions, exposure to the microbiome, aging and increasing welfare or advances in pre-analytic and analytic methods. When centrally organized, this will facilitate the Dutch laboratories in their duty to monitor and evaluate applied reference intervals conform ISO15189. Third, the participating laboratories represented the complete geographical area of the Netherlands, the four main clinical chemistry analyzer manufacturers and both plasma and serum matrices. This is of vital importance for the national implementation of the standardized reference intervals that were obtained from this project. Fourth, the Dutch EQA system uses, where possible, commutable, targeted materials. This is a minimum requirement for determining standardized reference intervals, applicable to all laboratories that show a Six Sigma score of at least 2 in the MUSE scoring system.

Some limitations also have to be acknowledged. First, as a consequence of using anonymous laboratory test results, clinical information was not available. We tried to select a 'healthy' population by (1) using laboratory tests ordered by general practitioners (that in general order blood tests to rule out disease, meaning that the *a priori* chance of illness is low), assuming that the majority of individuals are free of disease, (2) we excluded laboratory results when phlebotomy was performed at home, (3) used large numbers of test results in order to have confidence about the distribution of the central majority of the cohort [53], and (4) excluded results from related tests when results that were supposed not to belong to the reference population were detected. However, we cannot fully exclude the possibility of selection bias, especially for those tests that are routinely performed to monitor patients with chronic diseases (like risk management for cardiovascular disease) or are performed only for a very distinct indication. This is, in particular, the reason why we did not reach consensus yet for the implementation of the calculated reference intervals for CK and uric acid. As recommended in the GP guideline 'Cardiovascular Risk Management' of the Dutch College of General Practitioners [54], CK is commonly measured in patients using cholesterol synthesis inhibitor (statins), which are known for side effects such as myopathy [55, 56] and are used by more than 11% of the Dutch population [57]. In the case of uric acid, there is no universal definition for hyperuricemia. A cut-off defined by the solubility limit of uric acid (>0.42 mmol/L for men and >0.34 mmol/L for women) is commonly applied [58], based on the study by Campion and colleagues [59]. Others have already suggested that the reference intervals for uric acid may need to be redefined, because of a 'shift to the right' [60]. However, as uric acid is primarily requested for the diagnostic indication gout/arthritis, as is recommended in the Dutch GP Guideline 'Arthritis' [61], our dataset with results from patients visiting general practitioners might not be representative for the normal/healthy population for this test. Therefore, the preliminary calculated standardized reference intervals for CK and uric acid have to be evaluated in a healthy control group or population-based study such as LifeLines [62] before national implementation can take place. Second, the data mining method requires large numbers of individuals [53]. Although more than seven million test results were available from 16 participating laboratories, in some subgroups still very small numbers of tests results were available. Therefore, it was not always possible to propose reference intervals for the youngest age groups. A different representative population should be sought for the youngest age categories (0–1 year and 1–5 years). The CALIPER study determined reference intervals in children separately on several platforms, but did not standardize reference intervals across methods [36]. Zierk and colleagues applied an indirect method to generate pediatric reference intervals for biochemistry analytes analyzed on one platform during clinical care in a tertiary care center in Germany [63, 64]. The Canadian and German results can therefore not be translated to the Dutch medical laboratory setting. However, the continuous reference intervals that were provided by both projects are a topic of further study in the Netherlands.

In conclusion, with the use of a big data approach we were able to calculate age and sex stratified standardized reference intervals for 18 clinical chemistry tests that can be implemented by each Dutch laboratory with a Sigma score of at least 2 in the SKML EQA program and platform with a full traceability chain in place. Nationwide implementation of these established reference intervals has the potential to improve unequivocal interpretation of test results leading to better diagnosis, treatment, risk stratification and follow-up. As a next step, a fourth expert meeting with the ambassadors of the NUMBER initiative will be organized early 2018 to come to agreement for national reference intervals for CK and uric acid after evaluation in a healthy control group. In addition, promotion material will be developed and a detailed plan for the implementation of the reference intervals will be established (Figure 1, phase 5). The working group and ambassadors nationally and locally promote adherence to the standardized reference intervals in all Dutch laboratories according to a national implementation plan. The promotion of the use of the national reference intervals will not only lead to better exchangeability of laboratory results between different medical laboratories, but will also enhance the harmonization of methods, because the laboratory specialists are encouraged to calibrate their methods to achieve a bias of almost zero. Moreover, in cooperation with SKML and Calibration 2.000, our ambition is to formulate formal recommendations regarding more selective methods, e.g. enzymatic method for creatinine and the BCP method for albumin, to achieve the most reliable and exchangeable results for these parameters. Furthermore, we will set up a sustainable national surveillance system to structurally monitor constancy in reference intervals with time in cooperation with SKML using their existing digital infrastructure for all medical laboratories in the Netherlands (Figure 1, phase 6). When this is accomplished, not only biological effects on reference intervals but also the effects of evolution of test performance and test standardization on the reference interval will be periodically evaluated.

Acknowledgments: We thank the members of the NUMBER working group for providing anonymous clinical laboratory data and technical information on the laboratory tests, for their active participation in expert meetings and advice on the project outcomes.

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

Research funding: This project is supported by a grant from Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek.

Employment of leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization 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.

NUMBER working group: Dr. G. den Besten (Isala); Dr. V.M. Bruinenberg (Certe locatie Martini Ziekenhuis); Dr. M.W.H.J. Demmers (Gelre Apeldoorn); Dr. I.M. Dijkstra (Antonius Ziekenhuis); Dr. M.A.M. Frasa (LangeLand Ziekenhuis); Dr.ir. R.L.J.M. Herpers (Bernhoven); Dr.ir. M.B. Kok (Saltro); Dr. J.G. Krabbe (MEDLON); Dr. E. Kusters (HMC Bronovo); Dr. A. Leyte (OLVG); Dr. D. van Loon (Antonius Ziekenhuis); C.H. Meijer-Huissen (SALT); Dr. K. Mohrmann (Star-SHL); T.L. Njo (Saltro); Dr. W.P. Oosterhuis (Zuyderland locatie Heerlen); Dr. M.C.F.J. de Rotte (SCAL Medische Diagnostiek); Dr. R.C. Rotteveel (Atalmedial); Dr. C.H.H. Schoenmakers (Elkerliek Ziekenhuis); Dr. B.A. Wevers (Atalmedial); Dr.ir. J.J. van Zanden (Certe); Dr. J.C. Zant (Zuyderland locatie Heerlen).

References

1. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples – five year experience from the Netherlands. Clin Chim Acta 2012;414:234–40.
2. Berg J, Lane V. Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem 2011;48(Pt 3):195–7.
3. Jones GR, Barker A, Tate J, Lim CF, Robertson K. The case for common reference intervals. Clin Biochem Rev 2004;25:99–104.
4. Tate JR, Sikaris KA, Jones GR, Yen T, Koerbin G, Ryan J, et al. Harmonising adult and paediatric reference intervals in Australia and New Zealand: an evidence-based approach for establishing a first panel of chemistry analytes. Clin Biochem Rev 2014;35:213–35.
5. CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline – Third Edition. CLSI document C28-A3c. 2008:1–76.
6. Ceriotti F. Common reference intervals: the IFCC position. Clin Biochem 2009;42:297.
7. Ozarda Y, Ichihara K, Barth JH, Klee G, Committee on Reference I, Decision Limits IFfCC, 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.
8. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Martensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scand J Clin Lab Invest 2004;64:271–84.
9. Tate JR, Koerbin G, Adeli K. Opinion paper: deriving harmonised reference intervals – global activities. EJIFCC 2016;27:48–65.
10. Jones GR, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination – review and recommendations. Clin Chem Lab Med 2018. DOI: https://doi.org/10.1515/cclm-2018-0073. [Epub ahead of print].
11. Jansen RT, Kuypers AH, Baadenhuijsen H, van den Besselaar AM, Cobbaert CM, Gratama JW, et al. Kalibratie 2000. Ned Tijdschr Klin Chem 2000;25:153–8.
12. Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. J Clin Chem Clin Biochem 1985;23:829–39.
13. Koerbin G, Tate JR. Harmonising adult reference intervals in Australia and New Zealand – the continuing story. Clin Biochem Rev 2016;37:121–9.
14. Koerbin G, Sikaris K, Jones GR, Flatman R, Tate JR, Intervals AHCfCR. An update report on the harmonization of adult reference intervals in Australasia. Clin Chem Lab Med 2018. DOI: https://doi.org/10.1515/cclm-2017-0920. [Epub ahead of print].
15. Greg Miller W, Myers GL, Lou Gantzer M, Kahn SE, Schonbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57:1108–17.
16. Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, et al. A category 1 EQA scheme for comparison of laboratory performance and method performance: an international pilot study in the framework of the Calibration 2000 project. Clin Chim Acta 2014;432:90–8.
17. Thelen MH, Jansen RT, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multisample evaluation in laboratory EQA. Clin Chem Lab Med 2017;55:1509–16.
18. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–5.
19. Werkgroep Preanalyse van de NVKC, van Dongen-Lases EC, Eppens EF, de Jonge N, Roelofs-Thijssen MA. Richtlijn Veneuze Bloedafname. 2013. Available at: https://www.nvkc.nl/sites/default/files/2.pdf.
20. Cornes M, van Dongen-Lases EC, Grankvist K, Ibarz M, Kristensen G, Lippi G, et al. Order of blood draw: Opinion Paper by the European Federation for Clinical Chemistry and Laboratory Medicine (EFLM) Working Group for the Preanalytical Phase (WG-PRE). Clin Chem Lab Med 2017;55:27–31.
21. 'Kengetallen Nederlandse Ziekenhuizen 2014'. 2016. Available at: https://www.nvz-ziekenhuizen.nl/_library/33659/RapportageKengetallen2014.pdf.
22. Statistics Netherlands. 'Meeste Nederlanders wonen binnen 5 kilometer van ziekenhuis'. CBS StatLine 2009.

23. Statistics Netherlands. 'Nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers'. CBS StatLine 2016.
24. Jassam N, Weykamp C, Thomas A, Secchiero S, Sciacovelli L, Plebani M, et al. Post-standardization of routine creatinine assays: are they suitable for clinical applications. Ann Clin Biochem 2017;54:386–94.
25. Bachmann LM, Yu M, Boyd JC, Bruns DE, Miller WG. State of harmonization of 24 serum albumin measurement procedures and implications for medical decisions. Clin Chem 2017;63:770–9.
26. Nijsten MW, de Smet BJ, Dofferhoff AS. Pseudohyperkalemia and platelet counts. N Engl J Med 1991;325:1107.
27. Schindler EI, Brown SM, Scott MG. Electrolytes and blood gases. In: Rifai N, Horvath AR, Wittwer CT, editors. Tietz textbook of clinical chemistry and molecular diagnostics. 6th ed. St. Louis, Missouri: Elsevier, 2018:604–25.
28. Dietzen DJ. Amino Acids, Peptides, and Proteins. In: Rifai N, Horvath AR, Wittwer CT, editors. Tietz textbook of clinical chemistry and molecular diagnostics. 6th ed. St. Louis, Missouri: Elsevier, 2018:373–403.
29. Nutritional anaemias. Report of a WHO scientific group. World Health Organ Tech Rep Ser 1968;405:5–37.
30. Horn PS, Pesce AJ. Reference intervals: an update. Clin Chim Acta 2003;334:5–23.
31. American Diabetes A. 2. Classification and diagnosis of diabetes. Diabetes Care 2017;40(Suppl 1):S11–24.
32. WHO. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation 2006.
33. Jansen RT. The quest for comparability: Calibration 2000. Accredit Qual Assur 2000;5:363–6.
34. Jansen RT, Cobbaert CM, Weykamp C, Thelen M. The quest for equivalence of test results: the pilgrimage of the Dutch Calibration 2.000 program for metrological traceability. Clin Chem Lab Med 2018. DOI: https://doi.org/10.1515/cclm-2017-0796. [Epub ahead of print].
35. Plebani M, Panteghini M. Promoting clinical and laboratory interaction by harmonization. Clin Chim Acta 2014;432:15–21.
36. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, et al. Biochemical marker reference values across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian Health Measures Survey. Clin Chem 2015;61:1049–62.
37. Franck PF, Spaans A, Postma J. Het Nederlandse Referentielaboratorium voor Enzymen. Ned Tijdschr Klin Chem 2010;35:240–3.
38. Ceriotti F, Boyd JC, Klein G, Henny J, Queralto J, Kairisto V, et al. Reference intervals for serum creatinine concentrations: assessment of available data for global application. Clin Chem 2008;54:559–66.
39. Schumann G, Bonora R, Ceriotti F, Ferard G, Ferrero CA, Franck PF, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. International Federation of Clinical Chemistry and Laboratory Medicine. Part 6. Reference procedure for the measurement of catalytic concentration of gamma-glutamyltransferase. Clin Chem Lab Med 2002;40:734–8.
40. Schumann G, Bonora R, Ceriotti F, Ferard G, Ferrero CA, Franck PF, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. International Federation of Clinical Chemistry and Laboratory Medicine. Part 5. Reference procedure for the measurement of catalytic concentration of aspartate aminotransferase. Clin Chem Lab Med 2002;40:725–33.
41. Schumann G, Bonora R, Ceriotti F, Ferard G, Ferrero CA, Franck PF, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. International Federation of Clinical Chemistry and Laboratory Medicine. Part 4. Reference procedure for the measurement of catalytic concentration of alanine aminotransferase. Clin Chem Lab Med 2002;40:718–24.
42. Schumann G, Klauke R, Canalias F, Bossert-Reuther S, Franck PF, Gella FJ, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. Part 9: reference procedure for the measurement of catalytic concentration of alkaline phosphatase International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Scientific Division, Committee on Reference Systems of Enzymes (C-RSE) (1)). Clin Chem Lab Med 2011;49:1439–46.
43. Epstein M. Aging and the kidney. J Am Soc Nephrol 1996;7:1106–22.
44. van Bemmel T, Woittiez K, Blauw GJ, van der Sman-de Beer F, Dekker FW, Westendorp RG, et al. Prospective study of the effect of blood pressure on renal function in old age: the Leiden 85-Plus Study. J Am Soc Nephrol 2006;17:2561–6.
45. Reinders I, van Schoor NM, Deeg DJ, Huisman M, Visser M. Trends in lifestyle among three cohorts of adults aged 55–64 years in 1992/1993, 2002/2003 and 2012/2013. Eur J Public Health 2017. doi: 10.1093/eurpub/ckx173. [Epub ahead of print].
46. Slagter SN, van Waateringe RP, van Beek AP, van der Klauw MM, Wolffenbuttel BH, van Vliet-Ostaptchouk JV. Sex, BMI and age differences in metabolic syndrome: the Dutch Lifelines Cohort Study. Endocr Connect 2017;6:278–88.
47. Gussekloo J, van Exel E, de Craen AJ, Meinders AE, Frolich M, Westendorp RG. Thyroid status, disability and cognitive function, and survival in old age. JAMA 2004;292:2591–9.
48. Weverling-Rijnsburger AW, Blauw GJ, Lagaay AM, Knook DL, Meinders AE, Westendorp RG. Total cholesterol and risk of mortality in the oldest old. Lancet 1997;350:1119–23.
49. Le Couteur DG, Blyth FM, Creasey HM, Handelsman DJ, Naganathan V, Sambrook PN, et al. The association of alanine transaminase with aging, frailty, and mortality. J Gerontol A Biol Sci Med Sci 2010;65:712–7.
50. Ford I, Mooijaart SP, Lloyd S, Murray HM, Westendorp RG, de Craen AJ, et al. The inverse relationship between alanine aminotransferase in the normal range and adverse cardiovascular and non-cardiovascular outcomes. Int J Epidemiol 2011;40:1530–8.
51. Elinav E, Ackerman Z, Maaravi Y, Ben-Dov IZ, Ein-Mor E, Stessman J. Low alanine aminotransferase activity in older people is associated with greater long-term mortality. J Am Geriatr Soc 2006;54:1719–24.
52. Weykamp C, Franck P, Gunnewiek JK, de Jonge R, Kuypers A, van Loon D, et al. Harmonisation of seven common enzyme results through EQA. Clin Chem Lab Med 2014;52:1549–55.
53. Koerbin G, Sikaris KA, Jones GR, Ryan J, Reed M, Tate J, et al. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014;432:99–107.
54. Cardiovasculair risicomanagement (Tweede herziening). Huisarts en Wetenschap 2012;55:14–28.
55. Law M, Rudnicka AR. Statin safety: a systematic review. Am J Cardiol 2006;97:52C–60C.

56. Stroes ES, Thompson PD, Corsini A, Vladutiu GD, Raal FJ, Ray KK, et al. Statin-associated muscle symptoms: impact on statin therapy-European Atherosclerosis Society consensus panel statement on assessment, aetiology and management. Eur Heart J 2015;36:1012–22.
57. The Drug Information System of the National Health Care Institute (GIPdatabank Zorginstituut Nederland). Statin users in the Netherlands 2011–2015 (ATC-subgroup C10A) 2015. Available at: https://www.gipdatabank.nl/.
58. Becker AM. Asymptomatic hyperuricemia. UpToDate 2017. Accessed: 20 Oct 2017.
59. Campion EW, Glynn RJ, DeLabry LO. Asymptomatic hyperuricemia. Risks and consequences in the Normative Aging Study. Am J Med 1987;82:421–6.
60. Desideri G, Castaldo G, Lombardi A, Mussap M, Testa A, Pontremoli R, et al. Is it time to revise the normal range of serum uric acid levels? Eur Rev Med Pharmacol Sci 2014;18:1295–306.
61. Janssens HJ, Lagro HA, Van Peet PG, Gorter KJ, Van der Pas P, Van der Paardt M, et al. Artritis (Eerste versie). Huisarts en Wetenschap 2009;52:439–53.
62. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, Slaets JP, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 2008;23:67–74.
63. Zierk J, Arzideh F, Rechenauer T, Haeckel R, Rascher W, Metzler M, et al. Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. Clin Chem 2015;61:964–73.
64. Zierk J, Arzideh F, Haeckel R, Cario H, Fruhwald MC, Gross HJ, et al. Pediatric reference intervals for alkaline phosphatase. Clin Chem Lab Med 2017;55:102–10.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2018-0462).